université Paris Ouest Nanterre La Défense

UFR de Sciences Économiques, Gestion, Mathématiques et Informatique

École Doctorale Économie, Organisations, Société

Laboratoire EconomiX

INSTITUER LA PERFORMANCE

Une application au travail du médecin

Thèse pour le doctorat en Sciences Économiques

Présentée et soutenue publiquement par

Nicolas DA SILVA

Le mardi 09 décembre 2014

Directeur de thèse :

Philippe BATIFOULIER, Maître de conférences HDR, Université Paris Ouest Nanterre La Défense

Membres du jury

Olivier FAVEREAU, Professeur, Université Paris Ouest Nanterre La Défense

Maryse GADREAU, Professeure émérite, Université de Bourgogne

Florence JANY-CATRICE, Professeure, Université Lille 1 (*rapporteur*)

Bruno VENTELOU, Directeur de recherche au CNRS, GREQAM et INSERM Marseille (*rapporteur*).

Corinne VERCHER, Professeure, Université Paris 13, Sorbonne Paris Cité

UFR de Sciences Économiques, Gestion, Mathématiques et Informatique

École Doctorale Économie, Organisations, Société

Laboratoire EconomiX

INSTITUER LA PERFORMANCE

Une application au travail du médecin

Thèse pour le doctorat en Sciences Économiques

Présentée et soutenue publiquement par

Nicolas DA SILVA

Le mardi 09 décembre 2014

Directeur de thèse :

Philippe BATIFOULIER, Maître de conférences HDR, Université Paris Ouest Nanterre La Défense

Membres du jury

Olivier FAVEREAU, Professeur, Université Paris Ouest Nanterre La Défense

Maryse GADREAU, Professeure émérite, Université de Bourgogne

Florence JANY-CATRICE, Professeure, Université Lille 1 (*rapporteur*)

Bruno VENTELOU, Directeur de recherche au CNRS, GREQAM et INSERM Marseille (*rapporteur*).

Corinne VERCHER, Professeure, Université Paris 13, Sorbonne Paris Cité

Remerciements

Une thèse est avant tout une aventure collective.

Je commencerai par remercier mon directeur de thèse, Philipe Batifoulier. C'est grâce à son soutien inconditionnel que je suis en mesure de rédiger ces lignes aujourd'hui. Quoi que je puisse écrire sur lui, ce serait trop peu. Je me contenterai de dire qu'il un exemple pour moi, tant intellectuellement que moralement.

Les enseignements d'Olivier Favereau ont été déterminants ; ils ont éveillé en moi le goût de l'économie et de l'analyse critique. C'est donc un honneur pour moi qu'il ait accepté de participer à ce jury de thèse. J'ai également eu la chance de côtoyer Maryse Gadreau qui m'a permis d'éclaircir et d'approfondir mes ambitions théoriques. Florence Jany-Catrice et Bruno Ventelou m'ont fait l'honneur d'avoir accepté d'être les « rapporteurs » de ce travail et de m'apporter ainsi toute leur expertise. Je remercie enfin Corinne Vercher de venir représenter l'interdisciplinarité en participant à ce jury de thèse

Je remercie Pierre-Henri Bréchat qui est à l'origine de beaucoup de projets stimulants et le Service Médical de l'Assurance Maladie d'Ile de France, notamment Laetitia Fleury et Jean-Louis Vanhille qui ont contribué à rendre possible le travail empirique. Je tiens à saluer Nicolas Belorgey qui m'a appris les rudiments de l'entretien sociologique.

Je dois beaucoup à Ariane Ghirardello qui a passé de nombreuses heures sur ma thèse et m'a fait découvrir un versant du monde universitaire, enthousiasmant et addictif, dans lequel nous avons échangé des fous rires mémorables. J'adresse ma gratitude aux membres de la « cellule de crise » : Philippe Abecassis, Christine André, Jean-Paul Domin, Florence Gallois, Guillemette de Larquier, Amandine Rauly. Je vous suis reconnaissant pour cette aide inestimable et – surtout – pour les bons moments passés à vos côtés.

J'ai une pensée amicale envers mes collègues doctorants : Ali, Alzbeta, Anne, Audrey, Aziza, Frank, JF, Lara, Nader, Peixin, Raphael, M. Rosell, Silvia, Thomas, Xi (courage, c'est bientôt finit !!), Massimiliano, Tatiana et Tiana. Vous avez fait partie intégrante de ce « bout de vie ». Du bureau T206 au G301 en passant par les kilomètres de couloirs de Nanterre, j'ai été heureux de passer ces années à vos côtés. Merci à Marie qui a contribué plus qu'elle ne le pense à cette thèse. Merci aussi à Ingrid qui a le don d'égailler mes journées.

Gwendoline, tu es quelqu'un de rare.

Béranger, Julien, Keyne, Loïc, Serge, Willy, Xavier, Yohan, et les autres : je suis de retour.

J'adresse toute mon affection à ma famille. Je pense à Mila et Jorge qui ont été présents quand j'en ai eu besoin. Je n'oublie pas ma nouvelle famille qui a également été d'un grand soutien : Evelyne, Alain, Aldo, Annie, Clément, Ornella, Gabriel et le petit Jojo !

Merci à mes parents. Dans les ménages et les chantiers, depuis petit je vous ai toujours vu user votre santé dans le seul but d'offrir à vos enfants une vie meilleure que la vôtre. J'espère être digne de vous.

Mickaël, tu es un grand maintenant.

Enfin, mes derniers mots je les écris pour toi. Toi que j'ai trop souvent délaissé ces derniers temps. Anne, merci de me rendre heureux.

Pour Anne,

Que serais-je sans toi que ce balbutiement.

Sommaire

Introduction générale **11**

Première partie : Inciter les individus à la performance dans le paradigme *mainstream* L'introuvable originalité du travail médical

Chapitre 1 : Le paiement à la performance médicale : héritage de la norme concurrentielle .. **33**

Chapitre 2 : La performance en univers altruiste. Paradoxes de théorie économique et problèmes de politique économique **81**

Chapitre 3 : Payer plus pour motiver moins ? Performance médicale et *crowding out effect* **123**

Deuxième partie : Une évaluation du paiement à la performance médicale. Le difficile passage de la théorie à la pratique

Chapitre 4 : La performance du paiement à la performance médicale. Une revue de littérature **167**

Chapitre 5 : Le paiement à la performance médicale en France : les performants ne sont pas forcément les gagnants **217**

Troisième partie : Du paiement à la performance médicale à la régulation par les pratiques Vers le néolibéralisme sanitaire ?

Chapitre 6 : La médecine libérale française contre l'État social. De la régulation par les prix à la régulation par les pratiques **279**

Chapitre 7 : La qualité des soins est-elle soluble dans la quantification ? **327**

Chapitre 8 : Une analyse conventionnaliste de la régulation industrielle des pratiques médicales **377**

Conclusion Générale **419**

Références bibliographiques **429**

Introduction générale

Le jeudi 25 septembre 2014, Marisol Touraine, Ministre des Affaires sociales, de la Santé, et des droits des femmes, présentait le Programme national de réduction du tabagisme. Chaque année en France, 73 000 personnes meurent des suites de maladies liées au tabac. Selon la ministre, « il s'agit d'une véritable hécatombe, insidieuse, silencieuse, sournoise [qui représente] l'équivalent d'un crash d'avion de ligne, chaque jour, avec 200 passagers à bord ! » (Touraine, 2014, p. 2). La lutte contre le tabagisme est un enjeu majeur de santé publique dont la légitimité ne peut pas être remise en cause.

A cette occasion, plusieurs mesures visant la réduction de la consommation de tabac en France ont été présentées. Parmi elles, l'obligation de vente avec un « paquet neutre », l'interdiction de la publicité pour la cigarette électronique, l'interdiction de fumer dans une voiture avec un enfant de moins de 12 ans, etc. Ces décisions ont été amplement commentées.

Dans l'ensemble des outils présentés, l'un d'entre eux est passé plus inaperçu : rémunérer les médecins pour qu'ils accompagnent le sevrage. Alors que le rôle du médecin dans la sensibilisation face aux risques du tabac est avéré, les patients, même s'ils souhaitent vivement arrêter de fumer, ne reçoivent pas systématiquement des conseils de la part de leur médecin traitant. Pour y remédier, la ministre a annoncé qu'« une part de la rémunération des médecins [devra dépendre] des objectifs de santé publique qu'ils atteignent » (Touraine, 2014, p. 5). Plus précisément, il s'agit d'étendre la Rémunération sur objectif de santé publique (ROSP), déjà en vigueur pour les médecins traitants depuis 2012, à un nouvel objectif : la lutte contre le tabagisme.

La croyance assez largement partagée selon laquelle la rémunération du médecin n'est composée que des honoraires que lui verse le patient en fin de consultation s'avère être fausse. Les médecins sont aussi payés à la « performance ». Le dispositif de paiement à la performance médical (la ROSP) se présente comme une prime en fin d'année pour les médecins éligibles qui obtiennent un niveau minimal de performance. Cette performance est définie par un score composite, lui-même fondé sur le calcul d'une multitude d'indicateurs de qualité du travail médical.

L'objectif concernant le tabac n'ayant pas encore été mis en place, prenons deux exemples issus du dispositif appliqué depuis 2012. Premier exemple : les médecins sont éligibles à une prime à la performance si 80 % de leurs patientes âgées entre 50 et 74 ans ont participé au dépistage (organisé ou individuel) du cancer du sein. Second exemple : les médecins sont éligibles à une prime à la performance si 90 % de leurs prescriptions d'antibiotiques (en nombre de boîtes) est réalisée dans le répertoire des génériques. Dans un cas, la politique publique vise à lutter contre le cancer du sein. Dans l'autre cas, elle vise à réduire les dépenses de santé. Dans les deux cas, le mécanisme de paiement à la performance présente la vertu d'aligner les intérêts du médecin sur les objectifs de santé publique. Ainsi, en 2012, les 85 187 médecins concernés ont reçu en moyenne 4 003 € de prime à la performance, soit un coût total pour l'assurance maladie de 341 millions d'euros (CNAM, 2014).

L'universalité de la performance et ses résistances

Cette logique est bien connue des économistes. Elle repose sur une conception de l'homme économique, réticent à l'effort et opportuniste. L'*homo economicus*, égoïste et calculateur, ne fait rien pour rien et aucune forme d'organisation ne peut échapper à ce qui est présenté comme un constat anthropologique. Cette recherche universelle de l'intérêt personnel recèle son antidote : l'intérêt lui-même. En utilisant de façon adéquate l'intérêt personnel, on peut orienter le comportement des individus et transformer le vice en vertu (Laval, 2007, Favereau, 2004). Il convient d'inciter les agents à neutraliser leur capacité de nuisance en compensant financièrement leur peine ou leur effort. C'est dans cette perspective que l'État social peut jouer sur l'intérêt bien compris des individus pour les convertir au programme public de réduction des dépenses de santé (par la prescription de génériques) ou de promotion de la qualité des soins (par la prévention ou l'accompagnement). La mise en œuvre concrète de cette architecture se traduit par le développement des mécanismes incitatifs et l'État social en est alors réduit à être un « manipulateur d'incitations » (Batifoulier et al., 2007)

Si l'incitation est importante, elle n'est pas suffisante : il reste en effet à définir l'objectif de l'incitation. Dans ce que l'on pourrait appeler le « Moyen-âge de l'incitation », cette dernière pouvait prendre n'importe quelle forme. Par exemple, pour inciter un salarié à travailler durement, il suffisait de lui fournir un salaire élevé – la peur de perdre cette rente suffisant à le motiver. Le défaut majeur de cette situation est le manque d'objectivité sur ce qu'il convient d'inciter : que signifie « travailler durement » ? L'impossibilité de vérifier l'engagement du salarié peut alors s'interpréter comme une forme de « libéralité » qu'il

convient d'évincer. En effet, en incitant mal, les ressources rares seront nécessairement gaspillées.

L'originalité de la période contemporaine porte sur le contenu de l'incitation : la performance. Pour définir cette performance, il faut l'envisager sous son double ancrage étymologique (Jany-Catrice, 2012). D'un côté, la performance relève d'une dynamique, celle, par exemple, de l'acteur en scène (étymologie anglo-saxonne), de l'autre côté, la performance relève du résultat et de l'idée de perfection (étymologie latine). Au total, inciter à la performance, c'est vouloir faire de chacun un vainqueur dans une logique d'excellence. Or, pour évaluer cette performance de manière objective, il faut se doter d'un appareillage statistique considérablement développé. Dans un contexte de défiance généralisée, l'étalon de la performance ne peut être qu'un indicateur chiffré de résultat. Le chiffre prétend en effet à l'objectivité et va chercher à éviter le débat ou le jugement. S'il faut inciter, la seule incitation efficace est une incitation à la performance.

La logique de performance est sans frontière. Dans le secteur privé, l'évaluation du travail se fait de plus en plus systématiquement par l'intermédiaire du rapport au chiffre. Les tableaux de bord et autres systèmes d'information permettent de tenir à jour les performances de chaque salarié au sein de l'entreprise. La place primordiale de l'évaluation individuelle dans les entreprises s'appuie sur la notion de performance qui apparaît comme un « mot drapeau » qui tient sa capacité à fédérer de sa polysémie. La notion de performance est un dispositif de gestion des ressources humaines qui a une fonction stratégique et répond à des visions différentes de l'entreprise (Palpacuer et al., 2010**).** En ayant recours à l'évaluation de la performance du salarié, le management peut chercher à éliminer la subjectivité dans le jugement de la hiérarchie. La performance peut alors être mobilisée pour évincer des salariés jugés les moins performants, moins investis dans l'entreprise et les moins assujettis à sa culture comme on peut l'apprécier dans l'étude du licenciement pour motif personnel (Seignour et al., 2007).

La pratique du *benchmarking* et les mécanismes de paiement à la performance intensifient la logique de performance en organisant la comparaison continue entre les salariés. Si l'indicateur de performance ultime de l'entreprise est le profit, il semble plus difficile de mettre en chiffres le service public. Néanmoins, l'hégémonie des thèses du Nouveau Management Public a conduit aujourd'hui à ce que les méthodes du privé pénètrent la sphère publique. Plutôt que des indicateurs de rentabilité, les fonctionnaires sont évalués et jugés à l'aune d'indicateurs de qualité. La rationalité est toujours la même – inciter à la

performance pour utiliser au mieux les ressources et contrer les comportements opportunistes – mais, l'attaque est d'autant plus forte que ce secteur n'est jamais confronté au juge ultime de l'efficacité économique : le marché. À l'hôpital, à l'université où dans les commissariats, l'État est désormais « sous pression statistique » (Bruno et Didier, 2013).

Bien qu'elle se présente comme une mécanique de la rationalisation de l'usage des ressources – qui peut prétendre vouloir « gaspiller plus » ? –, la logique de performance crée partout où elle s'exerce des « résistances immunitaires ». En remettant en cause le sens du travail, les salariés et les fonctionnaires rejettent les principes de la performance sans pour autant être vulnérables à la critique classique de « résistance corporatiste ». L'opposition à l'emprise de la performance sur le travail ne se résume pas au refus de céder des avantages acquis.

Ainsi, de nombreux travaux ont montré que le règne de la performance engendre une souffrance au travail (Dejours, 1998, 2003, De Gaulejac, 2011, 2009). La compétition généralisée requiert une mobilisation de tous les instants qui laisse bien peu de place à la vie hors-travail – créant des désordres physiques et psychologiques entre vie professionnelle et vie personnelle. L'injonction à la performance s'accompagne paradoxalement d'une invitation à l'autonomie, si bien que le salarié devient son propre surveillant. Malgré les apparences de neutralité et d'impartialité, la performance place les corps et les esprits en souffrance. Au-delà de ce qu'elle fait aux personnes, d'autres critiques de la performance portent plus directement sur son ambition proprement économique. La logique de performance s'accompagne d'une logique de rentabilité « court-termiste » les profits d'aujourd'hui étant sanctuarisés quelles que soient les conséquences pour demain (Lordon, 2003, 2009).

Les résistances prennent une forme particulière au sein du secteur public : l'une des idées transversales est que l'esprit de service public n'est pas réductible à une mise en chiffre. Or, c'est la définition même des métiers qui est remise en cause par l'injonction à la performance. Dans cet optique, l'*Appel des appels* (Gori et al., 2009) défend les métiers « tissant le lien social » face à la logique de performance.

Le médecin comme « exemple exemplaire » des apories de la performance

Dans ce contexte global, le développement de la ROSP en médecine de ville n'est que la division médicale du mouvement généralisé vers « performance totale » (Jany-Catrice, 2012). Pourtant, loin d'être un cas parmi d'autres, l'exemple du médecin est exemplaire des

problèmes posés par l'universalisation du paradigme de la performance au travail. Dans le cas de la santé, la contradiction entre les indicateurs de performance et la définition profane du travail médical est facilement observable : alors que les indicateurs s'intéressent à des ratios de patients vaccinés contre la grippe ou de prescriptions de médicaments génériques, les représentations profanes définissent le travail médical comme la prise en charge de la douleur, l'écoute du patient ou encore le fait de « sauver des vies ».

Le domaine de la santé est envahi par des spécificités qui se prêtent mal à une analyse par les outils de la théorie économique dominante. C'est pourquoi, ces spécificités sont souvent qualifiées d'« anomalies ». La fiction de l'*homo oeconomicus* est difficilement tenable tant les notions de déontologie, d'éthique ou d'altruisme imprègnent l'activité médicale et les représentations que s'en font les patients. Il n'est par ailleurs pas aisé d'octroyer au patient le statut de consommateur rationnel dans la mesure où la santé se rapproche plus du concept de besoin (involontaire) que de préférences (volontaires) : on choisit rarement d'être malade. À rebours des autres champs de l'économie, l'histoire des principaux systèmes de santé montre qu'il est impossible de concevoir ce secteur économique en dehors de l'intervention de l'État. C'est dans cette perspective générale que l'historien de l'économie Mark Blaug (1998, p. 66) pouvait considérer que :

> *« Health economics would seem to be a perfect topic for heterodox dissent […] health economics is field which must make the average neoclassical economist squirm because it challenges his or her standard assumptions at every turn ».*

En conséquence, l'étude de la relation médicale est le lieu privilégié pour comprendre les apories de la performance. C'est la perspective générale de cette thèse.

1 Problématique et thèse

L'introduction de la logique de performance dans la régulation de la médecine libérale s'inscrit dans l'histoire longue des relations du système de santé, entre socialisation et libéralisation. Cette histoire est marquée par une question assez simple : comment intégrer les médecins dans l'organisation du système de santé de façon à défendre les intérêts des patients et sans remettre en cause la pérennité du financement public ? La problématique prend tout son sens lorsque l'on se rappelle que la profession médicale s'est farouchement accrochée à ses libertés face à l'État et au marché : elle se défend contre la fonctionnarisation mais exige le monopole de la dispense des soins (Hassenteufel, 1997).

L'émergence de la logique de performance dans la relation médicale marque un changement majeur de stratégie de la puissance publique. Alors qu'historiquement les négociations entre l'État et les médecins portaient exclusivement sur des problématiques de prix, depuis le début des années 1990, la régulation publique se fait par les pratiques. Le régulateur concède la liberté tarifaire et la liberté d'installation en échange d'un contrôle de plus en plus poussé du travail médical – c'est-à-dire le recul de la liberté thérapeutique et de la liberté de prescription. L'objectif du contrôle est alors de promouvoir la qualité des soins – notamment sur les enjeux de santé publique, tels que la lutte contre le tabagisme – et la réduction des dépenses – en évitant le développement des maladies et en favorisant la prescription de médicaments génériques.

La Rémunération sur objectif de santé publique (ROSP), venu remplacer et élargir le Contrat d'amélioration des pratiques individuelles créé en 2009, est la dernière étape de cette évolution. Il n'est en conséquence qu'une manifestation d'un phénomène bien plus large mobilisant de nombreux instruments : démarches qualité, référentiels de qualité, normes de bonnes pratiques, références médicales opposables, accréditations, agence indépendante de production de normes (Haute autorité de santé), etc. Ce « tournant métrologique » (Bloy et Rigal, 2012) de la profession conduit à multiplier les dispositifs d'évaluation chiffrée de la pratique médicale et à mettre en indicateurs le travail du médecin. Notre thèse propose d'interroger la pertinence de la régulation de la médecine libérale par les pratiques : est-ce que le déploiement de la « performance totale » en médecine libérale permet d'améliorer l'intégration de la profession au système de santé dans l'intérêt de l'ensemble des patients ? Si cette évolution n'est pas ou peu profitable aux patients, à quoi sert-elle ? N'existe-t-il pas une autre logique dans l'injonction permanente à la performance ?

Notre thèse est que la régulation par les pratiques mises au point par la puissance publique, en niant l'existence de rapports de force propre au capitalisme et en acceptant une naturalisation de la définition de la « bonne » santé, se fait au détriment des intérêts des patients et des médecins de première ligne. L'injonction à la performance ne conduit pas à améliorer la qualité des soins et à renforcer la maîtrise des dépenses de santé, contrairement aux objectifs annoncés. Par contre, dans l'esprit du néolibéralisme contemporain, la santé est assimilée à un bien comme un autre autour duquel il est possible de mettre en concurrence les producteurs et les consommateurs. Notre perspective critique s'appuie sur deux piliers principaux :

- La logique de performance implique une définition univoque du produit médical – le soin – qui n'est pas possible. Le soin apporté par le médecin à son patient est un service singulier dont la définition se nourrit des représentations normatives que se font les acteurs sur la santé. Notamment parce que la santé est un besoin et non une préférence, les acteurs ont développé des représentations excluant le soin du domaine marchand. Dans cette configuration, l'éthique médicale est une ressource indispensable pour l'action mais elle est menacée par l'industrialisation de la production des soins en germe dans la politique de régulation des pratiques. En désingularisant les soins et en érodant les valeurs hippocratiques de la médecine, cette politique de santé participe au néolibéralisme sanitaire visant à une mise en concurrence généralisée.

- Bien que légitime *a priori*, la focalisation sur la performance des médecins occulte le rôle de la configuration présente du capitalisme dans la définition des problèmes de santé. La conséquence en est l'éviction d'enjeux majeurs de santé publique. La régulation de la médecine par la performance répond à un moment historique de désocialisation de l'assurance maladie et de développement des inégalités en santé. Le choix public en faveur d'une régulation des pratiques médicale occulte de nombreux problèmes : liberté tarifaire et d'installation, mode de rémunération du médecin, absence d'organisation de la médecine ambulatoire, absence de coordination entre l'ambulatoire et l'hôpital, pénurie organisée des recettes de l'assurance maladie, etc.

2 Perspective théorique : de la critique du *mainstream* à l'analyse institutionnaliste

Une cartographie du mainstream

Cette thèse est construite à partir d'une critique des différentes variantes de la théorie économique dominante – le *mainstream*. En effet, ce dernier est multiple, ce qui nous conduit à décliner la critique en autant de variations que de métamorphoses du *maintream*. Un point de départ important est donné par Favereau (1989) pour lequel il est possible de qualifier les théories à partir de deux piliers définis par Arrow (1974) : (i) La rationalité de l'agent économique s'exprime par un critère d'optimisation (maximisation sous contraintes) (ii) Le marché est l'unique mode de coordination permettant la compatibilité des décisions individuelles et l'allocation optimale des ressources.

Le point de vue des économistes vis-à-vis de ces deux piliers permet alors de positionner leurs travaux par rapport à la théorie de l'équilibre général – constituant le cœur de la Théorie Standard (TS) en économie (Graphique 1.0.1.). La Théorie Standard Etendue (TSE) regroupe les travaux qui, acceptant de relâcher l'hypothèse de coordination exclusivement marchande, conservent néanmoins celle de rationalité optimisatrice. Il s'agit, par exemple, de la théorie des incitations qui met au cœur de ses travaux les problèmes d'asymétrie d'informations. Les individus optimisent toujours mais, la coordination peut se faire par d'autres moyens que le marché (le contrat, par exemple). À l'inverse, la Théorie Standard évolutionniste (TSe) rejette le premier pilier pour lui préférer le second. Le marché est dans ce cas le seul mode de coordination existant pour des individus non optimisateurs. Enfin, la dernière possibilité est de rejeter en bloc les deux hypothèses. C'est ce que font les Théories non Standards (TnS).

Graphique 0.1. Table d'orientation (Favereau, 2010)

Modes de coordination interindividuelle : Pluralité

Rationalité individuelle : optimisation	TSE Théorie Standard Etendue	TnS Théorie non Standard	Rationalité individuelle : satisfaction
	TS Théorie Standard	*TSe* *Théorie Standard évolutionniste*	

Mode de coordination interindividuelle : unicité

Si cette table d'orientation est absolument fondamentale pour comprendre l'évolution de la théorie économique au cours du 20ème siècle, elle mérite d'être actualisée. Avec le recul historique et la prolifération des travaux de la TSE, il devient important de donner de nouveaux repères permettant de distinguer les variations au sein même de la TSE. Pour cela, il nous semble judicieux d'étudier au sein même du premier pilier – l'optimisation sous contrainte – ce qui fait l'objet de l'optimisation.

La situation de référence de la TSE est donnée par la formulation canonique de la théorie de l'agence (Jensen et Meckling, 1976, Ross, 1973). Nous proposons de réunir ces travaux sous l'appellation de Théorie Standard Etendue 1 (TSE1) – 1 au sens de première génération. Les individus sont susceptibles de se coordonner autrement que par le marché et ils optimisent une fonction objectif qui ne fait intervenir que des éléments traditionnels déjà présents dans la TS (désutilité de l'effort, coût d'opportunité du loisir, revenus, etc.). Ce qui est de « première génération », c'est l'hypothèse d'optimisation – qui se fait selon des éléments traditionnels.

Suite aux critiques faites à l'hypothèse d'optimisation traditionnelle, l'objet de l'optimisation a ensuite évolué dans deux directions principales :

- La rationalité limitée : les capacités de calcul de l'individu sont limitées par des problèmes d'ordre cognitifs. Un nombre croissant de modèles de la TSE ont pris en compte l'argument ancien selon lequel l'individu de la théorie du choix rationnel ne peut pas faire les calculs que lui prêtent les économistes. L'optimisation est désormais contrariée mais le principe d'optimisation n'est pas remis en cause. On trouve dans cette catégorie, par exemple, les modèles de *search* dans lesquels la recherche d'information est coûteuse (Stigler, 1961).

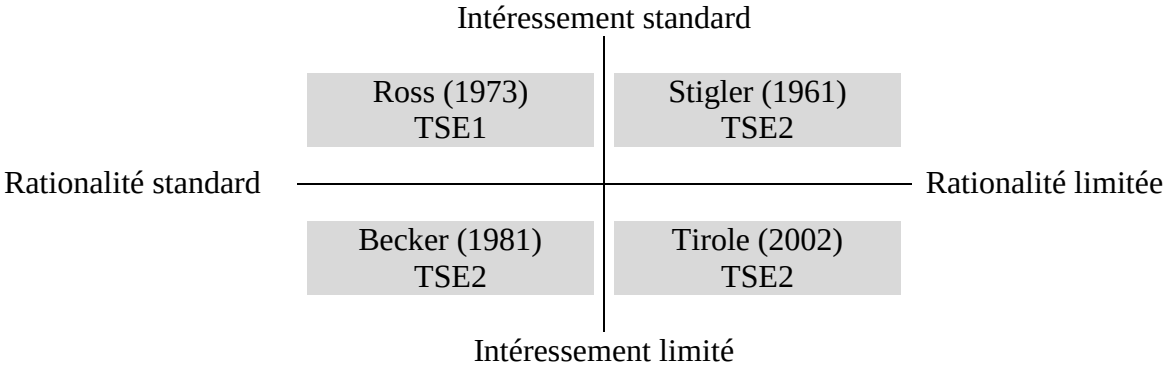
- L'intéressement limité : les individus ne sont pas uniquement intéressés à eux-mêmes. Il s'agit cette fois-ci de prendre en compte les critiques sur l'égoïsme exclusif de l'individu de la théorie du choix rationnel. De nouveaux modèles ont été imaginés dans lesquels les individus ont des « préférences sociales » et/ou des « contraintes sociales ». Les travaux de Becker sont exemplaires à cet égard (Becker, 1981).

Dans les deux cas, l'ambition de la TSE est de prendre en compte les critiques portées au pilier d'optimisation sous contrainte… tout en y restant fermement attachée. Nous parlerons de Théorie Standard Etendue 2 (TSE2) lorsque cette ambition est affichée – c'est à dire lorsque la modélisation essaie de prendre en compte, soit le problème de rationalité limitée, soit le problème d'intéressement limité, soit les deux. Cette évolution de la théorie économique *mainstream* ne s'attaque pas à son noyau dur : elle ne porte que sur des hypothèses secondaires (ce qui est optimisé) et non sur l'hypothèse principale (l'optimisation).

L'ambition plus générale de cette évolution est de produire de nouveaux modèles d'optimisation prenant en compte à la fois les contraintes cognitives et les problèmes liés au désintéressement. Il est dès lors possible de tracer une ligne d'horizon pour les travaux de la

TSE (Graphique 0.2.). Selon le programme de recherche exposé par Tirole (2002), en partant du modèle canonique de la théorie de l'agence (cadran Nord-Ouest), puis des remises en causes ponctuelles de l'hypothèse traditionnelle d'optimisation (cadrant Nord-Est et Sud-Ouest), il s'agirait désormais de réunifier la TSE autour des fondamentaux du comportement humain – en prenant en compte, avec « parcimonie », les remises en causes de l'homme économique (cadrant Sud-Est)[1].

Graphique 0.2. Table d'orientation des TSE



Une perspective institutionnaliste

La place faite dans cette thèse à la critique du *mainstream* ne doit pas laisser penser que notre perspective théorique se définit par opposition à lui, en « négatif ». Dans le même esprit qui animait Keynes au début des années 1930, notre démarche théorique s'interdit de contourner la « citadelle » (Keynes, 2002, Favereau, 2005). Contrairement à la plus part des « hérétiques » qui n'ont que « leur flair et leur instinct » pour justifier l'écart à l'orthodoxie, Keynes propose de trouver « la faille fatale » du « raisonnement orthodoxe » :

> *« […] j'ai été élevé dans la citadelle et je reconnais son pouvoir et sa puissance. […] Je ne serai donc pas satisfait tant que je n'aurai pas mis le doigt sur la faille de cette partie du raisonnement orthodoxe, conduisant aux conclusions qui, pour différentes raisons, me paraissent inacceptables » (Keynes, 2002, p. 219-220)*

Deux raisons majeures nous conduisent à s'inscrire dans cette stratégie de recherche. D'une part, il nous semble essentiel de maintenir et de renouveler continument le débat entre « hérétiques » et « orthodoxes ». C'est la condition nécessaire pour ne pas sombrer dans un autisme improductif. D'autre part, il nous semble que le règne de la performance a à voir avec le *mainstream* et sa propension à performer le réel. L'influence de l'orthodoxie (dont les raisonnements recèlent une « faille fatale ») sur le monde réel peut alors expliquer les effets

[1] Il va de soi que cette nouvelle cartographie n'est pas substituable à la précédente : elle lui est complémentaire.

pervers des politiques publiques. Même quand la pensée économique orthodoxe présente des failles importantes, elle continue d'alimenter les politiques. Voilà pourquoi nous ne pouvons pas nous satisfaire de l'ignorer.

En « positif », notre démarche relève de l'institutionnalisme. Si l'institutionnalisme est composé d'une pluralité de « familles » (Chavance, 2007), il est néanmoins possible de repérer un fondement commun, une unité paradigmatique, à ces différents courants (Postel et Sobel, 2009). Les institutions peuvent alors se définir comme l'ensemble des manifestations du collectif qui conditionnent, sans le déterminer, le comportement individuel des acteurs. Pour reprendre la formule de Postel et Sobel (2009, p. 48) plutôt que de rechercher des fondements micro à la macro, il s'agit pour nous de dévoiler les fondements macro de la micro.

Cette position épistémologique s'inscrit pleinement dans la tradition institutionnaliste en économie de la santé (Batifoulier, 2009, Domin, 2013, Gadreau, 2009, Davis, 2001, McMaster 2007, Hodgson, 2008). Au-delà du seul marché, d'autres institutions concourent à régler la coordination entre les acteurs : les règles de droit, les coutumes, les conventions, les normes, les principes éthiques, les représentations, etc.

Ici plus qu'ailleurs, la particularité du bien santé mets au premier plan le rôle des valeurs dans la coordination. Cette omniprésence des représentations normatives, nous conduit à mobiliser l'Économie des Conventions (Batifoulier, 2001, Eymard-Duvernay, 2006, Orléan, 2004) pour comprendre comment la politique publique influe sur l'action qu'il convient de mener quand on est médecin ou patient. Si les individus sont dotés d'une capacité de jugement que l'on peut qualifier d'éthique ou de normative (ils peuvent dire ce qui est juste au double sens de justesse et de justice), leur représentation de ce qu'il convient de faire est nourrie des évolutions de la politique publique. Le rôle des rapports de force est également central pour l'analyse des économies capitalistes. L'Ecole de la Régulation (Boyer, 1986, 2004, Boyer et Saillard, 1995) fournit des éléments précieux pour l'étude de ses rapports de force. Au final, dans une perspective d'économie politique, les institutions, comme la Sécurité sociale ou l'éthique médicale, sont le produit historiquement situé, donc plus ou moins durable, du processus de construction social du monde, entre accord et conflits.

3 Méthodologie

Durant le travail de thèse nous avons cherché la connaissance sur l'évolution du système de santé par ses « deux bouts ». Nous présentons ici les étapes et les obstacles qui ont façonné le devenir final de cette thèse – qui, elle-aussi, est le produit d'une construction socio-historique.

A la première extrémité de la chaine, nous avons conduit, avec le sociologue Nicolas Belorgey, plusieurs entretiens ouverts avec des médecins. Notons d'emblée que le nombre limité d'entretiens ne permet pas de les exploiter comme nous avions imaginé de le faire initialement. Nous avons réalisé quatre entretiens d'une durée comprise entre 45 mn et 1h15 avec deux médecins libéraux et deux médecins salariés en centre municipal de santé. Si cette expérience ne nous a pas conduits à des résultats robustes et présentables dans le cadre de cette thèse, elle nous semble néanmoins fondamentale dans la mesure où elle nous a permis d'effleurer la complexité de la rationalité médicale. Lors de ces entretiens, nous avons non seulement pu tester des hypothèses confectionnées en « laboratoire » mais aussi prendre connaissance d'hypothèses nouvelles qui n'apparaissent pas spontanément dans la littérature (théorique et empirique). La méthode de l'entretien permet d'accéder à une profondeur insoupçonnée de la réalité sociale qui n'est pas comparable avec la froideur de la donnée chiffrée.

En dépit de notre volonté de poursuivre ces entretiens, nous avons dû y mettre en terme pour deux raisons. La maîtrise de l'enquête sociologique exige un apprentissage que nous n'avions pas et qu'il nous était impossible d'acquérir en si peu de temps. Nous avons commencé les entretiens à la fin de l'année 2012 (au milieu de la thèse) au moment même où nous avons dû nous tourner vers la seconde extrémité de la chaîne de connaissance : l'assurance maladie, le « donneur d'ordre ».

L'équipe de recherche en économie de la santé du laboratoire EconomiX a été sollicitée par le président de la Caisse Primaire d'Assurance Maladie des Hauts-de-Seine (CPAM 92) pour une étude sur les médecins du département. L'un des objectifs de cette étude était d'analyser la réaction des médecins face aux sollicitations de la politique publique. Grace aux bases de données mises à disposition, l'idée était d'analyser statistiquement quels profils de médecin étaient signataire du Contrat d'Amélioration des Pratiques individuelles. Ce contrat,

signé individuellement sur la base du volontariat, est le dispositif de paiement à la performance précédent la généralisation de la ROSP en 2012.

Malheureusement, cette étude, à visée plus générale que les entretiens qualitatifs, n'a débouché sur aucun résultat empirique exploitable. Pourtant, encore une fois, elle a été une source d'information d'une richesse rare. En effet, en plus d'une première familiarisation avec les bases de données administratives et leur contenu, nous avons pu apprécier les contradictions internes à cette institution sur l'intérêt porté au paiement à la performance médicale. Si le président de la CPAM 92, représentant des syndicats de travailleurs et élu au titre du paritarisme, nous a sollicités pour faire cette étude, le directeur de la CPAM 92, détenteur du pouvoir de gestion, est resté 6 mois durant à mi-chemin entre désintérêt par manque de temps et exaspération face au pouvoir médical. Le chargé de recherche recruté par la CPAM sur proposition d'EconomiX a été le témoin privilégié de ces contradictions tant il a peiné à obtenir des autorisations d'accès à certaines données, du matériel performant ou même des informations sur les nomenclatures utilisées par les bases de données. Après presque 10 mois de travail, nous avons été contraints de prendre du recul.

Ce n'est qu'au début de l'année 2014 que nous avons enfin pu mener un travail empirique entièrement à son terme. Dans le cadre d'une convention de recherche entre la Chaire Santé de Science Po, EconomiX et le Service Médical d'Assurance Maladie d'Ile de France (SMAMIF), nous avons eu la possibilité de travailler en coopération volontaire et totale sur les résultats de la ROSP pour les médecins franciliens. L'intérêt partagé pour l'étude du dispositif de paiement à la performance médical, et l'expérience préalable à la CPAM 92, a permis de produire des résultats originaux et significatifs dans un délai pourtant très court. Durant l'intégralité de ce travail, nous n'avons pas eu d'accès direct aux bases de données. Les données médicales sur les assurés sont en effet confidentielles. Nous avons formulé les requêtes qui nous semblaient pertinentes en réunion, celles-ci étant systématiquement acceptées et traitées par les services de la SMAMIF. Les résultats sont présentés dans le chapitre 5 sous la forme de statistiques descriptives et d'analyse économétrique. Ils constituent à ce jour, la première évaluation statistique de grande ampleur du dispositif ROSP.

En dépit de son originalité, l'étude empirique ne prétend pas démontrer la thèse dans toute sa complexité. Notre travail empirique ne se concentre que sur un volet spécifique de la thèse : la prétention du ROSP à améliorer la performance des médecins. Dans ce cadre, l'approche exploratoire des données permet de mettre en évidence une variété importante de « réalités » qui s'écartent de ce que le *mainstream* prévoit. L'approche économétrique est

encore plus limitative : elle teste une hypothèse restrictive dans un ensemble ouvert d'hypothèses testables. Si ces deux approches novatrices nous apprennent que le ROSP n'est pas tel qu'on nous le présente, elles ne sont pas le socle de la thèse. Elles ne sont qu'un « constat statistique » qui doit être inséré dans une réflexion théorique plus large.

Notre travail cherche autre chose que ce que peut révéler l'exploitation des bases de données. C'est pourquoi nous accordons une force importante aux autres sources empiriques de la thèse. Parmi elles figurent évidemment les revues de littérature empiriques réalisées par nos soins. S'ils doivent faire l'objet d'une lecture critique, les travaux empiriques réalisés par d'autres ne comptent pas pour rien. Nous mobilisons également des données statistiques nationales fournies par diverses institutions compétentes (Direction de la recherche et des études, de l'évaluation et des statistique, Institut national des statistiques et des études économiques, Institut de recherche et de documentation en économie de la santé). Afin d'étudier la règlementation, nous nous servons des divers textes de lois et décrets d'application régulant le système de santé (Loi de financement de la Sécurité sociale, convention médicales, Contrat d'amélioration des pratiques individuelles, etc.). Enfin, nous mobilisons des données historiques indispensables à une perspective institutionnaliste.

4 Plan de la thèse

La thèse est rédigée telle qu'elle a été pensée et reflète en quelque sorte le parcours théorique de son auteur. Dès lors, le plan opère une montée en généralité à partir des conceptions *mainstream* de la performance jusqu'à l'analyse institutionnaliste de la réforme sanitaire, en passant par l'évaluation empirique des systèmes de paiement à la performance médicale.

Première partie : Inciter le médecin à la performance dans le paradigme mainstream

La première partie de la thèse s'intéresse à ce que pense la théorie *mainstream* de la régulation de la médecine libérale. Nous montrons que l'essentiel de son discours s'organise autour de l'idée selon laquelle la santé n'est pas un bien comme un autre. Si cette proposition est vraie, alors il n'est pas possible d'organiser la délivrance des soins sur le modèle du marché de concurrence pure et parfaite. La Théorie Standard (TS), qui prévoit la rémunération des facteurs de production à leur productivité marginale, doit alors être écartée. Pourtant,

l'évolution de la TS en Théorie Standard Etendue 1 va avoir pour effet de renforcer le pouvoir normatif du cadre marchand : toute défaillance de marché doit activer un dispositif incitatif permettant de se rapprocher de l'idéal de marché. La préconisation du paiement à la performance médicale pour contrer les problèmes liés aux asymétries d'information s'inscrit dans ce cadre théorique : le paiement à la performance est l'héritier direct de la norme concurrentielle. En promettant d'améliorer la qualité des soins et de réduire les dépenses de santé, cet instrument semble alors constituer un outil universel de politique publique.

Cependant, cette formalisation du comportement médical ne permet pas de souligner l'exemplarité supposée du médecin. Il reste un opportuniste comme un autre et le principe de paiement à la performance ne lui est en rien particulier. Face à cette conception du comportement médical, l'économie *mainstream* de la santé a cherché à intégrer des « préférences sociales » dans deux directions différentes. Le chapitre 2 s'intéresse à l'éthique médicale formalisée dans le langage de l'altruisme. Le chapitre 3 analyse un segment plus récent de la littérature portant sur les motivations médicales. Contrairement à l'éthique médicale, l'intérêt de la thématique des motivations porte sur l'hypothèse de substituabilité entre motivation intrinsèque (préférence sociale) et motivation extrinsèque (incitation financière). Inciter plus pourrait motiver moins.

Si cet effort de complexification, s'inscrivant dans la mutation de la TSE1 en TSE2, est appréciable, il manque l'essentiel de sa cible. En effet, nous montrons qu'en voulant rapporter les comportements anti-utilitaires dans la théorie du choix rationnel, la TSE2 produit des paradoxes de théorie économique qui débouchent sur des recommandations de politique publique contradictoires.

Deuxième partie : Les déboires empiriques de la performance médicale

En dépit des limites théoriques soulignées précédemment, les dispositifs de paiement à la performance se multiplient dans tous les systèmes de santé du monde. Partant de ce constat, la deuxième partie de cette thèse propose une évaluation du paiement à la performance médicale. À partir d'une revue critique de la littérature empirique, le chapitre 4 montre que cet instrument de politique publique est incapable de faire la preuve de son efficacité tant sur le plan de l'amélioration de la qualité des soins que sur celui de la réduction des dépenses. Dans de nombreux cas, le paiement à la performance est même responsable d'effet pervers remettant en cause la pertinence de son utilisation.

Bien que ces résultats invitent au scepticisme, la France a introduit ce nouveau mode de rémunération en 2012 dans le cadre de la Rémunération sur objectif de santé publique (ROSP). Le chapitre 5 propose une analyse empirique originale du ROSP. Deux idées principales traversent ce chapitre. D'une part, nous développons l'hypothèse d'un écart relativement grand – donc problématique – entre la notion abstraite de performance et la performance mesurée. Les médecins qui sont rémunérés par l'assurance maladie au titre de la performance ne sont pas forcément performants au regard des critères forgés par l'assurance maladie elle-même ! D'autre part, nous mettons en évidence que la performance mesurée n'est pas réductible à l'implication et aux qualités du médecin traitant : la performance dite individuelle a des ressorts collectifs. Dans cet esprit, notre étude économétrique montre que la performance du médecin – quelle que soit sa définition – est corrélée à l'état de santé et au niveau de revenu de sa patientèle.

Troisième partie : La politique de la performance dans le néolibéralisme sanitaire

La troisième partie prend au sérieux les limites du cadre d'analyse *mainstream*. Nous proposons un retournement de perspective épistémologique : plutôt que rechercher les fondements microéconomiques de la macroéconomie, nous recherchons les fondements macroéconomiques de la microéconomie. Les comportements des médecins trouvent leurs racines dans l'évolution de la politique publique qui forge un nouveau cadre normatif. Ainsi, nous privilégions l'analyse des institutions du système de santé pour comprendre les réactions des médecins face à l'introduction du paiement à la performance.

Le chapitre 6 propose une analyse des rapports entre la profession médicale et la puissance publique dans une perspective historique. Prenant en compte l'évolution du capitalisme, nous montrons que le paiement à la performance médicale n'est qu'un outil d'une évolution plus générale de la régulation de la médecine libérale. Alors qu'historiquement la puissance publique se bornait à négocier avec les médecins le prix des soins dans le but d'en généraliser l'accès, depuis le début des années 1990 celle-ci cherche plus volontiers à réguler les pratiques médicales – laissant aux médecins leur liberté tarifaire. À une époque où il n'est plus légitime de vouloir augmenter les dépenses publiques de santé, le régulateur veut contrôler les pratiques médicales, à partir d'indicateurs de production chiffrés, afin d'améliorer la qualité des soins et de maîtriser les dépenses. Le ROSP est un outil mais pas le seul de cette évolution.

Le chapitre 7 poursuit l'analyse de la pertinence de la nouvelle stratégie de la puissance publique en ciblant la façon dont se construit la nouvelle rationalité médicale à partir du processus de standardisation des soins. La mise en protocole du travail médical fait émerger une question fondamentale : la qualité des soins est-elle soluble dans la quantification ? Nous montrons d'abord que cette idée n'est en rien naturelle. Elle est le fruit d'une double évolution : c'est parce que la médecine fondée sur les preuves (*Evidence based medicine*) a imposé l'usage de normes chiffrées dans le champ médical, que la nouvelle gestion publique (*New public management*) a pu se développer en santé. La suite du chapitre propose une critique de l'objectif de quantification de la qualité de soins avec deux arguments principaux. D'une part, il n'est pas pertinent de naturaliser les résultats de la médecine fondée sur les preuves, tant du fait de conflits épistémologiques sur la définition de la maladie que du fait de l'encastrement sociohistorique de production de ses résultats. D'autre part, comme l'explique l'économie hétérodoxe des services, il est fortement problématique de considérer l'activité de soin par l'intermédiaire du paradigme du produit.

Le chapitre 8 propose de relire l'évolution de la régulation de la médecine libérale – des prix aux pratiques – à l'aune de l'Économie des Conventions. Nous montrons que cette évolution peut s'interpréter comme le passage d'une convention de qualité des soins que l'on peut appelée « inspirée / domestique » – fondée sur la singularité de la relation médicale et le rôle du jugement de l'expert en situation – à une convention de qualité des soins « industrielle » – fondée sur le respect de standards universels. Ce changement de convention de qualité des soins implique une évolution de l'éthique médicale. Alors que l'éthique médicale était fondée sur les valeurs déontologiques et la place du médecin comme dépositaire de la santé, il est désormais attendu du médecin une compétence technique et l'application de standards définis au niveau des agences de santé. Enfin, à un niveau plus général, le changement de convention de qualité des soins s'inscrit dans une modification plus large de l'éthique en santé. La convention de qualité industrielle, en homogénéisant le soin, est la condition nécessaire pour diffuser la norme concurrentielle caractéristique du néolibéralisme. Le soin devient peu à peu un bien comme un autre et le patient a maintenant les moyens de faire jouer la concurrence entre des producteurs homogènes. En modifiant la convention de qualité des soins, la puissance publique atteint la conception que le patient se fait de la santé.

<u>Première partie :</u>

Inciter les individus à la performance dans le paradigme *mainstream*

L'introuvable originalité du travail médical

Introduction de la première partie

La première partie de cette thèse constitue une critique de la théorie économique dominante du comportement médical. Malgré la reconnaissance d'une originalité du secteur des soins, on ne retrouve pas dans les résultats du *mainstream* de véritables traces d'une originalité de la régulation du médecin. Postuler que celui-ci doit être rémunéré au moins en partie à la performance est un résultat directement hérité du principe de rémunération à la productivité marginale des facteurs de production.

Le premier chapitre pose le problème liminaire de l'économie de la santé *mainstream*. Puisqu'il existe une originalité du bien santé reconnue en économie de la santé, on ne peut pas laisser faire le marché. Il n'est ni possible ni souhaitable de rémunérer le médecin à sa productivité marginale. Or, plutôt que de marginaliser le paradigme de marché tel qu'il est analysé par la Théorie Standard (TS), ici comme dans les autres domaines de la théorie économique, celui-ci va devenir la norme vers laquelle doit s'approcher le réel. Dans le cadre de cette « stratégie du Nirvana » définie par Demsetz (1969), puisque le réel est enlaidi par les défaillances de marché, il suffit de trouver des outils incitatifs suffisamment puissants pour réduire l'écart entre le monde tel qu'il est et l'idéal de marché de concurrence pure et parfaite.

S'étant développée sur cette base, la Théorie Standard Etendue 1 (TSE1) a fait des asymétries d'information le problème principal de la relation tripartite patient-médecin-assureur. Opportuniste et calculateur, le médecin dispose d'une rente informationnelle qu'il faut abattre dans le but d'améliorer l'allocation des ressources rares et la qualité des soins. Puisqu'il est impossible d'imposer une rémunération à la productivité marginale au médecin, le paiement à la performance est un moindre mal. Dans la mesure où cette forme de rémunération est liée arithmétiquement à un indicateur de productivité, il semble que ce dispositif soit le meilleur substitut à la « discipline de marché ». Le paiement à la performance, résultat du passage de la TS à la TSE1, est l'héritier de la norme de rémunération à la productivité marginale.

Le deuxième chapitre s'intéresse à l'abondante littérature portant sur la thématique de l'éthique médicale dont l'existence permet de revenir sur la particularité de la rationalité médicale. Prendre au sérieux cette éthique fragilise l'hypothèse d'un individu intéressé par l'intéressement – que lui offre, par exemple, le paiement à la performance. En effet, le monde médical est inondé de références à l'éthique ou à la déontologie interdisant au médecin de se

conduire comme un producteur classique et de le formaliser comme un simple maximisateur de revenu. De nombreux auteurs ont donc essayé d'intégrer ces comportements en amendant l'hypothèse d'optimisation traditionnelle dans le cadre de ce que nous avons appelé la Théorie Standard Etendue 2 (TSE2). Or, nous défendons l'idée que cette littérature est problématique à deux niveaux. À un premier niveau, en réduisant les considérations éthiques à la grandeur homogène d'utilité, c'est-à-dire en mettant sur le même plan l'intéressement et le désintéressement, ces travaux posent des paradoxes de théorie économique. À un second niveau, les paradoxes de théorie économique débouchent sur des problèmes de politique publique car ils produisent des recommandations contradictoires et/ou absurdes. En fin de chapitre, prenant au sérieux la particularité du bien santé, nous proposons une définition hétérodoxe de l'éthique médicale fondée sur la notion de besoin.

Tout en restant dans la TSE2, le troisième chapitre change de perspective en s'appuyant sur la notion de motivation issue de psychologie sociale. La motivation intrinsèque est une motivation non instrumentale dont l'existence remet en cause l'idée selon laquelle les incitations financières (relevant de la motivation extrinsèque) ne peuvent pas avoir d'impact négatif sur le comportement. Les motivations intrinsèques et les motivations extrinsèques peuvent être substituables : c'est l'hypothèse de *crowding out effect*. Augmenter le prix du travail n'a alors pas forcément pour effet d'augmenter l'offre de travail. L'originalité du médecin reposant dans l'existence de motivations intrinsèques, augmenter sa rémunération par un mécanisme de paiement à la performance peut avoir pour effet paradoxal de réduire son effort. La thématique des motivations est particulièrement avenante pour la théorie *mainstream*. Alors que l'éthique implique un jugement moral à cause de la présence du collectif, la motivation est un concept qui n'a pas besoin du social. Elle est alors plus souple et plus commode lorsqu'il s'agit de l'introduire dans le cadre de la TSE2 : si le modélisateur fait référence à des « préférences sociales », il n'est pas besoin de les justifier dans la mesure où la motivation est individuelle. Cependant, nous montrons que ce renouvellement théorique est sujet au même type de critique que dans le cas de l'éthique médicale. Inclure les motivations intrinsèques et l'hypothèse de *crowding out effect* dans la fonction d'utilité, pose des paradoxes de théorie économique et des problèmes de politique publique.

Chapitre 1

Le paiement à la performance médicale : héritage de la norme concurrentielle

Introduction

L'intuition générale de ce premier chapitre est que l'éventuelle particularité du secteur des soins médicaux au regard des autres domaines de la théorie explique l'éventuelle particularité de la régulation et de la rémunération du médecin au regard des autres secteurs d'une économie. L'idée défendue ici est qu'il existe une continuité structurelle entre la théorie standard (TS) et la théorie standard étendue (TSE) en ce qui concerne les recommandations de politique de rémunération du médecin. Malgré l'originalité des premiers travaux en économie de la santé, toutes les recherches contemporaines de la TSE peuvent se résumer à la recherche de réduction de l'écart entre le monde réel et la situation de marché de concurrence pure et parfaite – au fondement de la TS. Chaque proposition théorique est alors une version plus ou moins sophistiquée visant à rapprocher la rémunération du médecin à sa productivité marginale : les économistes ne résolvent que les problèmes qu'ils se posent.

Le paiement à la performance (en Anglais, pay for performance ; P4P) médicale est en conséquence le résultat logique de cette « stratégie du Nirvana ». La stratégie du Nirvana est, d'après Demsetz (1969), l'attitude théorique consistant à préférer comparer le réel à la norme idéale et irréalisable de marché de concurrence pure et parfaite plutôt que de comparer des arrangements institutionnels réels – ou réalisables – entre eux. De façon plus générale, en présentant le passage de la TS à la TSE1 par le concept de stratégie du Nirvana nous entendons dans ce premier chapitre2 décrire la démarche générale de la théorie *mainstream*. À chacune de ses mutations, certaines hypothèses auxiliaires sont relâchées afin de conserver le cœur paradigmatique : la norme marchande. Ici, c'est l'hypothèse d'unicité des modes de coordination qui est relâchée du fait de la mise en évidence de puissantes défaillances de marché dans le domaine des soins médicaux.

[2] L'étude de la TSE2 est réservée aux chapitres 2 et 3.

Le passage de la TS à la TSE1 est ainsi l'occasion d'observer la façon dont le *mainstream* construit ses problématiques. L'essentiel des questionnements dans le cadre de la TSE portent sur les asymétries d'information entre les parties prenantes. Puisque le médecin possède d'avantage d'information que le patient et l'assureur, il est rationnel pour lui d'extraire une rente informationnelle. De façon certaine, le médecin va alors exercer un opportunisme néfaste pour le bien-être collectif et il se pose le problème classique en économie de la qualité : en situation d'incertitude, il n'y a pas de marché, donc pas de coordination. Comment s'assurer de la qualité des soins en environnement incertain ? Et plus largement, comment expliquer que médecin et patient entretiennent des relations ?

La réponse est en fait très simple : le médecin est détenteur d'un savoir spécialisé. Une justification de l'asymétrie d'information est donc purement économique : l'investissement dans ce savoir complexe et évolutif est coûteux. Si le patient était aussi bien informé que le médecin, il n'irait jamais consulter. En fait, si l'on ne s'intéresse qu'à la partie diagnostic, le marché des soins est essentiellement un marché de l'information. Il s'agit pour le médecin, à partir de ses connaissances, de traduire des symptômes en diagnostic. Ce qui est échangé étant des informations – par hypothèse inégalement réparties –, pourquoi vouloir déduire toutes les recommandations de politiques publiques du « problème » d'asymétrie d'information ?

L'existence du médecin est la solution du problème d'asymétrie d'information. Plus généralement, dans le cas où ce qui est échangé est une information, s'intéresser à l'écart entre les observations et la situation de marché à information parfaite et complète devient absurde. Cela revient en effet à prendre pour étalon (Nirvana) une situation où le marché… disparaitrait. Ici, l'asymétrie d'information ne détruit pas les conditions de l'échange, elle les créé. Sans asymétrie d'information, pas besoin de médecin. La question pour l'économiste n'est alors pas tant de discuter des asymétries d'information que de discuter de la prise de décision en situation d'incertitude radicale3.

Nous voyons en conséquence que l'asymétrie d'information est un problème uniquement parce qu'elle n'existe pas dans la situation de concurrence pure et parfaite. Il ne s'agit d'un problème que parce que le *mainstream* prend pour norme la TS. Voilà la démarche générale du *mainstream*. En économie, on ne traite que les problèmes que se donnent les économistes : ceux de la stratégie du Nirvana. Dès lors, les problèmes que se donnent les économistes étant fondés sur une forte part d'arbitraire, les solutions à ces problèmes sont

[3] La troisième partie aborde cette question en profondeur.

également arbitraires. Le P4P et autres incitatifs n'ont que très peu de chose à voir avec une analyse du monde réel. En conséquence, l'originalité proclamée du secteur des soins médicaux est submergée par la puissance normative du marché de concurrence pure et parfaite. Le médecin n'a pas d'épaisseur propre et doit être rémunéré comme les autres… à la performance.

La première section explicite la démarche générale de la théorie économique dans le cas de la régulation du médecin. On montre en quoi l'économie de la santé est un exemple exemplaire des raisons pour lesquelles le cadre de la TS doit être dépassé (1). En dépit des appels à l'originalité du secteur des soins médicaux, la seconde section montre comment la TSE1 construit ses problèmes et comment elle aboutit à une recommandation de politique publique assez standard : l'introduction de P4P dans le cadre d'une rémunération mixte – entre partage du risque et incitation à l'effort (2).

1. En dépit de tout : la norme concurrentielle comme Nirvana

La théorie économique a une histoire. Nous pensons que pour comprendre les débats contemporains sur la régulation de la médecine ambulatoire, il est utile de se remémorer les premières discussions des économistes de la santé sur leur objet. En quoi les questions d'aujourd'hui sont-elles un renouvellement de celles d'hier ? Lorsque l'on s'intéresse aux textes fondateurs d'économie de la santé, il est frappant d'observer la constance de l'idée selon laquelle le soin n'est pas un bien comme un autre. Or, si le soin n'est pas un bien comme un autre, c'est donc que le secteur des soins médicaux n'est pas un marché comme les autres. Ce substrat théorique est fondamental pour ce qui nous occupe ici. En effet, répondre à la problématique du mode de régulation de la médecine ambulatoire passe nécessairement par la question du mode de coordination dans le secteur des soins médicaux. Le secteur de la santé est-il ou non organisé comme un marché ? Le soin est-il ou non un bien comme un autre ? Notre conviction est que la particularité ou non du bien a un impact sur la particularité ou non du mode de coordination dans le secteur et donc sur la particularité ou non de la régulation des offreurs de soin.

Dans cette section, nous commençons par étudier ce que serait le marché des soins médicaux à partir de la TS (1.1.). En conformité avec les résultats habituels, les facteurs de production doivent être rémunérés à leur productivité marginale. Le médecin doit alors être

payé à sa productivité marginale du travail, sans toutefois qu'il soit nécessaire de mobiliser les notions de contrôle, de sanction ou encore de discipline de marché. Ensuite nous rappelons les débats sur la particularité du secteur des soins médicaux (1.2.). Contrairement à ce que pourrait laisser croire une lecture des textes contemporains en économie de la santé, la question de savoir si l'activité de soin forme un marché n'est pas anodine. S'il n'existe pas de consensus chez les auteurs fondateurs sur la cause ultime de la particularité du secteur (conditions sur l'offre, conditions sur la demande, etc.), il semble émerger l'idée que le soin n'est pas un bien comme un autre – ce qui laisse entendre que le marché de la santé n'est pas un marché comme un autre. La section suivante propose une analyse plus systématique des défaillances de marché (1.3). Enfin, dans la dernière section, nous montrons que, malgré les défaillances de marché, le programme de l'économie de la santé a été le même que le programme de recherche de la théorie économique : la stratégie du Nirvana (1.4.). Cette approche du « Nirvana », selon l'expression de Demsetz (1969), consiste à nier une portée explicative au réel pour privilégier le monde idéal. Le monde réel est jugé à l'aune de l'écart avec des mondes idéaux construits par le théoricien. Plutôt que de prendre les conclusions qui s'imposent lorsque les défaillances de marché s'accumulent – substituer le programme de recherche néoclassique par un autre –, les économistes ont préféré modifier à la marge leur paradigme : ce sont les soubassements de la théorie standard étendue. En dépit des intentions liminaires, toute l'originalité du secteur des soins s'évapore.

1.1. Le marché des soins médicaux : l'ajustement marchand

Dans l'optique de la TS, la démarche princeps d'un chercheur s'intéressant à un secteur particulier de l'économie (ou l'économie dans son ensemble) consiste à utiliser le concept de marché de concurrence pure et parfaite. Notons d'emblée qu'à notre connaissance cette démarche n'a pas été entreprise de manière systématique pour le cas des services médicaux. Il peut alors paraître étrange au lecteur spécialiste d'économie de la santé de commencer par ce cas de figure. L'étrangeté du propos disparait si l'on se rappel du but poursuivit : montrer en quoi, bien que les économistes *mainstream* s'accordent pour dire qu'il n'y a pas et qu'il ne peut pas y avoir de marché des services médicaux, ils utilisent la norme marchande pour évaluer et modifier le monde réel. Pour s'assurer de cela, il faut rappeler, même brièvement, ce que serait un marché des services médicaux. C'est ce que nous faisons dans cette section.

En microéconomie standard, le consommateur réalise un arbitrage entre plusieurs biens étant données des préférences préétablies, non questionnées et non questionnables – les

individus sont supposés « libres ». Le consommateur est présenté comme un sachant : il sait ce qu'il veut. Il n'y a pas de problème d'information. L'utilité marginale d'un bien est décroissante et la préférence du consommateur pour un objet n'est pas inconditionnelle. Il est toujours prêt à substituer un peu plus d'un bien contre un peu moins d'un autre. Le consommateur étant sensible aux prix des biens, lorsque ceux-ci varient, sa demande varie en sens inverse. Au niveau agrégé, la demande pour les services médicaux est décroissante du prix. L'ajustement dépendra alors de l'offre agrégée adressée au consommateur.

1.1.1. Représenter le médecin et l'activité médicale

La question de l'offre de soin, celle qui nous intéresse plus particulièrement, est plus délicate. Trois possibilités nous sont offertes pour la traiter :

- Le médecin offreur de travail : Si l'on considère le médecin comme un offreur de travail, ce qui est le plus souvent le cas dans la littérature *mainstream*, il est soumis à un arbitrage travail-loisir classique (Sloan, 1975, Noether, 1986). Dans ce cadre, le médecin, selon ses préférences, arbitre entre une rémunération diminuée de la désutilité de l'effort et le loisir – supposé utile en soi, mais représentant un coût d'opportunité. La limite intrinsèque de ce choix de modélisation provient de l'absence de production. Ici, le médecin vend sa force de travail au consommateur qui est censé s'en servir pour se soigner. Or, dans le modèle néoclassique le travail n'est pas utile en soi. Il est nécessaire que le travail soit transformé en produit. Si l'on en reste là – et c'est ce que fait la littérature – on ne sait pas comment le travail du médecin se transforme en un produit homogène : le soin. Théoriquement, c'est à l'entreprise que revient ce rôle par l'intermédiaire de la fonction de production et le recours au capital. Cette dernière assure la modification du travail en un bien consommable. Voilà pourquoi, à la vision du médecin comme un offreur de travail, une approche rigoureuse de la théorie standard devrait trouver une place pour l'entreprise.

- Le médecin dans l'entreprise : Dans ce cas, le médecin est toujours soumis à un arbitrage travail loisir mais, cette fois-ci, ce n'est pas lui qui produit les services de soins : c'est l'entreprise – l'hôpital (Newhouse, 1970, Pauly et Redisch, 1973). Du point de vue du médecin rien ne change et la modélisation est plus proche de l'économie du travail standard. Le changement majeur porte sur la formation du prix : désormais, le capital entre en ligne de compte et l'ajustement offre-demande n'est plus un ajustement médecin-patient mais hôpital-patient.

- Le médecin entrepreneur individuel : Une troisième possibilité pour décrire ce qu'il se passe sur le marché des soins médicaux est de considérer que le médecin est aussi l'entreprise. Dans ce cadre deux nouveaux problèmes se posent. D'abord, le revenu du médecin est composé non seulement de son travail mais aussi du taux d'intérêt rémunérant le capital et du profit issu du programme de maximisation de l'entreprise. Ensuite, le questionnement se déplace sur la fixation du niveau de travail offert à cette entreprise. Dans la TS, les parties sont preneuses de prix parce qu'elles sont atomisées sur le marché. Or, lorsque le médecin est un entrepreneur individuel, celui-ci assure lui-même deux fonctions : il achète et vend son propre travail. Il n'y a donc pas de marché du travail-médecin ce pourquoi il n'y a aucune raison que le niveau de production individuel (chaque médecin) et agrégé (ensemble du secteur) des services médicaux soit à la hauteur de ce qui aurait été le cas s'il existait effectivement un marché du travail-médecin. Il faudrait théoriquement un marché du travail-médecin sur lequel les entrepreneurs-médecins viennent acheter le travail qu'ils vont utiliser pour produire les services médicaux. Cela rend possible le fait pour un médecin de produire des services médicaux avec le travail d'un autre médecin et, en même temps, de travailler pour un autre médecin-entrepreneur. Cette schizophrénie médicale ne paraît pas convaincante.

Au total, le modèle du médecin offreur de travail ne permet pas d'expliquer la transformation du travail en produit et celui du médecin entrepreneur individuel débouche sur une schizophrénie médicale insoluble. Le médecin dans l'entreprise est le cas le plus proche de l'idéal type de la TS (Tableau 1.1.). Si ce cadre ne pose pas de problème lorsque l'on s'intéresse à l'hôpital, dans le cas de la médecine libérale il faut imaginer que le médecin est le salarié de l'assureur[4] – et, en conséquence, on retombe dans le cas du médecin dans l'entreprise. Il s'agit d'une approximation douteuse mais, elle n'est pas totalement dénuée de fondements et elle est la seule permettant d'approcher les thèmes favoris de la TS.

[4] L'assureur n'étant pas, comme l'assurance maladie en France, en situation de monopole.

Tableau 1.1. La théorie standard et les problèmes du comportement médical

	Modélisation possible		
Problème de représentation de…	Médecin offreur de travail	Médecin dans l'entreprise	Médecin entrepreneur individuel
… la médecine libérale	Transformation du travail en produit	Le médecin n'est pas libéral, il est salarié de l'assureur	Schizophrénie médicale
… l'hôpital	Il n'y a pas d'entreprise	Situation classique (pas de problème)	Le médecin n'est pas à lui seul un hôpital

1.1.2. Régulation médicale et ajustement marchand

Si l'on retient le modèle du médecin dans l'entreprise, la problématique de la régulation de l'activité médicale reste classique. Par quels moyens l'entreprise pourra-t-elle contraindre le médecin à produire ce pourquoi il est rémunéré ? Quelle sera la quantité de services médicaux offerts ? A quel prix ? La réponse à ces questions, et à bien d'autres, passe par l'ajustement marchand et son principe concurrentiel. Dans cette perspective, les seuls problèmes de régulation sont des problèmes de marché, donc des problèmes de prix – le prix du travail étant la rémunération. Pour ce qui nous concerne, deux marchés sont en cause : le marché du travail des médecins et le marché des services médicaux.

- Le marché du travail des médecins : Le premier problème est celui de la fixation de la quantité et du prix du travail des médecins. La théorie néoclassique prévoit que la rémunération des facteurs de production, en situation de concurrence pure et parfaite, soit égale à leur productivité marginale. Ainsi, le travail du médecin, à l'équilibre, est rémunéré à sa productivité marginale (PmL),

$$R_f = Pm^L$$

Cette rémunération est optimale au sens où elle permet de résoudre le programme de maximisation du profit de l'entreprise. Mais, rien ne garantit que le médecin accepte n'importe quel niveau de rémunération. C'est à l'issue de son arbitrage travail-loisir qu'il saura à quels niveaux de rémunération il est prêt à travailler. Par la suite, à partir des fonctions d'offre et de demande agrégées de travail médical, on peut déduire le prix d'équilibre du travail médical.

Mais alors, comment l'entreprise peut-elle s'assurer que le médecin embauché produira l'effort pour lequel il est payé ? De façon assez surprenante, cette question n'a pas de sens dans le cadre de la TS. Sur le marché de concurrence pure et parfaite, les biens sont

homogènes, d'où l'exclusion par construction de la problématique de l'intensité de l'effort. Le marché de la TS est un mécanisme d'allocation des ressources qui ne connait pas les problèmes de qualité des biens. Seules les quantités et les prix sont en jeu. Autrement dit, les offreurs de travail n'ont pas prise sur le niveau de leur productivité marginale – ils ne peuvent pas « tirer au flanc ». Le seul problème pertinent concernant l'offreur de travail est la décision binaire de participation ou non au marché. Les biens étant homogènes, il ne peut pas y avoir de tromperie sur ce qui est échangé.

La notion d'effort n'est pas pour autant absente. En effet, lors de son arbitrage travail-loisir, le médecin offreur de travail subit une désutilité de l'effort. Cette hypothèse se justifie par l'idée de souffrance au travail. Celui-ci étant pénible par nature, seule une compensation peut convaincre les individus de travailler. Le problème sur le marché du travail est alors un problème de participation : étant donnés les prix et les préférences individuelles (dont la désutilité de l'effort et le goût pour le loisir), le médecin est-il prêt à participer aux échanges ?

- Le marché des services médicaux : Sur le marché des services médicaux, l'offre est croissante du prix et la demande décroissante du prix de sorte que l'équilibre sur le marché se trouve au point d'intersection des deux courbes. Lorsque le prix est supérieur au prix d'équilibre, les entreprises sont contraintes de réduire leur prix. Certaines d'entre elles peuvent même être amenées à quitter le marché car trop peu rentables. Si le prix devient inférieur au prix d'équilibre, alors les entreprises sont en position d'augmenter leur prix. Ce nouvel espace de profit est même de nature à inciter de nouvelles entreprises à entrer sur le marché. La concurrence est la règle d'or de l'ajustement marchand.

Ce mode de régulation a un impact sur l'offre de travail dans la mesure où plus le prix de marché est élevé, plus la demande de travail sera élevée. Dans cas de figure, l'élévation de la demande de travail produit une hausse de son prix, aboutissant, en fin de compte, à ce que de nouveaux offreurs de travail participent aux échanges. Les prix (du bien santé et du travail médical) ne résolvent donc pas tant un problème de discipline – d'incitation à l'intensité de l'effort – qu'un problème de participation – de compensation de l'effort. C'est la force de la TS que de résoudre le problème de discipline des salariés… en ne le posant pas.

1.1.3. Les propriétés d'optimalité de l'allocation marchande

À quoi bon un marché de concurrence pure et parfaite dans le cas des services médicaux ? D'où vient l'intérêt des économistes pour l'étude de ce type de marché ? D'après

la TS, la situation de concurrence pure et parfaite est une situation désirable en soi car elle permet de réaliser un idéal de justice : l'optimum de Pareto (pour des ressources et des conditions techniques données, on ne peut pas améliorer la situation de l'un des agents sans détériorer celle de l'un au moins des autres agents). Ce sont les deux théorèmes de l'économie du bien-être qui assurent que le libre jeu du marché concurrentiel aboutit à cet état social optimal : (1) Un équilibre de concurrence pure et parfaite est un optimum de Pareto. En d'autres termes, tout équilibre de marché correspond à une allocation des ressources optimale au sens de Pareto. (2) Toute allocation de Pareto peut être atteinte par un marché concurrentiel. En d'autres termes, tout optimum de Pareto est un équilibre.

Selon les préceptes de l'économie du bien-être, si le secteur des soins médicaux est un marché de concurrence pure et parfaite, alors aucune intervention n'est nécessaire pour assurer l'optimum parétien. Spontanément, l'ajustement marchand conduira à fixer la rémunération du médecin à sa productivité marginale. C'est en cela que ce niveau de rémunération peut être doté d'au moins deux vertus : (i) puisqu'il compense parfaitement l'effort du médecin, il récompense son mérite individuel et (ii) puisqu'il est associé à l'optimum de Pareto, il réalise un idéal de justice. Dans l'optique de la TS, la rémunération du travail à sa productivité marginale est juste à deux titres : elle est précise et elle est bonne.

Si le secteur des soins médicaux répond aux critères de l'ajustement marchand, alors il faut y laisser libre cours. Cependant, très tôt, les économistes vont insister sur l'originalité du secteur des soins médicaux : ce n'est pas un marché de concurrence pure et parfaite.

1.2. L'originalité de la santé chez les économistes pionniers

Malgré les divergences, lorsque l'on s'intéresse aux premières publications sur le sujet, il subsiste un fait saillant majeur en économie de la santé : le secteur des soins médicaux ne serait pas un marché comme les autres. Il est évident que la théorie de l'équilibre général n'a pas été élaborée dans le but de décrire le fonctionnement d'un secteur particulier de l'économie. Cependant, c'est bien la référence à cet idéal type du marché qui caractérise la démarche des économistes standards. Or, dans le cas de la santé, il est vite apparu un gouffre entre la théorie et l'observation la plus profane.

Pour s'en rendre compte, nous pouvons nous limiter à l'étude de trois textes majeurs du champ, encore aujourd'hui abondamment cités. Nous avons choisis ces trois textes parce qu'ils nous semblent représentatifs du débat occupant les économistes de la santé jusqu'à la

fin des années 1970. En se répondant les uns les autres ils donnent une bonne image des premiers pas de l'économie de la santé. Si avec le premier texte – celui de Kessel – l'écart au marché est lié à des contingences loin d'être insurmontables, avec les deux autres – ceux d'Arrow et de Pauly – la particularité du bien interdit de concevoir le secteur à partir du cadre marchand.

1.2.1. Reuben Kessel (1958), le marché entravé par les syndicats de médecins

L'article de Reuben Kessel (1958), « Price discrimination in medicine », paru dans le numéro inaugural du *Journal of Law and Economics*, pose d'emblée l'idée que l'offre de soin dans le secteur de la santé ne se conforme pas à l'organisation théorique d'un marché. Le pouvoir de marché des médecins, en situation de monopole, leur permet en effet de discriminer la demande de soins médicaux par les prix : les patients ne payent pas le même montant pour les mêmes soins médicaux. L'objectif de l'auteur est d'expliquer ce comportement qu'il considère comme anormal[5] en excluant ce que l'on peut appeler l'hypothèse éthique qui consisterait à dire que le médecin est animé de considérations morales le conduisant à ne pas faire payer les plus pauvres, par exemple.

Kessel ne donne aucun crédit à l'hypothèse éthique car, selon lui, les lois du marché sont correctes et inébranlables. Tout son travail est alors destiné à montrer que le comportement du médecin est effectivement un comportement de maximisation du profit qui transite par la discrimination par les prix et qui n'a pas besoin d'en appeler à une hypothèse éthique. Le comportement des médecins n'est pas celui d'un planificateur central assurant des transferts de revenus mais s'explique par le modèle du monopole discriminant[6].

Selon Kessel, le syndicat de médecin *American Medical Association* (AMA) constitue le cœur de l'explication de la discrimination par les prix[7]. C'est l'AMA qui, par le contrôle de l'enseignement délivré aux étudiants, peut dicter la politique de prix des médecins. Le comportement individuel des médecins serait contraint par l'organisation syndicale dominante de la profession. Les menaces de représailles du syndicat interdisent à un médecin isolé

[5] Anormal au sens de déviance par rapport à la norme : le marché de concurrence pure et parfaite.

[6] Dans un monopole discriminant, la discrimination par les prix signifie que l'offreur peut échelonner ses prix, pour un même bien et une même quantité, à partir du revenu des demandeurs. Il est ainsi en mesure d'extraire un surplus supérieur à celui de concurrence pure et parfaite en pratiquant des prix élevés sur les demandeurs à revenus élevés et en pratiquant des prix bas sur les demandeurs à revenus bas.

[7] Le travail de Kessel correspond à une application des théories du syndicat monopole au cas du monopole syndical de l'AMA. On pourra voir Cahuc (1990) pour une revue de littérature sur le rôle des syndicats dans la négociation salariale – et notamment lorsque le syndicat est en position de monopole.

d'adopter le prix de marché (ou même de réduire ses prix relativement à ceux du monopole discriminant). C'est pour ces raisons que le code de déontologie médicale interdit la publicité ou même la critique entre confères.

S'il n'y a pas de marché dans le secteur des soins, c'est du fait de l'entente entre les producteurs. L'écart par rapport au marché ne s'explique pas par les caractéristiques du bien. Il est ainsi relativement aisé de revenir à la situation socialement optimale en détruisant le pouvoir monopolistique de l'AMA. Par exemple, il serait possible de proposer aux patients des assurances santé prépayées afin de casser la possibilité même du monopole discriminant – le prix étant fixé par un tiers. Grace à ce type de politique économique, la liberté individuelle du médecin (consistant en la possibilité de maximiser son profit indépendamment du syndicat) et l'optimum social sont sauvegardés.

L'intérêt de cette approche pour notre propos, repose sur une première mise en évidence du conflit théorique sur la rationalité des offreurs de soins et, en conséquence, sur le degré avec lequel le secteur des soins médicaux peut être comparé au marché de concurrence pure et parfaite. Nous observons une ligne de fracture entre l'hypothèse éthique et l'hypothèse de maximisation du profit. Bien que Kessel préfère l'hypothèse de maximisation du profit, il est intéressant de noter que le secteur des soins n'est pas pour autant un marché de concurrence pure et parfaite. Au contraire, ce secteur est marqué par une situation de monopole discriminant consécutive à l'existence d'un syndicat de médecin tout puissant. Si l'analyse empirique ne permet pas de vérifier les préceptes théoriques, le cadre théorique du marché de concurrence pure et parfaite n'est en rien renversé. L'orthodoxie du propos sera radicalement critiquée par Arrow quelques années plus tard.

1.2.2. Kenneth Arrow (1963), le médecin n'est pas un « barbier »

Bien qu'il existe de nombreux textes en économie s'intéressant au secteur des soins médicaux antérieurs[8], l'article de Kenneth Arrow paru dans l'*American Economic Review* en 1963 est souvent considéré comme ouvrant la voie d'une « authentique » économie de la santé – il faut évidemment comprendre une économie de la santé *mainstream*. Le *Journal of Health Politics, Policy and Law* a d'ailleurs publié une édition spéciale en 2001 entièrement dédiée à l'article d'Arrow. Plus récemment encore a été publié un ouvrage entièrement consacré à

[8] On pourra se reporter, pour les textes les plus marquants, en plus de l'article de Kessel, à Hamilton (1932), Mushkin (1958), Ginzberg (1953, 1954) et Winslow (1951).

l'article de 1963 : son contexte politique de rédaction, les réactions qu'il a suscitées et sa postérité contemporaine (Tanti-Hardouin, 2011).

Il semble incontestable que la notoriété de cet article s'explique en partie par le prestige associé à son auteur[9]. Mais, il est tout aussi incontestable que ce texte rompt avec les modes de pensée antérieurs. L'ambition nouvelle développée par Arrow repose dans la comparaison systématique entre le marché tel qu'il est défini par la théorie néoclassique et l'observation profane[10] du secteur des services médicaux. Dans ce travail fondateur, l'auteur identifie plusieurs séries de problèmes tant sur l'offre que sur la demande de services médicaux ayant pour conséquence un écart vis-à-vis de la situation, pourtant théoriquement optimale, de marché.

D'abord, du côté de l'offre de soin, contrairement à Kessel (1958), Arrow défend l'idée que le comportement attendu du médecin est – et doit être – différent de celui d'un acteur marchand. Comme le produit et l'activité de production sont confondus, le client-patient est dans l'incapacité de tester son achat avant la consommation. S'il veut consommer, le patient doit, en quelque sorte, se jeter à l'eau sans savoir nager. C'est pourquoi la relation médecin-patient est caractérisée par l'existence de confiance. Il n'est pas socialement attendu du médecin qu'il agisse uniquement dans son propre intérêt mais qu'il agisse aussi dans l'intérêt du patient. Le médecin est sujet à une forme de désintéressement et il faut qu'il en soit ainsi car un lien purement monétaire serait inefficace. L'hypothèse éthique refait surface contre celle de maximisation du profit : le médecin n'est pas un « barbier ».

Bien sûr, nous dit Arrow, cette orientation spécifique du médecin est en partie contrebalancée par l'existence d'un aléa moral. Le médecin en sait plus que le patient et l'assureur sur la transaction qui s'effectue. Puisqu'il n'y a pas ou peu de contrôles sur le médecin, celui-ci peut prescrire plus que de raison de façon à augmenter artificiellement son revenu (au dépend du patient et de l'assureur) ou à faire plaisir à son patient (contre l'intérêt de l'assureur). Mais, cela ne remet pas en cause la spécificité de l'offreur de soin. Il est simplement nécessaire d'introduire le contrôle d'un tiers afin que les forces de marché soient, dans le cas de services médicaux, remplacées par un contrôle institutionnel bien plus efficace.

[9] Kenneth Arrow est l'un des premiers économistes à donner une formalisation aboutie du modèle d'équilibre général inspiré par Léon Walras. Il est en cela un fondateur majeur de la théorie néoclassique. Il se distingue également par ses travaux sur l'économie du bien-être et sur la prise en compte de l'incertitude dans les modèles néoclassiques. Mais, selon ses propres mots, il n'est en rien économiste de la santé. D'où le statut étrange de sa contribution à ce segment de la discipline : son seul article (ou presque) sur le sujet est aujourd'hui encore une référence incontournable.

[10] L'auteur parle de l'observation quotidienne (« everyday observation »).

Ainsi, Arrow rappelle l'existence de très fortes barrières à l'entrée de la profession. Le code de déontologie interdit la publicité, la concurrence sur les prix, et autres comportements associés à la coordination marchande. Si dans la théorie néoclassique ces institutions sont une entrave au bon fonctionnement du marché, selon l'auteur c'est l'existence de défaillances de marché (*market failures*) qui rend nécessaire ces écarts à la norme.

Du côté de la demande de soin, Arrow fait remarquer que, mis à part le cas de la prévention, celle-ci est totalement irrégulière et imprévisible. Le secteur des services médicaux est marqué par une double incertitude concernant la maladie (occurrence et effet du traitement) et concernant le médecin lui-même (asymétrie d'information). Cette situation de forte incertitude ne pouvant être résolue par le marché, d'autres mécanismes de coordination doivent prendre le relais. En ce sens, si les médecins sont sujets à l'éthique médicale, c'est pour nouer un lien de confiance avec les patients, sans quoi il n'y aurait pas d'échange. Puisqu'il y a une incertitude sur le produit, il faut la combler par une institution étrangère au marché[11]. Contrairement à ce que laisse supposer Kessel, l'existence de barrières à l'entrée et de normes médicales (notamment véhiculée par les syndicats) ne sont pas destinés à produire les conditions d'un monopole discriminant. Au contraire, il s'agit de construire la confiance des patients dans une situation où règne l'incertitude.

Ce qui est frappant dans le discours d'Arrow, c'est la causalité retenue pour expliquer les écarts au marché de concurrence pure et parfaite. Ce n'est pas parce que des institutions étrangères au marché existent que les marchés sont défaillants, c'est parce que les marchés sont défaillants que des institutions étrangères au marché existent. Arrow inverse la causalité implicitement mise en évidence par Kessel (1958). Il ne s'agit plus ici de rechercher les barrières au marché, mais de montrer comment des institutions sociales viennent compenser les défaillances de marché. Apparait au grand jour l'idée que si le marché des soins n'existe pas, c'est parce que ce marché a des limites intrinsèques et non parce qu'on le limite. La conclusion de politique économique est limpide : le laissez faire est intolérable.

1.2.3. Mark Pauly (1978), l'introuvable consommateur

Mark Pauly (1978) est le véritable père fondateur de l'économie de la santé *mainstream* par son opposition quasi-systématique à Arrow, chez qui il va critiquer tout ce qui est susceptible de remettre en cause trop fondamentalement la TSE1. Après de premières joutes

[11] En ce sens, Arrow anticipe les conclusions d'Akerlof (1970) à propos de l'incertitude sur la qualité.

théoriques à la fin des années 1960 (Pauly, 1968, Arrow, 1968), il critique vigoureusement dans son article de 1978 certaines des positions les plus originales d'Arrow. L'idée générale de sa critique est qu'effectivement il existe une particularité du secteur des soins. Pourtant, d'une part, l'origine de cette particularité est à trouver uniquement du côté de la demande et, d'autre part, la particularité du secteur n'est que marginale.

En effet, selon lui, regrouper sous le terme générique de services médicaux l'ensemble complet des activités des professionnels de santé est une erreur. Il faudrait distinguer dans cet ensemble différent types d'activités. Il propose trois groupes de services médicaux : (i) Des services qui sont achetés fréquemment par de nombreux ménages ; (ii) Des services plus spécifiques produits fréquemment par les offreurs mais qu'un consommateur type consomme relativement peu fréquemment (peut-être même une fois dans sa vie) ; (iii) Des services qu'un producteur produit et qu'un consommateur consomme relativement peu fréquemment. Dans ce cadre, argumente Pauly, les activités de type (i) et (ii) sont bien plus compatibles avec le modèle de concurrence pure et parfaite que celles de type (iii). Il s'agirait d'activités plus répétitives dont il serait possible de standardiser la production et à propos desquelles les consommateurs sont mieux informés.

En outre, concernant l'offreur de soin (médecin, hôpital ou tout autre), Pauly soutient que les suggestions d'Arrow à propos d'un comportement du médecin déviant de l'hypothèse de maximisation ne sont pas fondées empiriquement[12]. Comme les autres offreurs, le médecin maximise son profit. Plus généralement, les économistes de la santé ne conserveront d'ailleurs pas cette partie des développements du texte d'Arrow. La mémoire des économistes n'aura sélectionné que les développements sur l'aléa moral (Bloche, 2001).

Dès lors, revenant à Pauly, qu'y-a-t-il de réellement différent entre le secteur des soins et le marché ? Selon lui, l'originalité des soins provient du côté de la demande : le consommateur n'est pas parfaitement informé et cela implique une imperfection du marché de l'assurance. Tout individu rationnel cherche à se protéger contre les risques en s'assurant. Mais, réciproquement, les assurances refusent d'assurer les individus connus comme des mauvais risques (ou susceptible de l'être). Est ainsi exclue du marché de l'assurance toute une partie de la population qui théoriquement devrait en trouver une. L'équilibre de marché est hors d'atteinte. La résolution de ce problème par une assurance obligatoire (publique ou non)

[12] Notons néanmoins que – comme souvent – si les économistes critiques de l'hypothèse éthique dénoncent l'absence d'éléments de preuve empirique venant soutenir cette hypothèse, aucun d'entre eux ne prend la peine d'apporter une preuve empirique de l'hypothèse de maximisation du profit. C'est notamment ce qu'oublie de faire Pauly dans ce texte – comme s'il s'agissait d'une hypothèse naturelle.

est insatisfaisante. En effet, l'existence même d'une assurance (obligatoire ou non) cause une distorsion de la demande à cause du problème d'aléa moral. En résumé, c'est l'ignorance du consommateur qui est à l'origine de la particularité des soins.

Ce qui est important pour notre propos est la reconnaissance de la part de l'auteur d'une originalité du secteur de la santé relativement au modèle de marché néoclassique. Cette fois-ci la défaillance de marché provient du côté de la demande de soin. Les consommateurs étant faiblement informés et cette information étant coûteuse, le marché ne peut jouer son rôle. C'est, d'après Pauly, la différence essentielle et il récuse ainsi l'hypothèse éthique du côté de l'offreur. Indépendamment de cette question, l'écart entre le marché de concurrence pure et parfaite et le secteur des soins médicaux semble irrémédiable – au moins pour une partie des services médicaux (que Pauly ne définit pas précisément).

1.2.4. Le marché des services médicaux n'existe pas

Chacun à sa façon, ces travaux mettent en évidence la distance entre l'observation profane du secteur des soins médicaux et l'idéal type marchand (Tableau 1.2.). À cause de la forme particulière de l'offre de soin, de la demande ou bien à cause de l'incertitude, un marché de concurrence pure et parfaite n'est pas concevable pour décrire l'organisation des soins médicaux. Notons néanmoins que ces auteurs ne remettent pas en cause cet idéal normatif. Ils ont pour objectif, au mieux, de mettre en lumière les défaillances de marché, sinon, de montrer les barrières qui entravent la logique de marché. Rappelons l'apport majeur du texte d'Arrow pour qui la causalité entre marché et réalité est inversée : contrairement aux hypothèses habituelles, c'est parce que le marché est défaillant que l'organisation sociale produit des institutions coordinatrices hors du marché. Il s'agit là d'une césure importante avec les raisonnements antérieurs portés notamment par Kessel.

Tableau 1.2. Les textes fondateurs d'économie de la santé et l'idéal type marchand

	Perfection du marché	Cause de l'imperfection	Particularité du bien	Comportement du médecin
Kessel (1958)	Non mais réalisable	Hors marché (monopole)	Aucune	Hypothèse de maximisation du profit
Arrow (1963)	Non et irréalisable	Le marché lui-même	Incertitude, offre et demande	Hypothèse éthique
Pauly (1978)	Non et non entièrement irréalisable	Le marché lui même	Incertitude et demande	Hypothèse de maximisation du profit

Ces considérations préliminaires devraient être un puissant avertissement pour les économistes : l'analyse de l'organisation des soins médicaux doit s'éloigner – ou au moins se méfier – de la traditionnelle étude du marché de concurrence pure et parfaite.

1.3. Les défaillances du marché des soins médicaux

La méfiance des premiers économistes vis-à-vis du cadre de la TS s'est confirmée au fil du temps. En effet, l'histoire de l'économie de la santé est aussi une histoire des multiples défaillances de marché qui ont été peu à peu mises en évidence. Nous proposons dans cette section une analyse plus systématique de ces défaillances de marché. Il s'agit d'actualiser les intuitions des premiers économistes néoclassiques de la santé pour qui se pose la question de l'écart entre le marché et l'observation empirique.

1.3.1. Première faille : une concurrence monopolistique

L'écart à la théorie le plus souvent reconnu concerne l'existence d'un pouvoir de marché pour le médecin. L'une des hypothèses centrales du modèle de concurrence pure et parfaite est celle d'atomicité. Pour que le marché soit efficace au sens de Pareto il faut qu'il y ait un grand nombre d'offreurs et de demandeurs. Si cette hypothèse est facile à vérifier du côté de la demande, les choses sont moins simples du côté de l'offre. Il est en effet généralement considéré que les médecins sont en situation de concurrence monopolistique (McGuire, 1983, Dranove 1988, Frech, 1996 et Gaynor, 1994). Le monopoleur est *pricemaker*, ce qui signifie techniquement qu'il maximise son profit en ajustant ses quantités et ses prix. Il ne peut plus y avoir d'ajustement marchand par concurrence sur les prix.

Pour les patients, les médecins ne sont pas des substituts parfaits. De ce fait, il n'y a pas de demande pour les médecins en général (perspective de marché) mais pour un médecin particulier (perspective de monopole). Le patient est attaché à un médecin particulier tant qu'il lui assure une utilité minimum associée par exemple à une localisation géographique, une spécialité, un niveau de qualité, des goûts, etc. Dans ces conditions, au prix fixé par le monopoleur, le patient préférerait consommer moins de services de soin mais, le fait que les services soient non revendables permet au producteur d'exercer son autorité (Encadré 1.1.).

Si l'on suppose l'existence de différents types de patients avec des préférences plus ou moins prononcées pour les services santé, les patients avec une préférence plus forte pour ce type de bien paieront plus pour la même quantité de service que les patients avec ayant une préférence plus faible. Il y a alors une possibilité de discrimination par les prix. Les patients les plus modestes paieront moins parce que leur disposition à payer est plus faible (il n'est pas besoin d'émettre d'hypothèse sur l'élasticité de la demande).

On peut enfin imaginer que le régulateur fixe le prix p. Dans ce cas, le même programme de maximisation que précédemment s'applique (en prenant p comme constante). La diminution du prix par rapport au prix de monopole entrainera une augmentation de la quantité. Puisque le médecin ne peut extraire un surplus par l'augmentation du prix, il peut augmenter abusivement la quantité. Le patient accepte cette situation parce que cela correspond à un plus grand surplus pour lui – au détriment de l'assureur.

Encadré 1.1. La concurrence monopolistique (A partir de McGuire, 2000)

Dans le modèle qui suit nous présentons un modèle de concurrence monopolistique dans lequel le médecin a un pouvoir de marché et le patient quelques alternatives. Les services sont non revendables et hétérogènes. Ce point a des effets importants sur la discrimination par les prix et, plus généralement, sur la fixation des prix et des quantités. Soit la quantité de service (x), les bénéfices retirés par le patient de ses services en dollars (B(x)) et le bénéfice marginal du patient : $b(x) = B'(x)$ avec $b(0) > 0$; $b' < 0$. La fonction B(x) capture les chocs de santé implicitement donc B(0) peut être négatif. Si p est le prix des services de soin, le profit du médecin est :

$$\Pi = px - cx$$

Et le bénéfice net des patients :

$$NB(x) = B(x) - px$$

Si on note x* la solution de

$$b(x) = c$$

C'est le niveau efficient de x. Soit

$$NB^* = B(x^*) - cx^*$$

Le bénéfice maximum net du patient. Notons x^m le niveau de x qui maximise B(x), c'est dire que :

$$b(x^m) = 0$$

En concurrence monopolistique, le patient a des substituts : en général, le patient peut consommer les services de plusieurs médecins en même temps, et le bénéfice des services est une fonction des services consommés. Ici, on force un patient à choisir un médecin. On pose que le patient, en quittant le médecin, peut recevoir un bénéfice NB^0. Le patient utilise le médecin si et seulement s'il reçoit plus que NB^0. En modifiant NB^0, le modèle intègre la concurrence pure et parfaite et le monopole comme un cas limite. Si $NB^0 = 0$, le patient n'a pas le choix et on est dans une situation monopolistique. Si $NB^0 = NB^*$ le marché est parfaitement concurrentiel et le médecin n'a pas de pouvoir de marché. En général : $0 < NB^0 < NB^*$

Le prix et la quantité des services médicaux sont trouvés par la maximisation du profit du médecin, avec la contrainte sur le bénéfice du patient imposé par la concurrence entre médecins. Soit le programme :

$$L = px - cx + \lambda(B(x) - px - NB0)$$

La maximisation par x, p et λ donne :

$$Lp : x - \lambda x = 0$$

$$Lx : p - c + \lambda(b(x) - p) = 0$$

$$L\lambda : B(x) - px - NB0 = 0$$

La résolution donne : $\lambda = 1$, le vendeur gagne tout le surplus au-dessus de NB^0 et chaque relâchement de la contrainte va au profit. x est tel que $b(x) = c$ donc $x = x^*$. Le prix est défini de façon à extraire tout le surplus au-dessus de NB^0 :

$$p = \frac{B(x*) - NB0}{x*}$$

La combinaison choisie (p, x*) donne le profit :

$$(p-c)x^* = NB(x^*) - NB0$$

La quantité est toujours x*. Notons que le patient n'est pas un *pricetaker*. Au prix de p le patient préférerait consommer moins de services que x* mais le fait que les services soient non revendables permet au producteur de décider de la quantité.

Plus généralement, l'existence d'un monopole implique l'exercice d'un pouvoir de monopole pour le médecin. Ce dernier peut alors décider de fixer un prix des services médicaux supérieur au coût marginal. L'ajustement peut se réaliser par les quantités s'il existe une tutelle qui régule les prix. Cet écart par rapport à la situation de concurrence pure et parfaite n'est pas sans conséquence sur le bien-être collectif. En effet, la déviance par rapport à la norme (le marché) s'accompagne d'une réduction du bien-être collectif. C'est pourquoi il

est recommandé à la puissance publique d'adopter des politiques économiques visant à réduire le pouvoir de marché des médecins.

1.3.2. Une synthèse des défaillances de marché des services médicaux

Si le problème du monopole est important pour la théorie économique, il ne remet pas en cause l'idéal de marché. C'est bien plus les défaillances de marché mises en évidence par Arrow (1963) qui impliquent une remise en cause radicale du paradigme néoclassique. Aujourd'hui, la plupart des économistes de la santé s'accordent sur un certain nombre de défaillances (Fargeon, 2009) qu'il convient de rappeler.

- Le « concernement collectif » des services médicaux : Une première défaillance de marché importante porte sur l'existence d'une dimension collective de la consommation individuelle. Il est ainsi d'usage de parler d'un « concernement collectif » (Crozet, 1997) de la consommation de services médicaux. Les services de santé produisent en effet des externalités qui ne sont pas réintégrables dans le marché. Des agents recherchant séparément leur utilité individuelle dans la consommation (ou non) de services médicaux peuvent générer des coûts (externalité négative) ou des bénéfices (externalité positive) non pris en compte par le marché, provoquant alors une situation sous-optimale (au sens de Pareto). Cet écart par rapport à la situation de concurrence pure et parfaite justifie alors l'intervention publique (en dehors de la logique marchande). Afin d'illustrer ce phénomène il est de coutume de prendre l'exemple de la vaccination. Pour se protéger individuellement des maladies transmissibles et infectieuses, il est possible de se faire vacciner. Le coût individuel génère des externalités positives puisqu'il réduit pour tous la probabilité d'être infecté. À l'inverse, la décision de ne pas se faire vacciner au niveau individuel est génératrice d'externalités négatives puisque la probabilité d'être infecté augmente au niveau collectif. Ces externalités ne peuvent être réintégrées au marché. Il est inenvisageable de créer un marché sur lequel s'ajusterait le prix de la décision de se faire vacciner (ou non). C'est pour cela que la puissance publique intervient dans ces cas-là (obligation de se faire vacciner, financement du traitement, etc.). La discussion sur le concernement collectif des services médicaux encourage à les caractériser comme des biens mixtes, entre les biens privés et les biens publics (Encadré 1.2.). La vaccination est à ce titre exemplaire. Si la vaccination donne lieu à un prix (bien privé), les externalités engendrées plaident pour la considérer comme un bien public. Le cas des services d'urgence est également exemplaire : il n'y a pas d'exclusion par le prix (bien public) mais, dans certains cas d'encombrement, il peut y avoir rivalité (bien privé).

Encadré 1.2. Biens publics et biens privés

La fable du phare décrit très bien ce que signifie un bien public. Lorsqu'il est nécessaire de construire un phare sur les côtes d'un pays pour éviter la mort de marins et la destruction de bateaux et de leur contenu, plusieurs aspects sont à prendre en compte. Premièrement, il est impossible d'empêcher quelqu'un de profiter du phare, car tous le voient à l'horizon et il serait trop compliqué de rattraper les bateaux qui ne veulent pas payer. Ce phénomène s'appelle la non-exclusivité. Deuxièmement, le nombre de bateaux qui passent devant le phare ne diminue pas la qualité du bien pour les autres bateaux et n'augmente pas les coûts de production du phare. Ce deuxième phénomène s'appelle la non-rivalité. Ainsi, tous les biens qu'ils soient publics ou privés peuvent être identifiés selon qu'ils sont exclusifs ou non et rivaux ou non.

	Bien exclusif	Bien non exclusif
Rivalité	Bien privé	Bien public impur ou bien commun
Non Rivalité	Bien de club ou à péage	Bien public pur ou bien collectif pur

- Bien privé : bien divisible dont il est possible d'établir un prix couvrant au moins le coût de production. Le prix est la barrière à la consommation.
- Bien public impur ou bien commun : il peut être dégradé par la consommation (exemple : une plage convoitée).
- Bien de club ou à péage : la consommation est non rivale mais il peut y avoir exclusion par le prix (exemple : autoroute à péage).
- Bien public pur ou bien collectif pur : non rivalité et exclusivité (exemple : l'air que l'on respire).

Evidement tous ces exemples sont discutables (exemple : s'il y a trop de monde sur l'autoroute à péage, il y a rivalité) mais ils permettent de dresser une cartographie des différents biens économiques.

- Incertitude, asymétrie d'information et imperfection du marché de l'assurance : Le secteur des soins est caractérisé par des problèmes d'information plus importants que partout ailleurs. Le patient ne connait pas son état de santé ni quelle réponse médicale adopter face à cet état. La maladie et son degré de gravité sont imprévisibles et aléatoires. Par la remise en cause de l'intégrité physique de l'individu et de sa capacité à générer des revenus, l'autonomie du consommateur est en péril. Le consommateur de soin n'est pas un consommateur de voiture.

Comme l'avaient déjà remarqué Pauly et Arrow, ces problèmes d'information se répercutent sur le marché de l'assurance santé. L'asymétrie d'information entre assureur et assuré ainsi que le fait que le risque de santé est un risque de long terme mettent en échec le marché de l'assurance. En économie de la santé, la notion d'aléa moral est centrale. Le patient dispose d'un avantage sur l'assureur ce qui conduit à exploiter cette rente de situation soit par le biais de l'aléa moral *ex ante* (avant l'occurrence de la maladie, le patient se sachant assuré néglige la prévention) soit, par celui de l'aléa moral *ex post* (une fois la maladie survenue, le patient consomme plus que ce que son état de santé exige). Or, il est difficile (et même souvent interdit) de sélectionner les risques. L'incertitude sur les marchés provoque alors leur disparition : le marché de l'assurance est par nature inefficace.

- La relation patient/médecin marquée par l'incertitude : Ce n'est pas la santé qui fait l'objet de la transaction entre le médecin et le patient mais des services de soins (diagnostics, prescriptions, etc.) destinés à améliorer l'état de santé. Evidemment, le patient ne connait pas ce qui est susceptible d'améliorer son état de santé. Il doit s'en remettre au médecin dont il ne connait ni les intentions, ni les compétences. La thématique des asymétries d'information convient également à la relation médecin-patient. Celles-ci sont d'autant plus fortes que le patient n'est pas l'auteur de la demande de services médicaux : la demande est induite par le médecin13. Comment alors protéger la relation médicale et assurer la coordination ? Pour certains économistes suivant le chemin d'Arrow, les institutions seraient alors le socle de la confiance. La situation monopolistique, les barrières à l'entrée (études longues, diplômes exigeants, etc.), les codes éthiques, l'administration des prix, l'intervention des États, seraient des constructions sociales permettant d'assurer sans assurance un niveau de qualité minimal.

- Absence d'un output clairement identifié : Une autre défaillance de marché repose sur la difficile définition de la production du médecin. Il ne produit pas de la santé : les services de santé ne sont qu'une composante de l'état de santé. Il ne produit pas du tout : il rend des services dont les caractéristiques ne sont pas celles des produits (Gadrey, 1991). On peut brièvement lister plusieurs attributs des services : immatérialité (il ne peut pas être testé et/ou rendu), co-production avec l'usager (le patient peut même être considéré comme un input de la fonction de production), la production et la consommation sont simultanée (la constitution de stock est impossible), distinction de l'output (service en actes, plutôt à court terme) et de l'outcome (résultat en termes d'état de santé, plutôt à long terme), inscription sociale du service (la définition du produit et de sa qualité n'est pas objective, elle dépend du système de valeur/critères de jugement).

Toutes ces particularités vont à l'encontre de la TS et le marché semble être une catégorie à dépasser. C'est dans cette perspective qu'Arrow va proposer aux économistes un nouveau programme de travail.

[13] Nous explicitons ce point dans la seconde section de chapitre.

1.4. La stratégie du Nirvana en santé : le retour du marché ?

1.4.1. Le marché n'existe pas mais les économistes parlent de marché

Si de prime abord le *mainstream* s'est éloigné de l'idéal type marchand, il entretient une relation ambigüe aux concepts et outils issus de l'analyse de la concurrence pure et parfaite. Cette remarque n'est pas neuve. Au début des années 80, Darbon et Letourmy (1983) mettaient en évidence l'étrange relation des économistes de la santé avec le marché :

> *« Ces remarques [à propos des limites et incohérences des modèles économiques standards] indiquent l'ambiguïté de la place faite à la notion de marché par la micro-économie des soins. À plus d'un titre, cette notion n'est pas indispensable. Mais sa récurrence dans les travaux théoriques, et dans les débats sur le système de santé, suggère que ses fonctions idéologiques justifient peut-être son maintien. Il suffit de rappeler que la notion de marché des soins est utilisée dans des contextes où les principes du libéralisme sont officiellement préservés, alors même que la socialisation des dépenses de santé s'étend lentement. » (Darbon et Letourmy, 1983, p. 42)*

Ce que suggèrent ces auteurs, est que l'idéal type de marché reste présent dans la majorité des travaux en économie de la santé bien que les économistes néoclassiques disent s'en éloigner. Il y a comme un paradoxe qui voudrait que le secteur de la santé ne soit pas un marché mais qu'il soit étudié avec les outils conceptuels forgés pour la compréhension de la régulation marchande (offre, demande, concurrence, etc.). Selon nous, pour comprendre comment ce paradoxe *a priori* propre à l'économie de la santé est né, il faut faire un détour par l'évolution plus générale du *mainstream*.

1.4.2. Le programme d'Arrow et l'avènement de la théorie standard étendue

Le début des années 70 voit la théorie économique amorcer un virage majeur. Des deux piliers de la TS, seul celui portant sur la rationalité substantielle et intéressée subsiste. La multiplication des travaux portant sur les défaillances de marché – pas uniquement en économie de la santé – met en évidence qu'à côté du marché d'autres modes de coordination existent. C'est encore une fois Arrow (1970) qui va jouer un rôle majeur. Après avoir contribué à mener à bien la théorie de l'équilibre général walrasien, il énonce aux économistes la nouvelle orientation de leur programme de recherche.

Selon le prix Nobel, l'existence désormais prouvée de défaillances de marché rend le marché inefficace au sens de Pareto. Le pilier de coordination exclusivement marchande n'est plus tenable. Dans les cas où les marchés ne sont plus efficaces, d'autres institutions prennent le relais pour assurer la coordination. Il ne s'agit pas simplement de redonner une légitimité à

l'État face au marché mais, plus largement, de démontrer l'existence d'actions collectives en dehors du marché. Arrow parle d'institutions sociales capables de générer des normes de comportement efficaces lorsque les marchés sont défaillants. L'auteur distingue trois types d'institutions sociales assurant la coordination – en plus du marché : l'État, l'entreprise et les normes sociales (Arrow, 1970). Tout en conservant l'hypothèse de rationalité standard, cette extension des modes de coordination possible est l'acte de naissance de la Théorie Standard Etendue.

Paradoxalement, alors que le marché n'est plus considéré comme le seul mode de coordination, la théorie standard va s'arc-bouter sur ce qu'il lui reste de plus précieux : la théorie du choix rationnel (Favereau, 2010). Plutôt que de parcourir ces nouveaux champs d'investigation avec des yeux neufs, les économistes néoclassiques vont étendre l'hypothèse de rationalité à tous les domaines imaginables. De la sorte, ils vont étendre la sphère des activités marchandes au-delà du seul marché. Chaque interaction est potentiellement marchande :

> *« Le royaume du calcul rationnel est si étendu sur la surface de la terre que le soleil ne s'y couche jamais. Dans la mesure où marché et rationalité intéressés ont partie liée pour le sens commun, la relativisation du marché, sur un certain plan (micro ?), révèle en filigrane tout le contraire, une sorte d'absolutisation, sur un autre plan (macro ?) » (Favereau, 2010, p. 116).*

Ainsi, malgré les défaillances de marché, le *mainstream* a continué à plaquer le concept de marché là où il semblait ne plus avoir sa place. Les institutions ne sont considérées que comme une solution de substitution à une situation marchande qui reste théoriquement optimale. Quelles que soient les défaillances de marché, toutes les situations doivent être évaluées à partir de l'idéal type marchand. Si le premier pilier de la théorie néoclassique reste l'agent rationnel, le pilier de coordination marchande est substitué par le pilier de normativité marchande. Formellement, il n'est pas nécessaire de considérer le marché comme unique mode de coordination, il suffit implicitement de considérer que c'est un idéal normatif à atteindre en dépit de tous. Le marché rentre par les institutions : c'est la stratégie du Nirvana.

1.4.3. La stratégie du Nirvana : vers le marché des soins médicaux ?

Harold Demsetz (1969) fut le premier à critiquer l'ambition de la TSE à se servir des résultats de la théorie de la concurrence pure et parfaite pour discuter de ce qu'Arrow a appelé les institutions sociales (État, entreprise et normes sociales). Dans son texte, l'auteur cherche à évaluer la notion d'efficacité (*efficiency*) utilisée par les économistes lorsqu'ils qualifient différents types d'organisations. Le courant dominant présente chaque choix comme une

alternative entre une norme idéale – le marché de concurrence pure et parfaite – et un arrangement institutionnel imparfait. Tout se passe comme si le monde réel, enlaidi par les imperfections (externalités, asymétrie d'information, coûts de transaction, etc.), devait nécessairement tendre vers la situation théorique de marché. C'est le réel qui est inefficace.

Aujourd'hui cette méthodologie se donne à voir à travers l'usage du concept de *second best* (Lipsey et Lancaster, 1956). Dans un grand nombre de modélisations néoclassiques contemporaines, la démonstration implicite ou explicite s'opère en deux temps. Dans un premier temps, le modélisateur décrit ce que serait la situation qui l'intéresse si les hypothèses du marché de concurrence pure et parfaite étaient réalisées[14]. De là, découle la norme. Attention, il ne s'agit pas d'une norme au sens statistique du terme. Au contraire, la norme au sens statistique est la défaillance de marché. Ici, le mot norme est utilisé dans un sens normatif. C'est un jugement de valeur sur ce qu'il convient de faire ou d'encourager : la meilleure situation possible. Evidemment, le critère normatif est faible puisqu'il s'agit du critère d'optimum de Pareto. Mais, il convient aux économistes qui le qualifient de *first best*. Dans un second temps, il devient nécessaire de revenir au monde tel qu'il est, non tel que l'on voudrait qu'il fût. Dès lors, les défaillances de marché interdisent l'atteinte de l'idéal[15]. Tout n'est cependant pas perdu. La théorie économique va s'évertuer à définir les mécanismes incitatifs permettant de réduire l'écart entre le réel et la norme[16]. Cette situation, non totalement satisfaisante, est qualifiée de *second best*.

Selon Demsetz (1969, p. 2) commentant un texte d'Arrow (Arrow, 1962), la stratégie du Nirvana commet trois erreurs logiques :

(i) *The grass is always greener fallacy* : en comparant une situation réelle à la situation idéale, il est facile d'appeler à plus de perfection par incantation. Par exemple, à cause de l'aversion au risque, les entreprises sous-investissent. Il peut alors être tentant de recommander un investissement public. Mais, trop rarement, les situations réelles sont comparées entre elles. Dans ces conditions, si l'herbe semble toujours plus verte ailleurs, jamais la démonstration n'en est faite. Il peut sembler plus efficace de rémunérer le médecin à sa productivité marginale, mais ne vaut-il pas mieux comparer la situation réelle à une situation réalisable ?

[14] Cf. section 1.1.

[15] Cf. section 1.2. et 1.3.

[16] Cf. section 2., Chapitre 2 et Chapitre 3.

(ii) *The fallacy of the free lunch* : la TS suppose l'existence de marchés contingents. Dans le monde réel ceux-ci n'existent pas. Or, si l'on suppose que le coût de mise en place de ces marchés est supérieur au coût d'ajustement au risque, il n'y a pas de raison de considérer que l'absence de marché contingent est non optimale. À moins de penser, ce qui serait cocasse pour un économiste, que l'ajustement au risque est gratuit. L'approche du Nirvana repose sur une hypothèse implicite de non rareté. Penser que la rémunération optimale du médecin est la rémunération à la productivité marginale ne suppose-t-il pas que les coûts d'information et de contrôle pour l'entreprise soient nuls ?

(iii) *The people could be different fallacy* : dans la théorie néoclassique le problème d'aléa moral conduit à une inefficacité sur le marché de l'assurance. Mais, il est aussi possible de considérer l'aléa moral comme un coût comme un autre. Tout le minerai de fer n'est pas déterré parce qu'il est trop coûteux de le faire. Ce n'est pas pour autant qu'il est judicieux de dire que l'extraction de minerai est inefficiente parce qu'elle est incomplète. Certains risques ne sont pas couverts parce que le coût de l'aléa moral est trop grand. Le strabisme dirigé vers le Nirvana de marché peut rendre cette situation étrange et sous optimale mais elle ne l'est pas. Il est alors tentant de sombrer dans l'erreur selon laquelle les gens pourraient être différents. N'est-il pas parfois plus avantageux pour l'assureur de subir un aléa moral de la part du médecin que de le contrôler systématiquement ?

En résumé, pour Demsetz, la TSE – en tant que stratégie du Nirvana – commet des erreurs grossières qui peuvent s'expliquer par l'histoire de la théorie économique. Longtemps celle-ci a cherché à démontrer que la concurrence était plus efficace que les situations monopolistiques. Cette assertion est, selon l'auteur, crédible dans la mesure où il possible d'éliminer le monopole et qu'il existe au moins un autre arrangement institutionnel pour le faire : la concurrence. La TSE commence à s'égarer lorsqu'elle cherche à évincer des problèmes tels que l'aléa moral du monde réel. Evidemment, la TSE considère que l'aléa moral existe et qu'il est source de problèmes. Mais, en comparant la situation du monde réel à la situation de Nirvana, elle suppose implicitement qu'il est possible d'imaginer un monde sans aléa moral. En ce sens, l'aléa moral existe mais il serait possible de le supprimer ou de le réduire par de bonnes incitations. S'il est possible d'éliminer le monopole, il est absurde de vouloir éliminer l'aléa moral. Il est absurde de vouloir l'évincer de la réalité.

Pour le dire autrement, nous pensons que les défaillances de marché ne sont pas toutes de même ordre. S'il est envisageable d'imaginer des arrangements institutionnels autres que le monopole, il est chimérique de prétendre supprimer les asymétries d'information. Le travail

de Demsetz permet de repenser la critique de la TSE. Le problème avec cette approche porte sur la façon dont elle pose les problèmes. Selon elle, les éléments du monde réel qui sont absent de la TS – comme l'existence d'asymétrie d'information, les externalités, l'hypothèse éthique, etc. – sont systématiquement posés comme des problèmes. Or, si l'on anticipe sur la suite de la thèse, ces éléments ne sont pas que des problèmes mais peut-être aussi des solutions. Avant d'en arriver là, il faut conclure sur l'évolution de la TS en TSE en économie de la santé.

Alors que dès les années 1960 (Arrow, 1963, Demsetz, 1969), tout semblait en place pour remettre en cause le paradigme néoclassique dans le secteur des soins – le marché n'est pas une catégorie pertinente pour analyser le secteur et la stratégie du Nirvana est plus que douteuse –, finalement, le programme d'Arrow (1970) sera mené à bien. Toute interaction sera analysée à partir de la théorie du choix rationnel et de la comparaison au *first best* – la concurrence pure et parfaite. La position d'Arrow est d'ailleurs paradoxale pour les économistes de la santé. Si son texte de 1963 inverse la causalité entre défaillance de marché et instituions, ses travaux ultérieurs en dehors de l'économie de la santé vont avoir pour effet d'influencer les autres économistes vers… un retour en arrière. Le *first best* est bel et bien la situation de marché. Arrow conserve néanmoins une forte radicalité en économie de la santé que vont marginaliser ses successeurs au premier rang desquels on retrouve Mark Pauly. Oubliée l'orientation sociale du médecin, lui comme n'importe quel autre individu est un maximisateur agissant dans un cadre contractuel. La TSE sera d'abord une TSE1.

2. Le paiement à la performance médicale pour combler l'écart entre le Nirvana et le monde réel

Dans le cadre de la stratégie du Nirvana, l'analyse de la réalité est faite à l'aune de la norme du marché de concurrence pure et parfaite telle qu'elle a été rappelée dans la section précédente. La situation optimale – le *first best* – défend l'idée que le médecin, en tant qu'offreur de travail, doit être rémunéré à sa productivité marginale :

$$R_f = Pm^L$$

Où R_f est la rémunération de *first best* et Pm^L la productivité marginale du travail. Voilà un résultat très commun et, *a priori*, de peu d'intérêt. Il s'agit cependant de l'un des résultats

le plus problématique du *mainstream* dans la mesure où, en dépit de ses nombreuses incohérences, il reste jusqu'à ce jour la norme à atteindre.

Dans la perspective de la TSE, l'existence d'asymétries d'information (et d'autres défaillances de marché) rend impossible la réalisation du *first best*. Le passage de la TS à la TSE est alors marqué par le renversement de la relation salaire-productivité. Alors que dans le cadre de la TS, l'homogénéité du bien implique seulement un problème de participation, dans le cadre de la TSE, l'incertitude sur la qualité engendre un problème d'incitation à l'effort. Dans le premier cas, la productivité détermine le salaire, dans le second, c'est le salaire qui détermine la productivité. L'analyse économique se déplace d'un problème quantitatif à un problème qualitatif. La problématique générale de la TSE porte alors sur la question de l'incitation à l'effort : il est nécessaire de trouver un mode de rémunération permettant d'inciter à l'effort et donc de réduire l'incertitude sur la qualité du travail fournit (ou du bien échangé).

En dépit de ce renouvellement théorique, les économistes de la TSE vont analyser le comportement du médecin en général et les modes de rémunération existants en particulier à partir de la norme de concurrence pure et parfaite[17]. Les concepts de sanction marchande et de discipline de marché naissent à ce moment-là pour exprimer l'incapacité du réel à produire les incitations permettant d'atteindre le *first best*. Or, il n'y a pas de sens à parler de sanction marchande pour la situation de TS : en concurrence pure et parfaite il n'y a que des problèmes de participation au marché. Le problème d'intensité de l'effort naît avec la TSE et il est faux de postuler qu'en concurrence pure et parfaite les individus sont sanctionnés par le marché lorsqu'ils ne produisent pas d'effort – cette question ne se pose tout simplement pas. Paradoxalement, c'est donc la TSE qui invente la sanction marchande.

Pourtant, indépendamment de ce qui vient d'être dit, les économistes de la TSE vont rechercher à définir des mécanismes reproduisant la « sanction marchande » pour que le réel se conforme à la TS. Autrement dit, ils vont chercher une solution au problème de qualité dans une théorie ne discutant que de quantités. Dans cette optique, il faut se rapprocher de la règle d'or de régulation du médecin : la rémunération à sa productivité marginale. L'idée générale est relativement simple, si le médecin ne fait pas d'effort, il ne doit pas être payé ; s'il fait un effort, il doit être payé à la mesure de ce que cet effort permet de produire. Bien

[17]On pourrait alors s'interroger sur la pertinence à prendre pour norme une situation discutant l'allocation de quantités (TS) alors que les nouveaux problèmes de la théorie portent sur la production de qualités (TSE).

que le médecin soit réticent à l'effort et intéressé, il suffirait de reproduire la « discipline de marché » assurer une coordination de *second best*.

Nous verrons dans la suite de ce chapitre que le seul *second best* convainquant pour la TSE1 est l'introduction du paiement à la performance (P4P). Le P4P médicale est en effet destiné à inciter le professionnel à produire l'effort en établissant un lien arithmétique entre un indicateur de productivité et la rémunération. Il est alors possible d'interpréter toutes les recherches standards en économie de la santé comme des tentatives de se rapprocher de la rémunération de *first best*. Le P4P – auquel il faut ajouter une part de rémunération fixe constituant la contrainte de participation – est alors le résultat le plus abouti de cette tentative. C'est lui qui permet de se rapprocher au plus près de la rémunération à la productivité marginale. Le P4P est en ce sens l'héritier de la norme de concurrence pure et parfaite.

La démonstration s'opère en deux temps. Les sections 2.1. et 2.2. se désintéressent du mode de rémunération optimal du médecin. Leur objectif est de montrer comment les économistes se posent les questions dans le cadre de la TSE1. L'incertitude et les asymétries d'information sont au cœur des préoccupations (2.1) car, lorsque le médecin dispose d'un avantage informationnel, il peut manipuler la demande (2.2.). Une fois ce cadre d'analyse en tête se repose la question de la rémunération du médecin. Nous montrerons en quoi les modes de rémunération traditionnels sont jugés inefficaces à l'aune de l'idéal marchand (2.3.). Il sera alors temps d'en venir au P4P comme solution aux problèmes d'information qui intéressent la TSE1 (2.4.).

2.1. L'insupportable asymétrie d'information

Le *mainstream* ne résout que les problèmes qu'il se donne en référence à la TS. Les plus importants d'entre eux sont les problèmes d'information.

2.1.1. L'incertitude médicale

Comme l'avait annoncé Arrow (1963), l'incertitude est pour les économistes standards le facteur le plus important influençant le comportement du médecin. D'après Wennberg (1985), il est possible de distinguer plusieurs origines à cette incertitude : (i) incertitude liée à la classification du patient en termes d'état pathologique ou d'état de santé initial ; (ii) incertitude liée à l'effet du traitement ; (iii) incertitude liée aux préférences des patients qui ne peuvent pas être parfaitement connues des médecins. Bien qu'il soit envisageable de réduire

en partie l'incertitude, quels que soient les progrès dans l'accumulation d'information, il reste toujours une incertitude irréductible (*irreductible uncertainty*, Pauly, 1978) qui implique, pour tous, l'absence d'information.

Pour comprendre les conséquences de l'incertitude irréductible sur la TS, il est possible d'interroger son impact sur le bénéfice du patient. L'utilité marginale de la consommation de soins et l'utilité marginale de l'amélioration de l'état de santé son décroissantes. Du fait de l'incertitude irréductible, le patient et le médecin sont incertains quant à l'état de santé initial du patient et sur l'amélioration de l'état de santé consécutive à la consommation de services médicaux. Si les patients sont averses au risque (*risk aversion*) et que cette aversion est décroissante des revenus, la théorie économique prévoit que les bénéfices espérés des services médicaux croissent avec le niveau d'incertitude (Pratt, 1964).

Dès lors, l'existence d'une incertitude irréductible sur l'état de santé initial et/ou sur l'effet des traitements engendre pour le patient le désir de compenser ce risque en consommant plus de services médicaux. Le monde réel s'écarte du Nirvana de marché. Du point de vue de son état de santé, une partie de la demande peut sembler indigne d'intérêt, au sens où le coût est élevé relativement à la hausse espérée de l'état de santé. Mais, du point de vue de l'utilité *ex ante*, l'utilisation peut sembler être digne d'intérêt. En effet, puisque le patient ne sait pas quel sera l'effet du traitement, il est raisonnable pour lui de surinvestir dans le traitement pour s'assurer d'un résultat minimal alors que s'il connaissait parfaitement les effets du traitement, il consommerait moins de services médicaux. Les patients peuvent rationnellement chercher à s'assurer contre le risque qui pèse sur leur état de santé par la surconsommation de soins médicaux. L'aversion au risque peut ainsi augmenter la valeur perçue de la procédure de diagnostic et donc augmenter exagérément les dépenses – relativement au niveau optimal des dépenses de santé (Woodward et al., 1998). La conséquence de politique publique est claire : il est nécessaire de contrôler le niveau des dépenses parce que structurellement inflationnistes.

L'incertitude parait poser de gros problèmes à la théorie économique. Cependant, il faut noter que cette dernière ne traite jamais l'incertitude mais uniquement le risque. L'incertitude suppose qu'il n'existe pas de probabilité – même subjective – permettant de faire des calculs sur l'avenir. De même, les événements de la nature ne sont pas tous connus. Or, lorsque la TSE parle d'incertitude, elle le fait en attribuant une probabilité à des événements de la nature connus par avance. Elle traite donc de risque et non d'incertitude. Le risque peut alors faire

l'objet d'un traitement traditionnel (ou presque). Le plus gros problème de la TSE1 n'est donc pas tant l'incertitude – réduite au risque – mais l'existence d'asymétries d'information.

2.1.2. Les asymétries d'information

A ce problème d'incertitude, peut s'ajouter un problème d'asymétrie d'information. La théorie économique s'intéresse ici au fait qu'une partie de ce qui est inconnu du patient est (peut-être) connu du médecin. Cette situation d'asymétrie d'information n'est rien d'autre que le fondement de la théorie de l'agence (Encadré 1.3.). La relation patient/médecin constitue alors certainement un *gold standard* de la relation d'agence dont les économistes de la santé se sont largement emparée (Dranove et White, 1987, Mooney et Ryan, 1993, Gaynor, 1994, Blomqvist, 1991, Lundback, 1998, Zweifel et Breyer 1997).

L'asymétrie d'information concerne également la relation entre le médecin et l'assureur. La TSE1 considère que l'asymétrie peut être si forte qu'il n'y a aucun moyen pour une partie en dehors du contrat de savoir ce qu'un médecin a fait ou savait. Les patients ne peuvent pas rendre compte des agissements du médecin et certains aspects du traitement ne laissent pas de trace (par exemple, on ne peut vérifier l'existence d'une souffrance). Le problème principal – le problème du principal – est de s'assurer de la qualité des soins et de contrôler l'effort de l'agent. Puisque la « discipline de marché » est inopérante, il faut trouver un mécanisme différent permettant d'induire l'effort : ce sera l'incitation.

Encadre 1.3. Présentation canonique de la relation d'agence. Le modèle de Ross (1973)

La relation d'agence est l'une des thématiques les plus anciennes de la théorie économique. Si l'on remonte souvent aux travaux de Berle et Means (1932) pour dater les recherches sur cette question, il est possible de remonter bien plus loin si l'on considère qu'il ne s'agit rien d'autre que d'une relation de subordination. Dans le cadre de la TSE, il faudra attendre l'article de Stephen Ross (1973) pour avoir une première formulation du problème d'agence :

> « *We will say that an agency relationship has arisen between two (or more) parties when one, designated as the agent, acts for, on behalf of, or as representative for the other, designated the principal, in a particular domain of decision* » (p. 134).

L'agent doit choisir une activité (a) dans un espace des possibles (A). Un produit (p) est associé à chaque acte en fonction de l'état de la nature (n), inconnu au moment du choix, et de l'activité elle-même. Par hypothèse, les deux individus ont convenu d'une rémunération (f) pour les services de l'agent qui dépend de l'état de la nature et du produit. D'où :

$$f = f(p(a , n) n)$$

où l'agent va chercher à maximiser son utilité espérée (E). La solution du problème de l'agent implique le choix de l'acte optimal (a*), conditionné à la rémunération associée maximisant son utilité. Le principal, à partir de sa connaissance des préférences de l'agent, a pour seul problème de motiver l'agent à exercer l'effort maximisant sa propre utilité (π).

Plusieurs stratégies pour récolter les informations cachées par le médecin sont possibles. Par exemple, il est vraisemblable que plus le médecin fait d'effort, meilleurs seront les *outcomes*. S'ouvre alors la possibilité de payer directement le médecin sur les *outcomes* – en tant qu'indicateur imparfait de l'effort – afin de résoudre le problème du principal (Lu, 1999). Même si les *outcomes* ne sont pas contractualisables, ils sont observables par les clients/patients. Si le médecin est rémunéré pour produire un meilleur *outcome*, parce que cela augmente la probabilité que le patient revienne ou qu'il recommande le médecin à d'autres patients, le médecin est encouragé à entreprendre correctement les actions non observées pour améliorer la qualité. Il semble pourtant impossible de payer des médecins sur leurs résultats en termes d'état de santé établis par les patients. Il serait trop coûteux de valider un rapport de patient sur son propre état de santé. Le patient pourra toujours dire qu'il a encore mal et personne n'est en mesure – à bas coûts – de vérifier si le médecin a ou non correctement fait son travail.

Par ailleurs, même si l'on considère que le patient peut observer l'*outcome* de l'action du médecin, le patient ne sait pas si l'action du médecin était appropriée. En effet, l'*outcome* est lui aussi affecté par des facteurs non observables (incertitude, environnement, caractéristiques génétiques, par exemple) qui interdisent d'en faire un *proxy* de l'effort.

Comment alors s'assurer du bon comportement du médecin ? Si le médecin sait quelque chose qui peut bénéficier au patient, il faut trouver une solution pour que le médecin révèle son information avec précision afin que le patient puisse choisir le montant optimal de service de soin qu'il désir. Un résultat important – déjà présent chez Arrow (1963) – postule qu'une

solution potentielle de l'asymétrie d'information est le transfert de tous les risques au médecin. L'assurance idéale protégerait le patient contre la portion du risque sur l'état de santé qui pourrait être amélioré par l'action du médecin. Cependant, ce transfert de risque s'avère impossible parce que l'état de santé n'est pas contractualisable (White, 1987, Mooney et Ryan, 1993). Le patient pourrait en effet exploiter la situation en en voulant toujours plus (« ça fait toujours mal »).

Dans le cas général, lorsque les *outcomes* sont observables et contractualisables, il est souhaitable de transférer le risque sur l'agent. Si l'agent est averse au risque, il faut que le principal compense ce transfert de risque en le partageant : le contrat doit prévoir une prime de risque pour l'agent. Le médecin doit ainsi être rémunéré en partie pour assumer ce risque. Il s'agit de la contrainte de participation du médecin. Pour chaque patient, le médecin doit alors percevoir une prime indépendante des *outcomes*.

La situation d'asymétrie d'information est particulièrement saillante dans le cas de la santé mais elle ne lui est pas exclusive. Le médecin peut être considéré comme un expert (de la santé) de la même façon que le garagiste est un expert (de mécanique automobile). Le garagiste a toujours une incitation à dire que le problème du consommateur est plus important que ce qu'il est réellement pour augmenter la demande de services[18]. La concurrence entre experts (qui peuvent avoir des compétences hétérogènes) permet néanmoins de limiter le potentiel d'exploitation (Emons, 1997). La différence avec le secteur des soins est l'absence de concurrence. On peut alors s'attendre à ce que les processus de marché fonctionnent différemment dans le cas de la santé – les experts ne pouvant pas jouer sur les prix.

Dans le cas où le médecin est payé à l'unité, la demande peut par son comportement influencer le niveau de l'effort du médecin. Le médecin, rémunéré à l'acte, peut agir de manière à augmenter le volume de son activité et donc le volume de services utilisés et payés par le patient. Il encourt néanmoins le risque de décourager les patients à venir voir un médecin qu'ils jugent trop dispendieux (Pauly, 1980). Le patient peut également rejeter plus probablement une recommandation (de prescription, d'acte, de traitement, etc.) si le médecin a une mauvaise réputation – c'est-à-dire une réputation de sur-prescripteur (Dranove, 1988). Le médecin a ainsi intérêt à modérer l'expression de son intérêt par la sur-prescription à cause de la menace d'*exit* du patient (Rochaix, 1989). En effet, mécontent, le patient peut exercer un pouvoir de marché et aller à la concurrence.

[18] Nous effleurons ici le problème de la demande induite sur lequel nous revenons dans la section suivante.

Il est possible d'imaginer une rémunération indépendante du nombre d'actes. Dans cette optique, un offreur peut avoir une incitation à utiliser son avantage informationnel pour décourager le traitement (Newhouse, 1996). Ajoutons à cette question celle de l'incertitude : comme la mesure de l'*outcome* ne peut pas être un indicateur fiable de l'effort du médecin, le patient ne peut plus exercer de pouvoir de marché. Il est totalement captif. Au mieux, pour découvrir le niveau d'effort, le patient peut apprendre quelque chose de l'effort inobservé du médecin si les interactions sont répétées.

Il est aussi possible de réinterpréter l'effort réalisé comme une caractéristique inaltérable du médecin – et non comme un choix qu'il faut contrôler. Il est ainsi possible d'assimiler l'effort à la qualité entendue comme une caractéristique fixe du médecin. Les conséquences de cette caractéristique inaltérable sont assez originales. En supposant que les patients valorisent différemment cette caractéristique, l'efficacité économique requiert de trouver le bon médecin pour le bon patient. Si le *mix* de qualité peut être altéré à long terme en changeant la composition des médecins, l'efficacité requiert que la qualité soit récompensée.

Dans ce cadre, le médecin est un « bien d'expérience » (Gaynor, 1994). Le patient doit « essayer » le médecin et peut inférer sa qualité (Hoerger et Howard, 1995, Satterthwaite, 1979). Parce que l'apprentissage est imparfait et lent, les récompenses du marché à la qualité peuvent être inadéquates. L'asymétrie d'information sur le type du médecin est alors une base permettant d'expliquer la concurrence monopolistique dans le champ des soins (Pauly et Satterthwaite, 1981). Lorsqu'il y a plus d'un médecin sur un marché, un patient tendra à en savoir moins sur son médecin parce que l'information est dispersée. La croissance du nombre de médecin n'a donc pas pour effet d'augmenter la concurrence mais, en diluant la qualité de l'information des patients, elle augmente l'allégeance d'un patient à son médecin.

Puisque les patients doivent inférer la qualité par de bons *outcomes* et qu'il est toujours possible qu'un « mauvais » docteur soit chanceux, la valorisation de marché risque inévitablement de souffrir d'une régression vers la moyenne. Les mauvais médecins seront mieux évalués qu'ils ne le devraient et les bons médecins seront moins bien évalués qu'ils ne le devraient. Etant donnée l'uniformité (administrative) des prix, aucun patient ne pense que son médecin est en réalité en dessous de la qualité moyenne – ce qui est statistiquement faux (McGuire, 1983).

2.2. La fin du marché : l'hypothèse de demande induite

L'hypothèse de demande induite constitue un des domaines les plus développés en économie de la santé. Cette hypothèse provient du travail fondateur d'Evans (1974) qui s'appuie initialement sur une remarque profane : le médecin en sait plus que le patient sur ce qu'il convient de faire, ce qui produit une infantilisation du patient et une dépossession de son pouvoir de consommateur. Il est alors possible que ce soit le médecin lui-même qui fixe le niveau de la demande. Le médecin peut altérer les préférences du patient dans son propre intérêt. Cette possibilité menace le concept paradigmatique de marché et ébranle les implications normatives des économistes sur la politique de marché. En effet, l'économie positive et normative sont fondées en partie sur l'idée que les consommateurs ont des préférences stables et qu'ils exercent des choix souverains. Les combinaisons prix/quantités peuvent être interprétées en termes d'intérêt pour le consommateur et d'efficacité de marché.

Il faut mesurer les implications de politique économique aux États-Unis des années 1970 pour comprendre la portée subversive de cette hypothèse. Longtemps la politique économique dans le secteur des soins de santé s'est focalisée sur le côté de la demande. La problématique générale autour de la *National health insurance* était liée au degré de subventionnement des prix qui doit être contenue dans la politique nationale. Plus précisément : combien les assureurs doivent-ils rembourser ? En prenant appui sur les développements de Pauly sur la question de l'aléa moral, il a longtemps semblé clair que pour s'approcher du niveau de consommation optimal des soins, il fallait mettre en place une politique de partage des coûts avec les patients. Puisque les patients ont tendance à abuser des assurances, il faut réduire le taux de couverture des assurances.

Or, si l'hypothèse de demande induite est juste, si les médecins peuvent induire la demande, la politique publique construite sur la théorie du consommateur est un danger de santé publique (Reinhardt, 1989, Rice, 1998). En effet, si le médecin induit la demande, ce sont les conditions d'existence même du marché qui – encore une fois – s'effondrent.

L'hypothèse de demande induite, en rompant avec la théorie de l'offre et de la demande, est donc très importante pour la politique publique. Par exemple, réduire le nombre de chirurgiens peut avoir pour effet de réduire le nombre d'interventions chirurgicales non nécessaires (Schroeder, 1992). Plus généralement, réduire la consommation passe par la réduction de l'offre et non par l'effet prix – comme le supposent les politiques de co-paiement.

S'il est facile de comprendre l'intuition de cette hypothèse, il est plus difficile d'en donner une définition. Celle de McGuire (2000) permet de montrer les problèmes liés à la demande induite :

> *« Physician-induced demand exists when the physician influences a patient's demand for care against the physician's interprÉtation of the best interest of the patient » (McGuire, 2000, p. 504).*

Deux distinctions importantes sont à faire pour délimiter les frontières de la demande induite :

- La première distinction oppose l'induction et l'agence. Selon Fuchs (1978), pour juger de l'induction ou non, il est nécessaire de se référer à la norme de consommation optimale. En ce sens, il est possible que le médecin influence le patient sans qu'il y ait induction. Si un médecin influence la demande d'un patient pour l'amener à son niveau de consommation optimal, ce ne serait pas de l'induction – uniquement de la conséquence de la relation d'agence. Cette dernière situation n'est qu'un transfert d'information mutuellement avantageux comme il est prévu par la théorie de l'agence. L'influence du médecin n'est pas un problème, c'est ce qui est attendu de lui. Pour qu'il y ait induction, il faut que la prescription de services de soins du médecin soit telle que le même patient, mais bien informé, aurait choisi un autre niveau de consommation (Eisenberg, 1986). À partir de cette définition Pauly (1980) utilise le concept d'agent parfait (*perfect agent*). Un médecin agissant comme un agent parfait aide le patient à demander

> *« […] exactly those quantities of care of various types that the patient would have chosen if he had the same information and knowledge the physician has » (Pauly, 1980, p. 5).*

L'influence sur la demande n'est pas suffisante pour établir l'induction, il faut une influence indue. La question empirique majeure devient alors de savoir comme délimiter cet excès. Les économistes considèrent que s'il y a un changement dans le rendement de l'activité du médecin (par exemple une hausse des honoraires) qui stimule un changement dans la pratique (plus de chirurgies recommandées), il y a preuve d'induction.

- La seconde distinction importante porte sur la différence entre demande et utilisation. Ce point est saillant lorsque l'on s'intéresse aux questions de partage des coûts et/ou aux systèmes de *managed care*. Dans ce cadre, un médecin peut influencer l'utilisation des soins sans pour autant influencer la demande de soins. Par exemple, un patient traité en HMO peut recevoir moins de traitement que le niveau optimal de consommation. Cela peut être

interprété comme une induction à la baisse de la demande. Mais, il est également possible de voir dans cette situation une preuve de rationnement. Si tel est le cas, c'est l'utilisation des services de soins qui a été affectée. Il n'y a pas induction de la demande au sens où la fonction liant le prix et la quantité désirée est inchangée. Ce qui change est l'utilisation des soins – par l'intermédiaire de l'organisme rationnant les soins.

Selon le texte fondateur d'Evans (1974), le médecin maximise sa fonction d'utilité, celle-ci étant dépendante du revenu et du pouvoir d'induction. Graphiquement, l'induction implique un déplacement de la courbe de demande. N'importe quel vendeur a intérêt à une demande plus élevée et un médecin rationnel cherchent à maximiser son revenu induirait sa demande infiniment. Il devient alors nécessaire d'introduire une limite à l'induction. Afin de limiter la possibilité d'induction – et d'éviter le cas absurde d'une extension infinie de l'induction et des revenus –, les économistes considèrent que l'induction a un coût. Il y aurait une désutilité de l'exercice du pouvoir d'induction.

Selon Stano (1987), il est possible de faire une analogie entre la publicité et le pouvoir d'induction. L'induction et la publicité ont un coût et les dépenses d'induction et de publicité sont limitées par leurs rendements décroissants. Dans le cas du médecin, l'induction a un coût soit parce qu'elle est déplaisante psychologiquement[19] (Evans, 1974, Fuchs, 1978, etc.) soit parce qu'elle est une contrainte à la maximisation du profit (Dranove, 1988, Stano, 1987). Cette présentation correspond bien à la définition de McGuire (2000) : seules les influences sur la demande qui retirent le patient du point de consommation optimal imposent des coûts psychiques au patient (Encadré 1.4.).

[19] Nous reviendrons sur cette possibilité-là dans le chapitre 2.

Encadré 1.4. L'hypothèse de demande induite. Le modèle de McGuire et Pauly (1991)

Le modèle de McGuire et Pauly (1991) cherche à présenter simplement la possibilité pour les médecins d'induire la demande des patients dans un contexte à deux secteurs d'activité. L'utilité du médecin dépend de son revenu Y et de l'induction I.

$$U = U(Y, I)$$

Avec

$$Y = N(m_1 x_1(i_1) + m_2 x_2(i_2))$$
$$I = N(i_1 + i_2)$$

Et

$$U_Y > 0; U_I < 0; U_{YY} < 0; U_{II} < 0$$

Le médecin voit N patients qui utilisent les services 1 et 2. La quantité de chaque service est donnée par xi et est affectée par le niveau d'induction (ii). On a :

$$i.x' > 0; x'' < 0$$

m est la marge pour chaque service égale à la différence entre l'honoraire du médecin et le coût du service. Le médecin choisit i1 et i2 qui maximise son utilité. Les conditions du premier ordre donnent :

$$m_1 x_1' = m_2 x_2' = -\frac{U_I}{U_Y}$$

Le rendement marginal de l'induction pour chaque service (deux premiers termes de l'équation) doit être égalisé au coût marginal de l'induction (dernier terme de l'équation). Cette équation permet alors d'interpréter la modification du nombre de patients par médecin (N) comme un résultat (a) d'une augmentation du nombre de médecins ou (b) de la modification des honoraires.

Une baisse de N consécutive à l'augmentation du nombre de médecins affecte uniquement le troisième terme de l'équation : $-\frac{U_I}{U_Y}$. La baisse de N implique une diminution globale de l'induction et des revenus – impliquant une réduction de la désutilité marginale de l'induction (UI) et une augmentation de l'utilité marginale des revenus (UY), d'où la réduction de $-\frac{U_I}{U_Y}$. Dès lors, pour assurer l'égalité de l'équation donnant les conditions du premier ordre, le montant d'induction dans chaque service doit s'accroitre.

Si l'on imagine que l'assureur réduit le prix du service m1 – le prix du second service restant inchangé, il y a alors un effet revenu et un effet substitution. Effet revenu : $-\frac{U_I}{U_Y}$ va décroitre à cause de la hausse de UY.

Cela va engendrer une induction pour les deux services 1 et 2. Il y a également un effet substitution qui vient à cause de la baisse du prix dans le secteur 1. Cet effet tend à réduire l'induction dans le secteur 1 et à croitre l'induction dans le secteur 2 pour rétablir l'équilibre. Dans le secteur 2, l'induction augmente et, dans le secteur 1, l'induction dépend de la valeur de l'effet substitution par rapport à l'effet revenu.

L'hypothèse de demande induite ne fait pas que renouveler la discussion autour de la politique de santé. Elle permet aussi de donner un cadre théorique à la loi de Roemer présentée pour la première fois dans les années 1960. D'après cette loi empirique, il existe une corrélation entre la densité de lits d'hôpitaux et le taux de séjours à l'hôpital. Selon la formule de l'auteur, un lit construit est un lit rempli. Etendue à l'ensemble des services médicaux, cette loi s'apparente à l'hypothèse de demande induite. Au lieu de voir l'augmentation du nombre d'offreurs faire pression à la baisse sur les prix, on observe que, lorsqu'un nouveau médecin vient augmenter la densité médicale, les prix augmentent. Il est alors tentant d'expliquer ce phénomène par l'hypothèse de demande induite.

Cependant, il n'existe pas de consensus sur l'existence et l'importance de la demande induite. Dans leur récente revue de la littérature, Léonard et al. (2009) mettent en évidence les

limites empiriques qui empêchent de conclure sur l'existence ou non du phénomène d'induction. En dépit de l'hétérogénéité des études (stratégies de recherche, variable dépendante, mesure de l'utilisation des services médicaux, etc.), il existe une corrélation positive entre la densité médicale et la consommation des soins dans la majorité des textes étudiés. Ces confirmations de la loi de Roemer ne sont pas pour autant des éléments de preuve de l'existence d'une demande induite. Plusieurs limites empiriques empêchent de conclure sur ce point. D'une part, les *outcomes* étudiés ont varié considérablement du point de vue de l'état de santé des patients, de la qualité des soins et de l'impact économique. D'autre part, il est difficile de démêler l'induction des médecins de celle des patients. Enfin, lorsqu'elle est détectée la demande induite est souvent de faible magnitude.

Le problème principal est de définir le niveau de consommation optimal. En quoi est-il possible de dire que telle ou telle hausse de la consommation médicale est injustifiée ? Malgré l'absence de preuves empiriques de ce phénomène, il semble, en observant les travaux théoriques des économistes de la santé, que l'existence théorique de l'induction soit suffisante pour en déduire l'existence réelle. En effet, tous les modes de rémunération traditionnels du médecin sont examinés à l'aune du pouvoir discrétionnaire du médecin.

2.3. Les modes de rémunération traditionnels du médecin marqués d'inefficacité

La TSE1, en mettant l'accent sur les comportements potentiellement opportunistes des acteurs, trace une ligne directrice pour la politique publique. Il faut maîtriser les coûts et la qualité des soins en limitant l'opportunisme médical. Il faut alors étudier les incitations fournies par les différents modes de rémunération traditionnels[20] pour, le cas échant, les substituer par d'autres. Trois grands modes de rémunération du médecin existent : le salariat, la capitation et le paiement à l'acte. Ceux-ci ont tous fait l'objet d'un examen théorique au regard de l'analyse standard (par exemple, Albouy et Déprez, 2009 ; Grignon et al. 2002). Mais, d'après les canons de la théorie du choix rationnel, aucun d'entre eux ne permet de maximiser le bien-être collectif (Tableau 1.3.). Ils fournissent tous des incitations perverses qui ne permettent pas d'atteindre ni la situation de marché ni le *second best* – qui reste à définir.

[20] La littérature parle de modes de rémunération « traditionnels » car ce sont ceux utilisés par la plupart des systèmes de santé. La « tradition » n'est pas théorique mais empirique : c'est ce qui existe dans le monde réel.

Tableau 1.3. L'évaluation des modes de rémunération traditionnels par la TSE1

	Salariat	Capitation	Paiement à l'acte
Exemple de système de santé	France (hôpital), Finlande, Danemark, Royaume-Uni (hôpital)	Royaume-Uni (soins de ville)	France (soins de ville), États-Unis (soins de ville)
Résolution de l'incertitude	Le médecin est incité à participé à la relation de soin, pas le patient et pas l'assureur	Le médecin est incité à participé à la relation de soin, pas le patient et pas l'assureur	Le médecin est incité à participé à la relation de soin, pas le patient et pas l'assureur
Résolution de la demande induite	Le médecin n'a pas intérêt à induire la demande	Le médecin n'a pas intérêt à induire la demande	Maximise le potentiel d'induction de la demande
Effet sur la qualité des soins	Faible incitation à répondre à la demande des patients, faible incitation à la qualité, pas d'incitation à la productivité	Incitation à répondre à la demande des patients, faible incitation à la qualité, pas d'incitation à la productivité, incitation à la prévention	Incitation à répondre à la demande des patients, incitation à améliorer la productivité
Effet sur le coût de soins	Pas d'incitation à la production, maîtrise du niveau de dépense, système moins coûteux	Faible incitation à la production, maîtrise du niveau de dépense, système moins coûteux	Risque de surproduction, risque de demande induite, nomadisme médical niveau de dépense connu *ex post*

2.3.1. Le salariat

Le salaire est un revenu fixe versé à échéance régulière basé sur une durée de travail elle aussi fixe. Les conditions de travail et de rémunération sont déterminées au moment de la signature du contrat et sont indépendantes de l'intensité de l'effort ou des résultats de l'activité pendant le temps de travail. Il y a une totale déconnexion de la rémunération et de l'acte médical. Ce mode de rémunération est bien connu des économistes *mainstream* qui n'ont de cesse de présenter ses effets désincitatif : il n'inciterait pas à faire d'effort mais à « tirer au flanc » (Cordonnier, 2000). Dans un contexte d'asymétrie d'information, il est impossible d'observer l'effort du médecin et donc de sanctionner le comportement opportuniste. Le salariat est alors une anomalie profonde et il est difficile de comprendre la rationalité de l'entreprise qui met en place de type de rémunération. Etant sûr de percevoir son salaire quel que soit son niveau d'effort, le médecin exercera toujours un effort nul. C'est théoriquement le mode de paiement traditionnel le moins incitatif. Il n'y a aucune valorisation de l'effort puisque rien ne vient récompenser ou sanctionner le comportement du médecin.

Néanmoins, en dehors de l'économie de la santé, le salaire peut inciter à l'effort s'il constitue une rente (Milgrom et Robert, 1997). Par exemple, dans le modèle à salaire d'efficience de Shapiro et Stigliz (1984), les entreprises offrent un salaire supérieur au salaire d'équilibre afin que les salariés aient intérêt à produire l'effort maximal. Au niveau agrégé, l'élévation du salaire au-dessus du niveau optimal produit un chômage involontaire – résultant logiquement du déséquilibre entre offre et demande de travail. Ce salaire est un salaire d'efficience dans la mesure où il créé une menace crédible pour les salariés : s'ils ne travaillent pas, ils risquent de perdre leur rente (part du salaire au-dessus du salaire d'équilibre), et même de devenir chômeur. Cette littérature a ainsi mis en évidence de multiples raisons pour lesquelles le salaire, sous certaines conditions, peut être incitatif.

Par ailleurs, la littérature en économie de la santé accorde un avantage majeur au salariat médical. Dans la mesure où le salaire est un paiement prospectif – l'entreprise connait le montant dû avant le début de l'exercice –, il est possible pour la tutelle d'anticiper et de maîtriser ses coûts. Il suffit pour cela d'anticiper les mouvements de main d'œuvre (retraites, départs volontaires, promotions, mutations, etc.) et l'évolution de la masse salariale (augmentations annuelles, promotions, avancement dans les grilles de rémunération, etc.). Cet avantage est loin d'être négligeable surtout lorsque l'assureur se trouve dans une situation où il cherche à maîtriser ses dépenses. Mieux vaut dans ce cas pouvoir anticiper le volume de rémunération dans un cadre salarial que de devoir attendre la fin de l'exercice pour observer et rémunérer l'activité effectivement réalisée. Il est plus facile d'encadrer la masse salariale par la négociation collective autour des augmentations salariales que d'être dépendant du niveau d'activité agrégé décidé individuellement par chaque médecin – ce point étant d'autant plus important que l'induction de la demande est possible.

En dépit de ces avantages de contrôle des dépenses, le salariat médical n'est pas pour la TSE1 une option viable dans la mesure où il est associé à une logique de moindre effort, désastreuse pour la qualité des soins. Dans cette perspective, chaque euro dépensé dans le salaire est un euro de trop car il n'y a aucune incitation à l'effort. L'opportunisme systématique du médecin est un frein à la solution salariale. Au-delà de la question de l'opportunisme, il faut également noter la dimension exclusivement instrumentale du travail dans la TSE. Le seul intérêt du travail est sa rémunération. Ici, les hypothèses d'intéressement et de désutilité de l'effort – notamment au fondement des modèles d'arbitrage travail-loisir – sont poussées à leurs conclusions les plus radicales : la tâche n'a jamais d'intérêt en soi si bien

que sans menace de sanction financière, le médecin préférera toujours laisser le patient souffrir que de le soigner.

2.3.2. La capitation

Avec la capitation, le médecin perçoit un forfait par patient inscrit à son cabinet quelles que soient la quantité et la qualité des soins délivrés (de nulle à infini). Par exemple, un médecin disposant d'une patientèle de 400 patients et ayant une rémunération de 20 euros par mois et par patients, obtiendra une rémunération mensuelle à la capitation de 8 000 €. Qu'il ait réalisé 100 ou 1 000 consultations au cours du mois écoulé ne change rien à sa rémunération. Le montant du forfait peut évoluer selon les caractéristiques du patient (âge, sexe, milieu socio-économique, état de santé, etc.). Tout comme le salariat, la rémunération est indépendante de l'intensité ou des résultats de l'activité médicale. Cependant, ce mode de rémunération incite les médecins à maximiser le nombre de patients inscrits sur leur liste (dans la limite éventuelle d'un plafond). Cela peut créer une concurrence entre les médecins pour attirer une plus grande patientèle cette concurrence ayant des effets positifs. En effet, pour attirer la patientèle, les médecins sont incités à produire des soins de plus grande qualité sans quoi les patients votent avec leurs pieds. En outre, si les médecins veulent réduire leur effort à long terme, il y a une incitation à produire des soins préventifs qui vont réduire le retour des patients. Cette (faible) incitation n'existe pas dans le cas du salariat car rien n'indique que les patients se montreront fidèles – alors que dans le cas de la capitation, les patients n'ont souvent pas le choix.

D'après la TSE1, ce mode de rémunération demeure problématique dans la mesure où il n'y a pas de lien entre le nombre d'actes par patients et revenus. Une fois le patient inscrit sur la liste du médecin, la rémunération du médecin étant déconnectée du nombre de visites, il y a une incitation à produire peu de soins (dans l'exemple donné plus haut, mieux vaut produire 100 actes que 1 000 à nombre de patients donné). Par ailleurs, afin de conserver sa patientèle dans une logique de concurrence, le médecin peut également être incité à satisfaire les demandes des patients, même si elles sont déraisonnables. Il peut alors y avoir un risque de surconsommation au détriment de l'assureur.

De plus, si le médecin est responsable d'un budget global (comme au Royaume-Uni), il est au contraire incité à rationner le soin (au détriment des patients). Dans ce cas de figure, le médecin doit gérer un budget au mieux et est autorisé à conserver le résidu. Il y alors un conflit entre la rémunération du médecin et la qualité délivrée au patient. Ce conflit peut se

traduire par un phénomène de sélection adverse : pour maximiser son résidu, le médecin peut choisir d'écarter les patients les plus malades (dont les soins sont plus réguliers et plus coûteux). Enfin, il faut noter que, lorsque le médecin à un rôle de *gate keeper*[21], on observe un phénomène de report excessif vers d'autres structures médicales (spécialistes et hôpitaux). Le médecin étant payé *per capita* quoi qu'il arrive, il est incité à faire peu d'effort et donc à reporter sa charge de travail (le patient) vers ses confrères.

Au final, le paiement à la capitation peut avoir théoriquement des effets ambigus (positifs et négatifs) sur la qualité des soins. Comme dans le cadre du salariat, il permet une meilleure maîtrise des dépenses car il limite l'induction de la demande et il repose sur un paiement prospectif. Mais, les effets en termes de qualité des soins restent assez néfastes.

2.3.3. Le paiement à l'acte

Le paiement à l'acte est un forfait payé à chaque réalisation d'un acte médical. La rémunération dépend du nombre d'acte et du prix des actes. Ces derniers peuvent être fixés (comme en France avec le secteur 1) ou libre (secteur 2). De façon assez tautologique, les économistes *mainstream* insistent sur l'effet majeur du paiement à l'acte : il incite à produire des actes – il est donc inflationniste. L'effet sur les coûts est très important puisque le médecin est incité, d'une part, à améliorer sa productivité horaire (nombre d'actes par heure) et, d'autre part, à accroitre son temps de travail (nombre d'heures travaillées dans la journée). C'est le mode de rémunération le plus propice au comportement de demande induite. Il y a alors un risque de surproduction de soins coûteux pour la collectivité. La dépense est cette fois-ci rétrospective (le niveau des dépenses est connu *ex post*), ce qui signifie que la tutelle n'a pas de prise sur l'évolution de son budget.

L'augmentation de la productivité horaire du médecin peut avoir pour conséquence une diminution du temps de consultation préjudiciable au patient. Si l'on retient l'hypothèse de corrélation positive entre durée de consultation et qualité des soins, l'augmentation de la productivité horaire implique nécessairement une baisse de la qualité des soins. Mais, dans le cadre du paiement à l'acte, le patient a la possibilité de consulter plusieurs médecins pour la même pathologie (nomadisme médical). Si l'effet en termes de qualité peut être positif grâce à la dimension concurrentielle du mécanisme, cela implique pourtant des coûts supplémentaires

[21] Le rôle du *gate keepper* est, en première ligne du système de soin, de contrôler l'accès aux structures en aval souvent plus coûteuses.

pour la tutelle. Le médecin garde un intérêt important à la satisfaction de la demande de ses patients. Dans le cas contraire, il risque de voir sa patientèle s'exiler vers la concurrence.

Il ressort de l'analyse du paiement à l'acte que celui-ci incite à produire des soins de meilleure qualité – bien que l'asymétrie d'information implique que la qualité est nécessairement sous-optimale. Cependant, cet effet positif ne peut pas masquer les effets néfastes sur les dépenses. Il s'agit du mode de rémunération le plus coûteux car structurellement inflationniste.

2.4. Un peu plus près des étoiles : le P4P médicale

Le résultat majeur de la discussion théorique autour des modes de rémunération traditionnels est que ceux-ci sont incapables de maximiser le bien-être collectif. Le médecin reste toujours en position d'utiliser son pouvoir discrétionnaire au détriment de l'assureur et du patient. Schématiquement, soit le médecin est incité à dilapider les ressources (paiement à l'acte) soit il est incité à produire une faible qualité de soin (capitation et salariat). Pour comprendre ces résultats et avancer dans la recherche du mode de rémunération optimal du médecin, il faut revenir aux fondements de la TS. L'approche walrasienne montre qu'à l'équilibre général de concurrence pure et parfaite les facteurs de production sont rémunérés à leur productivité marginale. Ce résultat repose sur deux hypothèses centrales que sont la désutilité d'effort et l'intéressement : les individus ne veulent pas travailler, ce pourquoi il faut les intéresser.

Alors que dans la TS, le mécanisme concurrentiel permet d'éviter la sous-rémunération et la sur-rémunération des facteurs de production, dans un univers où la répartition de l'information est asymétrique, il est impossible d'observer la productivité marginale des individus et de rémunérer l'effort. L'employeur ne pouvant observer la productivité de son salarié sera incité à le rémunérer en dessous de sa productivité marginale et le salarié, sachant cela, est incité à tirer au flanc. On se retrouve dans une situation de dilemme du prisonnier où, chacun cherchant à tirer la couverture à lui, l'équilibre qui va s'imposer conduit à un salaire faible et un effort faible, alors qu'il existe une autre situation pareto-dominante (Eber, 2006, Liebenstein, 1982).

Il faut alors trouver un mécanisme permettant d'inciter à la coopération. L'une des solutions possibles pour échapper à cette situation est d'inciter le salarié en le rémunérant en fonction de sa production. Plutôt que d'observer directement l'effort du salarié, la mise en

place d'indicateurs contrôlables permet de le rémunérer au plus près de sa productivité marginale. Dans ce cadre, le mode de paiement optimal est le P4P. Selon cette conception, il est nécessaire de lier la performance observée à la rémunération. Le risque pesant sur le niveau de sa rémunération incite l'individu à l'effort. L'idée générale est de faire peser un risque financier sur le médecin pour qu'il agisse de façon attendue par la tutelle. Ce risque doit interdire (ou diminuer) le recours au pouvoir d'induction coûteux pour la tutelle tout en incitant à l'effort nécessaire à une prise en charge de qualité. Il y a un partage du risque entre l'employeur et l'employé permettant de maximiser le bien-être collectif. Le P4P prend le relai de la concurrence.

Ce résultat est l'un des plus importants de l'économie du travail (Lazear, 2000a, 2000b, 2007). Au-delà de l'économie de la santé, tout ce qui ne correspond pas à la rémunération à la productivité marginale est une rigidité sur le marché du travail qu'il convient de supprimer. Il s'agit encore là de la stratégie du Nirvana : l'atteinte de l'optimum est contrainte par des barrières institutionnelles (salaire minimum, temps de travail, congés payés, protection sociale, etc.) ou informationnelles (incertitude, asymétrie d'information, etc.). En conséquence, il convient soit de détruire les barrières lorsque cela est possible soit de mettre en place le P4P de façon à inciter les individus à atteindre une productivité proche de celle de concurrence pure et parfaite. Il faut se rapprocher du Nirvana.

Cette démonstration vaut pour l'offreur de travail qu'est le médecin. Puisqu'il n'est pas possible de vérifier son effort en raison d'asymétries d'information importantes, il est nécessaire de mettre en place un contrat incitatif. Pour cela, il faut créer des indicateurs censés mesurer la performance du médecin. Sa rémunération sera alors dépendante du taux de réalisation des indicateurs définissant sa performance. Cette stratégie ne constitue pas une stratégie propre au monde médical. Le médecin n'est pas un acteur singulier au sein d'un secteur d'activité singulier.

Bien que l'idéal d'un paiement totalement corrélé à la performance des acteurs soit au premier abord séduisant, celui-ci est remis en cause par l'existence de barrières puissantes. Dans le cadre plus général de l'offre de travail, certains obstacles ne permettent pas la mise en place du P4P. Par exemple, Azariadis (1975) a montré que, dans certaines situations, le paiement d'une partie de la rémunération sous forme de forfait peut être une stratégie optimale à la fois pour l'entreprise et le travailleur. En effet, lorsque l'agent a une aversion au risque supérieur à celle du principal, le versement d'une rémunération fixe permet au travailleur de s'assurer contre les aléas de la conjoncture et à l'entreprise de verser un forfait

inférieur à la rémunération d'équilibre. Mais, pour inciter l'agent à l'effort, une partie de sa rémunération doit rester dépendante d'indicateurs de productivité. Ainsi, faire reposer un risque financier trop important sur le salarié peut être préjudiciable.

Dans le cas de la rémunération du médecin, il est également possible de mettre en évidence les mêmes obstacles à un P4P intégral. L'analyse économique standard souligne alors l'utilité d'un paiement mixte. Il s'agit de mélanger les différents modes de paiement afin de combiner les effets positifs des multiples formes de rémunération. La littérature a démontré l'intérêt de la capitation dans la mesure où ce mode de rémunération engendre un partage des risques entre le médecin et l'assureur permettant de vérifier la contrainte de participation du premier. Dans ce contexte, il peut être optimal pour la tutelle de mettre en place une rémunération mixte, c'est-à-dire au moins deux types de rémunération avec un degré de forfaitisation différent (Baker, 1992, Ellis et McGuire, 1990, Ma, 1994, Newhouse, 1996, 2002, Ma et McGuire, 1997, Pauly, 2000). Plus précisément, il serait optimal de combiner une rémunération fixe à la capitation et une rémunération à la performance (Encadré 1.5.).

Encadré 1.5. Case mix, capitation et P4P. Le modèle de Baker (1992)

Le modèle de Baker (1992) s'interroge sur la façon optimale dont un assureur (principal) devrait contracter avec un médecin (agent). Par hypothèse, l'agent est mieux informé et toutes les dimensions de la performance ne sont pas contractualisables. Le bénéfice que le patient retire du traitement dépend de la qualité de l'effort du producteur, e. L'assureur veut maximiser la qualité des soins minorée du coût payé au producteur. Cet objectif est imparfaitement corrélé au coût des soins, m.

$$m = \mu e$$

Et,

$$\mu \sim N(1, \sigma^2)$$

Où σ^2 mesure la variation du case-mix non capturé par l'ajustement au risque.

L'acheteur paye le producteur un montant fixe R par patient plus une fraction r, $0 \leq r \leq 1$, des coûts. Le taux de remboursement définie le partage des coûts, s, avec :

$$s = 1 - r$$

Le producteur reçoit alors R + rm et subit une désutilté de l'effort c(e). L'effort choisi à l'équilibre est :

$$e^* = \frac{r\mu}{c}$$

L'effort est croissant du remboursement et décroissant de la désutilité marginale de l'effort. Il dépend aussi de la façon dont l'effort impacte la performance μ. Le choix optimal de remboursement revient à maximiser :

$$r - \frac{r^2(1+\sigma^2)}{2}$$

A l'optimum on trouve :

$$r^* = \frac{1}{1+\sigma^2} < 1$$

si la variation du case-mix est supérieure à 0. Le payeur ne doit pas rembourser les coûts dollar par dollar mais seulement une fraction. Lorsque $\sigma^2 > 0$, le taux de P4P dans la rémunération totale doit être inférieur à 1. Il doit y avoir un partage des coûts ($0 < s^* < 1$) : le paiement doit être mixte.

L'assureur reconnait ainsi que les coûts (m) et la variabilité des cas (σ^2) sont une mesure imparfaite de la performance. Les deux objectifs de l'amélioration de la qualité et du contrôle des coûts appellent deux instruments : un bonus lié à la qualité et une rémunération fixe.

La littérature standard présente un autre motif pour lequel un mode de rémunération mixte serait plus efficace : le P4P intégral est sous-optimal parce que ce mode de rétribution encourage à la sélection des patients (Eggleston, 2000). Le médecin est incité à choisir les patients qui lui permettent de réaliser plus facilement ses objectifs. En excluant les patients les moins utiles à la réalisation des objectifs, le médecin maximise ses chances d'être performant et ainsi sa rémunération. Cela pose un problème d'égalité d'accès aux soins incompatible avec la recherche du bien-être collectif. Une solution est alors de coupler le P4P avec un forfait par patient. Ce paiement à la capitation permet de modifier le comportement du médecin vis-à-vis des patients les moins utiles à la réalisation de ses objectifs : plutôt que de ne pas les soigner, il a intérêt à les prendre en charge afin de percevoir la capitation.

Un paiement mixte, avec capitation partielle, présente également des avantages lorsque l'on considère l'activité médicale comme une activité multitâche (Eggleston, 2005). Deux problèmes se posent pour rémunérer à la performance une activité multitâche. D'abord, il faut

autant d'indicateurs que de tâches différentes. Dans le cas contraire l'individu est incité à délaisser les tâches non rémunérées. Ensuite, il faut que chaque tâche puisse faire l'objet d'une mesure précise et qu'elle soit contractualisable. Si ce n'est pas possible, si une tâche est difficilement mesurable, alors l'indicateur sera déficient et l'incitation mal dirigée. Dans le cas de la production de soins ces deux conditions ne sont pas réunies. Les indicateurs du P4P étant imparfaits, il devient optimal d'introduire une part de capitation. Cette forme de paiement mixte possède l'avantage d'équilibrer les incitations à la qualité contractualisables et les incitations à la qualité non contractualisables.

Plus généralement, pour la TSE1, il est possible de démontrer « clairement » la supériorité théorique d'un paiement mixte à la fois sur le paiement à l'acte et sur la capitation (Levaggi et Rochaix, 2003). Le mode de paiement optimal est alors un mélange entre capitation et P4P.

Conclusion

L'idée fondamentale qui a guidé la rédaction de ce premier chapitre est qu'en identifiant l'originalité du secteur des soins médicaux, il serait possible de comprendre la forme particulière de rémunération et de régulation de la médecine. Fondée sur l'intuition que le champ de la santé porte en lui quelque chose de particulier relativement aux autres domaines économiques (automobile, immobilier, foncier, etc.), nous avons vu que la singularité des soins médicaux a été reconnue très rapidement par les auteurs pionniers de la discipline. C'est le cas notamment avec l'article majeur d'Arrow (1963). Les défaillances de marché appellent un renouvellement paradigmatique… donc un renouvellement des outils de régulation de la médecine.

Cependant, ce renouvellement n'a jamais eu lieu. Tout au mieux la théorie standard a été reformulée en théorie standard étendue. Paradoxalement, alors qu'en levant les hypothèses de la TS, celle-ci perd toute utilité, le *mainstream* s'est reconstruit en… généralisant la fiction de marché de concurrence pure et parfaite.

La théorie standard reste un idéal normatif qui désormais dépasse le cadre délimité du marché. C'est la stratégie du Nirvana, critiquée dès 1969 par Demsetz.

Tout se passe comme s'il s'agissait d'atteindre la lune à bord d'une caravelle. Les valeureux explorateurs ont beau jeu de voguer de plus en plus vite, de naviguer à vent

contraire et de braver les tempêtes, ils n'atteindront jamais la lune ! D'ailleurs, et c'est là tout le problème, la différence entre nos marins et les économistes de la TSE est que si la lune existe, le marché de concurrence pure et parfaite n'existe pas. Comme l'a montré Demsetz, l'une des principales raisons à cela tient au fait qu'il n'existe pas d'arrangements institutionnels susceptibles d'éradiquer certains types de défaillance de marché (asymétrie d'information, incertitude, etc.).

Pourtant, la théorie économique s'est développée et a produit de nouveaux résultats. En vertu de la stratégie Nirvana – substituant le pilier de la coordination marchande par le pilier de la normativité marchande –, tous ces résultats ne peuvent être qu'une déviation par rapport au *first best* : la rémunération du médecin à la productivité marginale du travail. La réalité doit être analysée à partir de ce critère normatif. Partis de modes de rémunération traditionnels sous-optimaux, la théorie standard dispose maintenant d'une théorie de la rémunération du médecin. Le salariat, la capitation et le paiement à l'acte sont tous les trois inefficaces car ils engendrent de mauvaises incitations. Le médecin dispose d'un avantage informationnel qu'il est susceptible d'exploiter et son pouvoir d'induction doit être contrôlé. En conséquence, la rémunération de *second best* doit combiner une part fixe (capitation) – visant à s'assurer de la participation des médecins – et une part variable, le P4P – destiné à inciter à l'effort :

$$R_s = C + \sum p_i t_i$$

Où R_s est la rémunération de *second best*, C est la part à la capitation et $p_i t_i$ le P4P (calculé comme la somme des produits pour chaque indicateur i du taux de réalisation t_i et du prix associé à cet indicateur p_i). Si le paiement à la capitation apparaît comme une contrainte illégitime au regard de la situation de concurrence pure et parfaite (il s'agit d'une rémunération liée à la variation des cas et à l'aversion du risque des médecins), le P4P est lui une approximation rapprochant le monde réel de la productivité marginale. En ce sens, et en suivant en cela les travaux d'économie du travail, le médecin est un travailleur comme un autre et il doit être rémunéré comme les autres. Puisqu'il n'y a plus grand-chose de spécifique au secteur des soins médicaux, il n'y a aucune raison pour que le médecin soit rémunéré différemment.

Chapitre 2

La performance en univers altruiste. Paradoxes de théorie économique et problèmes de politique économique

Introduction

En dépit de la reconnaissance d'une particularité des services médicaux, le premier chapitre a mis en évidence le fait que le passage de la Théorie Standard (TS) à la Théorie Standard Etendue (TSE) en économie de la santé n'a pas modifié substantiellement les recommandations de politiques économiques à destination du médecin. Celui-ci restant au final un acteur économique comme un autre, il faut le rémunérer comme les autres. Le paiement à la performance n'est alors qu'un pis-aller rendu obligatoire par la difficulté d'atteindre la situation théorique de marché de concurrence pure et parfaite.

Cette approche repose néanmoins sur une vision d'un médecin maximisateur de profit ; vision assez lointaine de l'image profane que l'on se fait du monde médical. En effet, dans le champ des soins médicaux, les considérations éthiques sont omniprésentes. Chaque profession médicale a une morale professionnelle, par exemple fondée sur un code de déontologie. Aux États-Unis, l'*American Medical Association* stipule dans son *Code of Medical Ethics* la conduite à tenir par chaque médecin. L'*European Council of Medical Orders* a pour objet de promouvoir au sein de l'Union Européenne une médecine de qualité respectueuse des besoins des patients. Ces « Conseils de l'Ordre » existent non seulement pour réguler la pratique des médecins mais aussi celles des infirmiers, des pharmaciens ou encore des sages-femmes.

Chez les médecins, le Serment d'Hippocrate est un engagement à respecter une orientation éthique. Les valeurs affichées dans l'éthique médicale, telles que la bienfaisance, la non-malfaisance, la justice et le respect de l'autonomie, induisent des responsabilités spéciales pour les professionnels de santé (Gillon, 1994). Il y a également une forte croyance auprès des patients du fait que les médecins usent de ces considérations éthiques dans leur travail quotidien (Glannon et Ross, 2002). Si tous les travailleurs peuvent être dotés d'une

éthique professionnelle, le médecin apparaît comme un cas exemplaire de telles dispositions. L'organisation de la médecine en profession n'est pas étrangère à cette exemplarité : les grandes professions jouissent de capacité d'autorégulation et d'autocontrôle en échange de leur engagement éthique (Hassenfeutel, 1997, Jaunait, 2004).

La question qui se pose aux économistes est alors de savoir s'il faut prendre au sérieux l'existence d'une éthique (ou d'une morale, d'un altruisme, etc.[22]) des médecins ? Ou bien, comme tend à le supposer la TSE1, l'« *homo medicus* » est-il un *homo economicus* comme un autre ?

La réponse à cette question est d'une importance capitale d'un point de vue purement théorique mais aussi parce qu'elle a potentiellement un impact sur la façon dont il convient de rémunérer le médecin. Un médecin suivant un code éthique répond-t-il de la même façon aux incitations qu'un médecin uniquement maximisateur de profit ? Si la réponse est négative, l'éthique médicale est alors en mesure de remettre en cause la recommandation de politique publique visant à payer les médecins par un mixte de capitation et de paiement à la performance. La TSE1 vacille : si le médecin agit de façon altruiste, il n'est pas nécessaire de le rémunérer pour qu'il délaisse son égoïsme. Si le médecin se soucie de son patient, il n'est pas nécessaire de le payer pour qu'il s'affranchisse de la poursuite de son seul intérêt personnel. Si l'altruisme médical désigne la capacité du médecin à agir avec empathie pour son patient, il est déjà performant. Pourquoi l'inciter alors à la performance et de quelle performance parle-t-on si ce n'est pas celle qui consisterait à agir pour le seul bien-être du malade ?

La théorie économique n'est pas restée sourde à ces objections et elle a pris en compte l'existence d'une éthique médicale. Cette introduction de dispositions sociales ou éthiques dans le calcul des agents marque une évolution du *mainstream*. En s'intéressant désormais aux limites de l'optimisation (intéressement limité et rationalité limitée), la TSE1 mue en TSE2. En ce sens, elle conserve la norme marchande comme idéal indépassable tout en intégrant les « anomalies » de comportement du médecin : la TSE2 poursuit la stratégie du Nirvana.

L'intégration de la composante éthique de la médecine dans l'économie de la santé peut être considéré comme emblématique de la TSE 2. Cette éthique médicale est d'ailleurs régulièrement mentionnée pour justifier l'originalité du bien santé. De ce point de vue, Arrow

[22] Les termes sont utilisés ici sans les distinguer, ce que nous ferons plus bas.

(1963), dans son article fondateur de l'économie de la santé, a souligné l'importance de l'éthique médicale pour qui veut étudier les médecins[23] :

> *« It is clear from everyday observation that the behavior expected of sellers of medical care is different from that of business men in general. These expectations are relevant because medical care belongs to the category of commodities for which the product and the activity of production are identical. In all such cases, the customer cannot test the product before consuming it, and there is an element of trust in the relation. But the ethically understood restrictions on the activities of a physician are much more severe than on those of, say, a barber. His behavior is supposed to be governed by a concern for the customer's welfare which would not be expected of a salesman. In Talcott Parsons's terms, there is a "collectivity-orientation," which distinguishes medicine and other professions from business, where self-interest on the part of participants is the accepted norm » (Arrow, 1963, p. 949).*

Etant donné qu'elle est un facteur de coordination essentiel, l'éthique médicale ne peut être exclue de l'analyse économique. Le problème est que la façon dont l'éthique professionnelle est introduite dans les frontières de la rationalité économique n'est pas claire. La solution proposée par la TSE2 est d'enfermer l'éthique professionnelle dans le calcul d'optimisation. L'attitude éthique du médecin est alors formalisée en termes d'altruisme médical. Ce concept est censé prendre en compte les valeurs du médecin avec l'hypothèse que le paradigme de l'*homo economicus* est un point de départ suffisant (même s'il doit être modifié). Les valeurs sont incorporées par l'intermédiaire du concept d'utilité. Les individus sont dotés de préférences « classiques » (dans un contexte d'arbitrage travail/loisir par exemple) auxquelles sont ajoutées des préférences « sociales ». Suivant en cela la définition traditionnelle de l'altruisme, travaillée notamment par Harsanyi (1955) et Becker (1981), l'altruisme médical est caractérisé par l'intégration du bien-être du patient dans la fonction d'utilité du médecin. L'altruisme est défini comme l'interdépendance des fonctions d'utilité.

Cette représentation de l'altruisme consiste à internaliser la fonction d'utilité du patient (ou une variable *proxy* comme le revenu ou l'état de santé) dans celle du médecin. Un médecin altruiste reste un maximisateur d'utilité, cette dernière incluant désormais le bien-être du patient. Cette définition présente deux sérieux problèmes que nous développons dans ce chapitre.

Le premier problème est que la TSE2 ne propose pas un modèle d'altruisme authentique parce qu'elle réduit l'altérité à une source d'utilité comme une autre. Cet altruisme est instrumental : il ne prend pas en compte les actions faites dans l'intention de bénéficier aux autres indépendamment de sa propre utilité (Caillé, 2005, Chanial, 2008). À l'inverse,

[23] On se reportera au chapitre 1 pour une discussion de l'apport d'Arrow à l'économie de la santé.

l'altruisme authentique, ignoré par la TSE2, est plus proche de la définition commune ou ordinaire de l'altruisme. Est altruiste l'action qui est entreprise sans que l'utilité reçue par son auteur soit en jeu. Cela ne veut pas dire qu'en agissant de manière altruiste il soit impossible d'augmenter son propre bien-être. La question ne se pose tout simplement pas.

Le second problème que l'on trouve dans la TSE2 est que les comportements éthiques et altruistes (que l'altruisme soit instrumental ou non) sont sans cesse confondus. En ce sens, il suffirait qu'une action soit altruiste pour qu'on puisse la qualifier d'éthique. Dans les modélisations aucune distinction n'est faite entre ces deux concepts. Or, il est très facile d'imaginer des situations où une action altruiste peut être non éthique. Par exemple, le complice d'un meurtre peut aider celui qui torture et assassine la victime sans que sa propre utilité ne soit en jeu. C'est une action altruiste. Mais est-elle éthique ?

Les jugements éthiques sont similaires à l'altruisme dans la mesure où il y a indépendance d'intérêts de ceux qui les font. Cependant, les jugements éthiques diffèrent radicalement des actes altruistes au sens où ils impliquent une dimension universalisable et inéluctable : il n'y a pas de porte de sortie à un jugement éthique (Joyce, 2006). Même s'il a agi de manière altruiste, le complice d'un meurtre devra trouver difficile d'éviter le sentiment de culpabilité consécutif à ses agissements. En réduisant l'éthique à l'altruisme et l'altruisme à l'altruisme instrumental, la TSE2 s'interdit et interdit toute discussion sérieuse sur l'éthique médicale.

Puisque la TSE2 n'est pas en mesure de nous fournir une théorie convaincante de l'éthique médicale, nous proposons en fin de chapitre un fondement possible à une éthique médicale non instrumentale. La notion de besoin de santé est au centre de ce projet. L'objectivité des besoins de santé est alors en opposition frontale avec la subjectivité des préférences dont il est question dans la fonction d'utilité (Hodgson 2009, 2013).

Les besoins de santé tendent à être involontaires et inégalement distribués. Beaucoup de maladies, blessures ou accidents qui créent des besoins de santé sont imprédictibles et indépendants des choix ou des responsabilités de ceux qui en souffrent. Les besoins de santé sont donc proches des besoins vitaux (Batifoulier et al., 2013). Bien que les besoins de santé soient également des besoins sociaux – au sens où ils sont en partie contingents aux conditions sociales des individus, notamment aux inégalités de santé –, la reconnaissance de la dimension universelle du besoin est une des fondations de l'éthique en santé.

Cette éthique en santé n'est pas propre aux médecins, elle est partagée unanimement. La reconnaissance du besoin de santé requière une protection des patients contre la douleur et la souffrance de la part des professionnels de santé. Il s'agit d'une obligation morale que la collectivité délègue aux soignants. C'est de là que naît un *ethos* professionnel avec une dimension morale inéluctable : l'éthique médicale.

La première section montre comment la théorie économique de l'altruisme instrumental a été importée dans l'économie de la santé pour résoudre certains problèmes théoriques issus de l'hypothèse de demande induite. L'incorporation d'un argument « social » dans la fonction d'utilité du médecin est non seulement une avancée dans l'analyse économique mais aussi une nécessité logique pour la crédibilité de la théorie du choix rationnel quand elle cherche à expliquer le comportement médical (1.).

La deuxième section développe le fond de la critique : la tentative d'introduire l'altruisme dans l'hypothèse d'optimisation[24] créé des paradoxes de théorie économique qui créent à leur tour des dilemmes de politique économique (Batifoulier et Da Silva, 2014a et b). Donner aux motivations éthiques le même statut qu'aux autres motivations utilitaires vide l'éthique médicale de toute référence aux valeurs. Cela aboutit à des propositions théoriques qui défient la logique. Expliquer l'éthique médicale avec une conception instrumentale de l'altruisme mène par ailleurs à la formulation de propositions de politique économique absurdes (2.).

La troisième section cherche à reconstruire une théorie de l'éthique médicale à partir de la notion de besoin de santé. Nous définissons le besoin de santé à partir de trois sources théoriques distinctes : le besoin vital, le besoin objectif et le besoin social. L'universalité du besoin de santé ainsi défini est la source d'une éthique en santé s'articulant autour de la notion de souffrance. L'éthique médicale – c'est-à-dire l'éthique des professionnels de santé que sont les médecins – est alors conçue comme s'inspirant de l'éthique en santé tout en ne s'y résumant pas (3.).

[24] Le passage de la TSE1 à la TSE2 dans la théorie du comportement médical.

1. La « préférence éthique » du médecin et le calcul rationnel

Cette première section s'essaie à une description générale de la façon dont l'éthique médicale est introduite dans la TSE. Il s'agit de la naissance de la TSE2 en santé. Qu'elles soient traduites, à la suite de Becker, dans le langage de la théorie de l'altruisme ou intégrées dans l'hypothèse de demande induite, les considérations éthiques pénètrent désormais le *corpus* de la TSE (1.1.). Mais cette formalisation de l'éthique médicale est problématique dans la mesure où elle ne donne pas au médecin une épaisseur propre – comme aurait pu le laisser envisager les développements des premiers économistes qui, tel Arrow (1963), revendiquaient une originalité du bien santé et du comportement médical (1.2.).

1.1 De la demande induite aux deux formes de l'altruisme médical

Le concept de demande induite (Evans, 1974) est particulièrement pertinent en économie de la santé. Dans la plupart des cas, le médecin est producteur et consommateur de soins dans la mesure où il traduit la maladie ou la blessure du patient en une consommation médicale par la prescription d'un traitement. La demande n'est pas indépendante de l'offre : elle est tributaire du médecin tant quantitativement que qualitativement. Dès lors, il devrait y avoir une corrélation positive entre la densité médicale et la consommation médicale : un plus grand nombre de médecins dans un espace géographique donné devrait impliquer une hausse de la consommation médicale plutôt qu'une baisse (comme ce serait le cas sur un marché parfait) et le montant total d'honoraires devrait lui aussi être en augmentation. Cet avantage du médecin face au patient est si puissant qu'il peut aller jusqu'à remettre en cause l'existence même d'une relation médicale. Cependant, la TSE2, en faisant place à l'hypothèse d'intéressement limité, modère la capacité d'induction du médecin par l'existence de l'éthique médicale. Les obligations sociales du médecin auxquelles les théoriciens de la TSE2 font référence presque systématiquement sont une réponse à ce pouvoir original du praticien.

La fonction d'utilité du médecin UM = UM (YM, WP) présente en effet des arguments classiques YM (en matière de revenu, de temps de travail, etc.) et des arguments moins habituels WP reflétant les préférences sociales du médecin, susceptibles d'inclure sous différentes formes des arguments éthiques. L'utilité du médecin incorpore ainsi le bien-être du patient et elle permet d'expliquer le comportement médical en matière de prix et de

quantité. L'utilité mesure alors la satisfaction des préférences, y compris la « préférence éthique » du médecin (Batifoulier et Thévenon, 2003).

La sensibilité éthique du médecin est identifiée à une préférence exogène ou un goût dont on ne cherche pas les origines : *De gustibus non est disputandum*. Comme les autres individus, les médecins ne sont pas responsables de la formation de leur préférence. Cette préférence éthique donnée *a priori* peut apparaître dans la fonction d'utilité sous deux formes génériques (Batifoulier, 2009).

Dans le premier cas, on insistera d'abord sur le pouvoir discrétionnaire du médecin et l'éthique sera mentionnée pour prendre en considération la désutilité de l'exercice d'un tel pouvoir. Certains modèles d'induction fonctionnent sous cette forme : WP = I avec U'I<0 (par exemple, Woodward et Warren-Boulton, 1984 ; Wilensky et Rossiter, 1981, 1984). L'utilité de l'induction (I) est alors positive mais décroissante et l'éthique est invoquée pour justifier cette décroissance (Encadré 2.1.). Elle se présente comme limite au pouvoir de création de demande.

Au nom de l'altruisme, le médecin va brider l'exercice de son pouvoir discrétionnaire. Il va s'interdire la sur-prescription et l'augmentation artificielle du prix (multiplication des consultations, dépassement d'honoraires, etc.) payé par le malade, car cette « médecine maximale » lui est coûteuse sur le plan psychologique (Domenighetti 1995). Le médecin est ainsi modélisé comme maximisant une utilité sous des contraintes qu'il s'impose à lui-même. L'altruisme médical est ici assimilé à une préférence qui corrige les autres préférences. Il agit comme contrainte au *self interest* et aliène les autres arguments de la fonction d'utilité au respect de cette contrainte, ce qui réduit la souveraineté du choix du médecin.

Encadré 2.1. L'éthique médicale comme la désutilité du pouvoir d'induction.

Le modèle de Rossiter et Wilenski (1984)

La théorie de la demande induite pose un problème important aux économistes de la santé partisans de la TSE. Quand s'arrête le pouvoir discrétionnaire du médecin ? Si les médecins sont maximisateurs, jamais. Une solution possible pour résoudre ce problème est de mobiliser le concept d'éthique médicale comme autorégulation du comportement opportuniste du médecin. Dans ce modèle, les auteurs cherchent à décrire la sensibilité éthique du médecin en l'assimilant à la désutilité de l'exercice du pouvoir d'induction. Plus précisément, il s'agit de montrer que le médecin n'agit pas uniquement dans son intérêt égoïste et qu'il tient compte du fardeau du patient.

La fonction d'utilité du patient se présente comme suit[25] :

$UM = UM(YM, I) = UM(R, Y, I)$ où

$I = c.H.y^i$ avec

R, le revenu du médecin,

Y, la charge de travail du médecin convertie en volume de production,

c, le taux de co-assurance,

H, le prix total des services (payé par le patient et par l'assureur)

y^i, le nombre d'actes médicaux induits par le médecin ;

$$\frac{\partial UM}{\partial R} > 0 \; ; \; \frac{\partial UM}{\partial Y} < 0 \; ; \; \frac{\partial UM}{\partial I} < 0$$

Les deux dernières inégalités impliquent que, dans ce modèle, le médecin éprouve une désutilité à l'usage de son pouvoir d'induction (I) ainsi qu'une désutilité, plus traditionnelle, à l'effort. La désutilité de l'induction est justifiée, entre autres choses, par le fait que le médecin sait que l'augmentation irraisonnée de sa production a un impact négatif sur la qualité de sa prestation. Le modèle est complété par l'équation du niveau de production et celle du revenu du médecin :

$Y = y^i + y(c.H, d)$;

$R = H.Y - C(Y, M) + \mu$ avec

C, la fonction de coût,

M, le coût des facteurs de production achetés par le médecin,

d, un indice de densité médicale supposé constant à court terme,

μ, un revenu indépendant de l'activité de production de soin.

Lorsque le médecin maximise son utilité sous contraintes, les conditions du premier ordre donnent :

$$\frac{\partial UM}{\partial H} = UM'_R\left[y^i + y(c.H, d) + H.y'_H - C'_y.y'_H\right] + UM'_Y.y'_H + c.y^i.U'_I = 0$$

et

$$\frac{\partial UM}{\partial y^i} = UM'_R[H - C'_Y] + UM'_Y + c.H.UM'_I = 0$$

En conséquence, le médecin fixe son prix de manière à égaliser les désutilités marginales de la charge de travail et du pouvoir d'induction à l'utilité marginale du revenu. Le médecin a donc intérêt à induire la demande tant que l'accroissement d'utilité marginale du revenu qui en est consécutive est supérieure à l'accroissement des désutilités marginales issues de l'exercice du pouvoir discrétionnaire et de l'effort. Les auteurs pensent ainsi parvenir à démontrer que le médecin maximise son utilité lorsqu'il agit dans l'intérêt du patient et non uniquement dans son intérêt égoïste.

[25] Les notations UM, YM, RM et I sont les mêmes que dans le corps du texte.

Le second cas est davantage en phase avec la théorie économique de l'altruisme en prenant en considération positivement la volonté du médecin d'agir dans l'intérêt du patient. L'éthique est alors explicitement prise en compte dans la fonction d'utilité (WP = E), ce qui reflète le concernement du professionnel pour le bien-être, l'état de santé ou encore de la demande d'un patient représentatif. Dans ce cas, l'utilité du médecin croît avec une telle éthique U'E > 0.

Selon l'imagination du modélisateur et ce qu'il cherche à expliquer, on peut décliner la notion de patient sous différentes formes. Le médecin peut être soucieux du bien-être du patient moyen et quelconque (Richardson, 1981). Il peut aussi être sensible à des types particuliers de patients, par exemple les plus pauvres. On développera alors des modèles de discrimination par les prix où la satisfaction du médecin dépend du revenu de son patient (Ruffin et Leigh, 1973). La segmentation de la clientèle conduit à fixer des tarifs différents suivant les patients (voir l'encadré 2.2). Cette « éthique discriminatoire » s'oppose aux injonctions des codes de déontologie. Le médecin peut également être sensible au patient intéressant d'un point de vue clinique parce qu'il favorise une médecine plus attractive intellectuellement et source de prestige. L'éthique peut alors s'exprimer en fonction du niveau de sévérité de la maladie (Zweilfel, 1981). Le curseur éthique se déplace ainsi en fonction des diverses caractéristiques des patients.

Encadré 2.2. L'éthique médicale discriminatoire. Le modèle de Ruffin et Leigh (1973)

Dans ce modèle, Ruffin et Leigh (1973) mobilisent l'existence d'une éthique médicale pour expliquer la discrimination par les prix. Les médecins facturent la production de façon croissante du revenu des patients. Les médecins seraient ainsi partagés entre charité et concurrence. Un vendeur est qualifié de charitable si le coût marginal du service rendu est supérieur au revenu marginal perçu de ce même service. En d'autres mots, il fait payer le prix de la production en dessous de son coût. Comme il s'agit de service de santé, le coût désigne ici la désutilité de l'effort (évaluée en termes monétaires).

On note Q_1 la quantité de services vendus aux patients les plus pauvres (*poor*), Q_2 la quantité de services vendus aux patients ayant un niveau de revenu standard. (*customary*) et Q_3 la quantité de services vendus aux plus riches (*rich*). On note le prix en fonction de la catégorie de revenu, P_i. Le profit R est la somme des revenus moins les coûts de production C :

$$R = \sum_{i=1}^{i=3} P_i Q_i - C$$

Par hypothèse, chaque médecin est capable d'identifier la classe de revenu de chaque patient. Il n'est pas nécessaire que le médecin connaisse le revenu exact de chaque patient ; il suffit simplement d'une approximation permettant de procéder au classement entre « pauvre », « standard » et « riche ». Chaque médecin peut vendre la quantité désirée aux patients (au prix P_1, P_2 et P_3). Les médecins restent donc en concurrence sur chaque segment de marché bien que les marchés des services charitables et non-charitables soient surement les moins parfaits. Le coût de production, C, est donné par :

$$C = C(Q) = C(\sum Q_i)$$

Les préférences sont différentes selon les médecins mais ils ont tous une fonction d'utilité quasi-concave :

$$U = v(Q_1, Q_2, Q_3, R)$$

La quasi-concavité permet de procéder à une maximisation mais c'est une hypothèse trop générale pour avoir du sens. On y ajoute alors trois hypothèses sur U_1, U_2, U_3, U_4.

$$U_1 = \frac{\partial v}{\partial Q_1} - \frac{\partial v}{\partial Q_2} > 0$$

U_1. Une augmentation des services de charité (Q_1) avec une variation compensatrice dans les services standards (Q_2), à volume de travail constant, augment l'utilité totale.

$$U_2 = \frac{\partial v}{\partial v_2} < 0$$

U_2. Une mesure de la désutilité du travail.

$$U_3 = \frac{\partial v}{\partial Q_3} - \frac{\partial v}{\partial Q_2} < 0$$

U_3. Une augmentation des services pour les riches compensée par une variation des services pour les patients standards (à travail constant) réduit l'utilité.

Les signes de U_1 et de U_3 reflète l'éthique médicale. Les médecins agissent comme s'ils (*as if*) retiraient plus de bénéfices monétaires quand ils servent les riches que quand ils servent les pauvres.

$$U_4 = \frac{\partial v}{\partial R} > 0$$

U_4. L'utilité marginale du revenu est croissante.

Chaque médecin maximise son utilité sous contrainte de revenu :

$$R - \sum P_i Q_i + C(Q) = f(Q_1, Q_2, Q_3, R) = 0$$

Après maximisation sous contrainte, les conditions d'optimalité sont :

$$\frac{U_1}{U_4} + P_1 \leq C' - \frac{U_2}{U_4}$$

U_1/U_4 représente l'utilité marginale des services de charité divisée par l'utilité marginale du revenu soit le taux marginal de substitution entre le revenu et les services de charité. Le membre gauche de l'équation représente donc le revenu marginal total : incluant le prix unitaire et le revenu psychique des services de charité. Si Q_1 est

différente de 0 on obtient l'égalité entre revenu marginal et coût marginal. En effet, C' représente le coût marginal de production et U_2/U_4, le coût marginal de l'effort (ou encore le taux marginal de substitution entre effort et revenu).

Le prix des services standards :

$$P_2 \leq C' - \frac{U_2}{U_4}$$

Le prix des services pour les riches :

$$\frac{U_3}{U_4} + P_3 \leq C' - \frac{U_2}{U_4}$$

Les conditions sur le revenu et les quantités offertes :

$$R = \sum P_i Q_i - C(Q)$$

$$Q_i \geq 0$$

$$R \geq 0$$

Au final, Ruffin et Leight (1973) proposent un modèle où l'éthique explique la discrimination par les prix. L'éthique médicale est ici associée à un goût pour les pauvres et un dégout pour les riches expliquant que les premiers payent moins que les seconds.

1.2.Un altruisme médical sans épaisseur propre

L'éthique ainsi formalisée met en avant le souci de l'autre. Si elle est appliquée au cas du médecin, elle ne lui est pas dédiée. La bienveillance ou l'altruisme sont des comportements génériques que l'on peut trouver chez chaque individu et qui ne sont pas spécifiques au médecin. Le médecin exerce une bienveillance envers son patient de la même façon que l'instituteur envers son élève ou le parent envers son enfant. L'éthique médicale est alors assimilée à un altruisme générique, dans un environnement spécifique. Le médecin n'a pas d'épaisseur propre (Davis et McMaster, 2007) alors que la spécificité de la médecine tient à l'existence d'un *ethos* professionnel (Hodgson, 2009). Cet altruisme médical a des « airs de famille » avec différentes notions forgées dans des contextes tout à fait différents. On trouve ainsi dans la littérature d'autres termes en lieu et place de la notion d'altruisme pour qualifier le comportement du médecin. Le « concernement » du médecin pour son patient est proche de celui que l'on retrouve dans la famille, pour les biens collectifs ou encore dans certaines configurations d'économie industrielle (Tableau 2.1.).

Tableau 2.1. Divers termes pour qualifier l'altruisme médical

L'altruisme médical comme …	*Care*	*Externalité*	*Humanitarian spillover*
Origine	Économie des services et de la famille	Économie publique	Économie industrielle et spatiale
Définition	Prêter attention et « soins » aux autres	L'utilité du médecin dépend positivement de celle de son patient	Effet de débordement humanitaire
Portée pour l'interaction patient médecin	Le médecin préfère le particulier (son patient) au général. Le *care* est conditionnel aux personnes	Le bien-être du patient (état de santé, consommation de soins) rejaillit sur le bien-être du médecin	Le médecin attend des retombées de son souci du patient.

Quels que soient les mots employés et la forme de la formalisation de l'éthique professionnelle comme altruisme, il s'agit de récuser l'exercice plein et entier d'un pouvoir discrétionnaire. Supposer qu'un tel pouvoir n'est pas contesté aboutit à un résultat absurde faisant du médecin un agent économique comme un autre.

> *« In its extreme form, it assumes that there is no limit to the physician's ability to induce demand. But this leads to the absurd result that as long as the price of health services exceeds marginal cost, physicians will continue to induce demand until the patients' full income is extracted » (De Jaegher et Jegers, 2000, p. 234).*

De Jaegher et Jegers (2000) font remarquer que dans ces modèles, le comportement du médecin qui induit la demande n'est pas si éloigné de celui de la firme qui essaie de modifier la demande. La publicité a pour rôle principal de diffuser des informations élogieuses sur un produit avec pour objectif l'induction de la demande à un niveau différent que celui qui aurait été sans publicité ou avec une information parfaite. La différence réside dans le fait que si la firme subit un coût pécuniaire (le financement de la publicité, le *marketing*, etc.), le médecin subit un coût moral : la violation des principes éthiques.

La prise en compte d'une éthique, au moins sous forme d'un « coût moral » est donc indispensable à la formalisation. Elle est même instrumentalisée dans ce sens car elle permet de dérouler le calcul économique traditionnel – dans un domaine qui *a priori* n'était pas accueillant pour un tel calcul – sans faire du médecin un entrepreneur individuel. L'argument

éthique est plus malléable, et *in fine* plus orthodoxe qu'une hypothèse de revenu cible[26] où le médecin se fixe une limite à la maximisation de son revenu. La cible atteinte, il ne cherche plus à induire de la demande. Cet argument peut plaider pour l'abandon de la rationalité maximisatrice au profit d'une rationalité procédurale à la Simon où les individus se fixent des seuils de satisfaction

De fait, l'analyse de la rationalité économique du médecin s'accompagne quasi systématiquement de la prise en compte d'une éthique. Elle fait partie de la boite à outil du modélisateur. Dans ce sens, cette microéconomie de la santé est aussi inévitablement une microéconomie de l'éthique. L'éthique ainsi formalisée est donc loin d'être une anomalie et le médecin est un acteur économique comme un autre. Cette impérieuse nécessité de l'éthique pose néanmoins certains problèmes.

2 Altruisme médical instrumental : paradoxes de théorie économique et problèmes de politique publique

La conception réductrice de l'éthique développée par la TSE2, a des implications pour la compréhension du comportement médical. Nous pouvons identifier plusieurs niveaux d'analyse. La première série de difficultés est théorique : en réduisant l'éthique à l'altruisme et l'altruisme à l'altruisme instrumental, la théorie *mainstream* de l'altruisme conduit à un nombre important de paradoxes. L'altruisme devient de l'égoïsme et les problèmes éthiques sont en fait ignorés. La seconde série de problèmes est liée à la politique économique. Dans ce cas, la critique est déplacée vers un régulateur central qui souhaiterait réguler les médecins altruistes. Une telle régulation est difficile parce que l'altruisme n'est jamais pur. Les formalisations du comportement médical cherchent à exclure le cas d'un médecin totalement altruiste (UM=WP=E) et celui d'un médecin totalement égoïste (UM=YM), avec YM, les préférences individuelles du médecin et WP, ses préférences sociales. Le régulateur central doit alors faire face à des médecins dont le comportement est « quelque part » entre l'altruisme et l'égoïsme.

La façon dont la TSE 2 intègre l'éthique du médecin dans le calcul rationnel soulève de nombreuses difficultés théoriques. Elle débouche aussi sur une inaptitude à proposer des

[26] Cette hypothèse est notamment utilisée par Evans (1974) afin d'expliquer les limites de l'usage du pouvoir d'induction. On pourra aussi voir Sweeney (1982) ou Rizzo et Zeckhausser (2003).

recommandations pertinentes de politique économique. Nous proposons d'appeler « paradoxes » ces difficultés théoriques pour signifier qu'elles défient la logique, se contredisent elles-mêmes ou expriment une opinion absurde. Les paradoxes de théorie économique ne sont pas sans conséquences sur les recommandations de politique publique (Batifoulier et Da Silva, 2014a et b).

C'est ici le cœur de la critique de la TSE2 lorsqu'elle aborde l'éthique médicale. En effet, ce qui est le plus dommageable ce n'est pas tant l'existence de paradoxes de théorie économique – les théoriciens *mainstream* et les autres peuvent bien disserter sans fin entre eux – que le fait que ces paradoxes engendrent de graves problèmes de politique économique. On ne sait plus quoi faire. La TSE2 n'est alors plus utile – ni pour comprendre le monde ni pour agir sur lui. Le tableau 2.2. résume notre perspective critique. Dans les sous-sections suivantes nous décrivons chacun de ces paradoxes ainsi que les problèmes de politique économique qui en découlent.

Tableau 2.2. Des paradoxes de théorie économique aux problèmes de politique économique

Paradoxes de théorie économique	Problème de politique économique
Altruisme médical egocentrique	Signer un chèque en blanc aux médecins
Altruisme médical sadique	Protéger le patient du médecin
Altruisme médical masochiste	Protéger le médecin du patient
Prime à l'égoïsme	Sous-payer le médecin altruiste

2.1 Altruisme médical égocentrique

2.1.1. L'altruisme instrumental : la dissolution du jugement dans le calcul

La première des critiques faite à l'altruisme médical est celle qui est formulée à l'encontre de la théorie économique de l'altruisme dans son ensemble. L'éthique professionnelle, intégrée dans une fonction d'utilité n'a pas de valeur en soi. Elle n'est qu'un instrument pour une fin donnée. Le moyen éthique (ou désutilité de l'induction) ne se justifie que par la fin recherchée : ne pas faire du médecin un maximisateur sans limite.

Les arguments de la fonction d'utilité, qui se prêtent au même type de calcul, sont alors interchangeables du point de vue de leurs conséquences en termes de satisfaction du médecin. La force de cette modélisation de l'altruisme médical est en effet de réduire les avantages et inconvénients du respect d'une norme comme l'éthique médicale à une seule grandeur

homogène : l'utilité, en intégrant les valeurs dans une analyse coût-avantage. Cette force est aussi une faiblesse car ceci revient à définir l'éthique par défaut et à dissoudre le jugement moral dans le calcul (Haussmann et Mc Person, 1993, Ben-Her et Putterman, 1998, Folbre et Goodin, 2004). L'analyse rend tous les arguments de la fonction d'utilité commensurables et il est toujours possible d'opérer un *trade off* entre différents arguments. Or, en comptant avec le même numéraire des préférences d'ordres différents, on peut aboutir à des résultats contradictoires. Ainsi, dans la composante individuelle de la fonction d'utilité, l'effort du médecin est compté négativement (désutilité de l'effort) mais dans sa composante sociale, il devient positif puisque le médecin est soucieux du bien-être du patient et y consacre temps et énergie. L'effort privé qui entraîne une désutilité, et l'effort social qui est à la base de la reconnaissance du patient, ne sont pas de même niveau[27].

A l'inverse de l'altruisme « altercentrique » (Khalil, 2003), cet altruisme médical est égocentrique et/ou égoïste car il est « bon » d'un point de vue personnel et non en tant que tel[28]. Le médecin maximise son bien-être en cherchant le bien-être du patient. Les vertus morales ne sont engendrées qu'à partir de l'intérêt. La sympathie pour l'autre est jugée au regard des bénéfices qu'elle engendre. Les acteurs maximisent toujours d'abord leur propre utilité et non celles des autres. Le bien reste rattaché irréductiblement au bien-être subjectif. On peut alors refuser le label « altruisme » à cette conception (Rose-Ackerman, 1996, p. 713) ou préférer le mot *commitment* (Sen, 1987)[29] pour dépasser le cadre de la simple sympathie et de l'interdépendance des utilités de façon à donner une dimension morale à l'altruisme. L'engagement réellement altruiste conduit à préférer une action qui peut entraîner une baisse de son utilité alors qu'une autre action, à sa portée, pourrait augmenter cette utilité.

[27] Une voie possible pour résoudre ce paradoxe conduirait à adopter une autre approche des préférences (méta-préférences ou préférences réflexives) s'écartant de l'axiomatique de la théorie du choix rationnel.

[28] Khalil distingue trois types de théories de l'altruisme : (i) l'altruisme égoïste (*egoistic*) est l'attitude d'un individu portant assistance à un autre uniquement s'il en attend un bénéfice futur ; (ii) l'altruisme égocentrique (*egocentric*) est une situation où la fonction d'utilité inclut celle du receveur ; (iii) l'altruisme altercentrique (*altercentric*) découle d'un trait de personnalité de type pro-social qui n'est pas généré à partir d'une intention, c'est un gène moral.

[29] Le *commitment* peut se définir négativement comme une action indépendante de la question des gains et des pertes ou encore une action qui ne relève pas du calcul coût/avantage.

2.1.2. Quand l'altruisme instrumental rencontre la médecine libérale : un chèque en blanc ?

Cet altruisme égocentrique débouche sur un altruisme paternaliste où le médecin prend à sa charge la santé de son patient car il est soucieux de son bien-être. Il renonce de lui-même à ses avantages pour exprimer les intérêts du patient. En phase avec la tradition beckerienne, le médecin est perçu comme un dictateur bienveillant qui sait ce qui est bon pour le patient (comme le « bon père de famille »). Chacun y trouve son compte car le « bonheur » du médecin est réalisé au travers du « bonheur » du patient. En contrepartie, le médecin est assuré d'un retour sur son investissement éthique si le patient se conduit comme un « enfant »[30] : il se montrera coopératif et fidèle au médecin.

Cette conception de l'altruisme médical est en phase avec celle de la médecine libérale. Le paternalisme médical constitue en effet une doctrine de la profession médicale qui, comme l'économiste de la santé, insiste sur les inégalités de position entre patient et médecin. Le malade doit s'en remettre à celui qui sait et dont le jugement n'est pas brouillé par la souffrance. Le droit d'ingérence se déduit naturellement de l'inégalité de position. Le paternalisme se justifie ainsi par la lucidité et la rationalité du médecin face un patient dont la sécurité réclame la soumission[31]. Il y a donc bien pouvoir de la profession médicale, mais canalisé volontairement par une responsabilisation du médecin qui veut le bien de son patient. C'est en ces termes qu'est définie l'éthique dans le rapport *Éthique et Profession de santé* :

> *« Nous comprenons l'éthique comme la mise en question du pouvoir et de la puissance par la responsabilité pour autrui… En d'autres termes, il s'agit d'interroger le pouvoir par le devoir » (Cordier, 2003, p.18).*

Les conclusions de politique économique sont immédiates : donner du crédit à la notion d'éthique professionnelle dans ce cadre paternaliste revient à signer un « chèque en blanc » aux médecins, seuls capables de connaître ce qui est bon pour la santé du patient et naturellement orientés vers un comportement bienveillant. Il est alors nécessaire de se donner les outils capables de renforcer la liberté et l'autonomie du médecin. S'il est dans la mission du médecin d'adopter le principe *patient comes first* alors la politique économique, qui veut

[30] Ou *rotten kid*, enfant « pourri gâté » au sens où il est égoïste au départ mais devient altruiste sous l'effet de l'altruisme du parent, assimilé par Becker à des incitations qui modifient le calcul de l'enfant.

[31] Voir l'analyse de Jaunait (2005) à partir notamment des écrits de Louis Portes, le premier président du conseil de l'ordre.

défendre l'intérêt du patient au nom du bien commun qu'est la santé, n'a plus qu'à s'effacer derrière l'autorégulation professionnelle.

2.1.3. Préférences individuelles contre éthique collective

Cette conception assure la coordination entre le patient et le médecin simplement par le jeu des préférences individuelles. Le médecin altruiste ratifie les préférences du patient sans procéder à un jugement moral. Dans ce cadre, les préférences du patient ne peuvent être que des désirs et non des besoins. Or, la subjectivité des désirs rend les préférences individuelles plus fragiles et remet en cause le paternalisme médical désormais suspect aux yeux de la société. Le médecin altruiste peut en effet satisfaire les désirs du patient au prix de la violation des codes éthiques en vigueurs. L'exemple extrême de l'euthanasie permet de comprendre l'argument : un médecin peut aider un patient à mourir pour des raisons altruistes, mais en quel sens s'agit-il là d'une décision éthique ?

On peut également prendre l'exemple de l'interruption volontaire de grossesse dont le statut est régulièrement remis en question. En France, avant la loi Veil du 17 janvier 1975, l'avortement était interdit et passible d'une peine de prison. Le Serment d'Hippocrate interdisait également la pratique jusqu'à sa version de 1966 : « je ne remettrai à aucune femme un pessaire abortif », disait-il. Quel devait alors être le principe d'action du médecin ? Fallait-il respecter la loi ? Ou bien fallait-il, de façon altruiste, satisfaire les demandes des patientes ? Ces demandes étaient-elles des préférences – des désirs – ou des besoins ? La violation de l'interdit était-il synonyme de violation de l'éthique ?

En l'absence de compétences éthiques, le médecin ne peut pas différencier l'éthique et l'altruisme (qu'il soit authentique – c'est-à-dire indépendant de la question des intérêts – ou non). C'est parce que les patients expriment des besoins objectifs et universels avec une puissante dimension morale que la société est capable de les satisfaire et de déléguer ce travail aux médecins. Le souci de l'autre exprimé par la profession médicale n'est pas simplement une fonction de l'utilité du patient : il est lié à une dimension morale universelle. Le pouvoir médical a pour explication principale cette obligation d'agir. Or, la théorie économique *mainstream* n'en dit rien.

Paradoxe de théorie économique 1 : L'altruisme médical est intéressé et le médecin est bienveillant parce qu'il y a intérêt.

Problème de politique économique 1 : Dans cette conception, il faut laisser s'autoréguler la profession médicale car c'est la seule à savoir ce qui est bon pour le patient. Il n'y a pas de problème de coordination autres que les barrières à la libre entente.

2.2. Altruisme médical sadique

2.2.1. De l'« autrisme » au sadisme médical

À des niveaux d'analyse différents, discours médical et formalisation économique se rejoignent pour ramener l'altruisme médical à la prise en compte du patient. Cependant, rien ne garantit que le médecin soit naturellement bienveillant et la modélisation de l'altruisme à partir de l'interdépendance des fonctions d'utilité est insuffisante pour garantir cette bienveillance. Elle réduit l'altruisme à un « autrisme » sans valeurs (Mahieu, 1998, Ballet et Bazin, 2006).

La présence du patient dans la fonction d'utilité du médecin n'est en effet qu'une extension de la rationalité individuelle à la présence de l'autre. Le fait de veiller sur l'autre s'apparente à une externalité (*caring externality*) qui peut être positive ou négative. Elle n'a pas besoin de mobiliser une norme sociale quand il suffit d'une partielle identification avec le bien-être de l'autre. Cet altruisme est alors neutre *a priori* et il ne produit que des utilités économiques où l'action est jugée par rapport aux gains qu'elle engendre pour soi. Si l'on reprend et simplifie l'écriture de la fonction d'utilité du médecin UM= UM (YM, WP), alors l'égoïste est celui qui est insensible à l'autre $\partial UM/\partial WP = 0$. C'est l'inverse dans le cas altruiste $\partial UM/\partial WP \neq 0$. L'envie, la jalousie, la malveillance sont des formes de cet altruisme ($\partial UM/\partial WP < 0$) qui produisent des utilités négatives. Il faut alors faire un pas supplémentaire pour postuler un médecin bienveillant ($\partial UM/\partial WP > 0$). Dans ce cas, le médecin est gagnant à se soucier du patient.

Mais ce souci de l'autre, ainsi formalisé, développe une conception très particulière de l'altérité : il peut suggérer l'existence d'un comportement médical sadique. Comme il peut gagner de l'utilité en agissant dans l'intérêt du malade, le médecin peut avoir intérêt à la souffrance du patient qui l'amène à consulter et à solliciter sa bienveillance (McMaster, 2007). Ce médecin bienveillant gagne à s'investir dans le processus thérapeutique. Il a alors besoin de l'existence de la maladie pour retirer un profit de l'interaction avec le patient. La maladie du patient sert de valorisation à l'intérêt lucratif du médecin :

> *« […], in extremis, a "caring" physician may enter upon an infinite regress in gaining utility from easing a patient's pain, only to desire to establish the status quo ante in order to vicariously experience the process again, presumably after calculating the discounted disutility of inducing the initial state! Under this conception medical cures are the last thing physicians wish: therapeutic treatments are ruled out, and medical procedures are reduced to some form of intermittent palliative episodes of "care" followed by periods of a toleration for a patient's worsening medical condition, to be followed by yet further episodes of palliative "care", and so on. » (McMaster, 2007, p. 18).*

C'est à travers la souffrance du malade que le médecin est soucieux des intérêts de son patient. La souffrance est ici le moteur de l'action intéressée. La relation médicale est alors vidée de son sens. Un médecin qui profite de la souffrance du patient est bien peu bienveillant. Il active la suspicion du patient qui n'a pas intérêt à se fier à lui. La confiance médicale est alors impossible et la relation thérapeutique inexistante. Ce type de configuration qui exacerbe la logique de l'intérêt instrumental et égoïste conduit à des situations du type dilemme du prisonnier où toute coopération, si elle est raisonnable, est rationnellement impossible (Encadré 2.3.).

Encadré 2.3. Le sadisme médical dans la configuration du dilemme du prisonnier

On peut imaginer une situation où un patient a le choix entre faire confiance et ne pas faire confiance à son médecin (s'engager sincèrement dans la relation thérapeutique, prendre régulièrement ses médicaments, etc.) et le médecin a le choix entre entretenir le mal du patient ou le soigner rapidement au risque de perdre le bénéfice de consultations répétées.

Si le médecin décide de soigner le patient, il s'assure une rémunération financière (F) qui est nulle s'il ne soigne pas le patient. Dans le cas où le médecin soigne, le patient est assuré de guérir rapidement (GR) ce qui lui procure une utilité supérieure à la situation où le médecin profite de sa souffrance. En effet, on considérera que, même si le médecin profite de sa douleur, le patient perçoit une source d'utilité du fait d'une guérison lente (GL). Si le médecin décide de ne pas soigner le patient et, au contraire, d'exercer son sadisme – consécutif à la forme particulière d'éthique médicale dont il est ici question –, il perçoit une « récompense morale » (S) supérieure à la récompense financière. Reste la stratégie du patient : s'il fait confiance au médecin, il subit une perte d'utilité consécutive à l'effort de suivi du traitement thérapeutique (E) ; si, à l'inverse, il ne fait pas confiance au médecin, c'est ce dernier qui subit un coût (C) lié aux difficultés que lui fait le patient.

Si on prend les valeurs, S=GR+3 ; F=-C=2 ; GL=-E=1 on obtient une situation de dilemme du prisonnier (Tableau 2.3.). Dans cette configuration, l'équilibre de Nash est donné par les stratégies de meilleure réponse « ne pas faire confiance / entretenir le mal » (cadran B). Or, l'équilibre est Pareto dominé, par la configuration du cadre C (« confiance » / « soigner le mal »). C'est donc la présence d'une éthique médicale qui rend l'équilibre sous-optimal.

Tableau 2.3. Le dilemme du prisonnier avec un médecin sadique

		Patient	
		Confiance	Pas Confiance
Médecin sadique	Entretenir le mal	A : (S, GL-E) = (3, 0)	B : (S-C, GL) = (1, 1)
	Soigner le mal	C : (F, GR-E) = (2, 2)	D : (F-C, GR) = (0, 3)

2.2.2. Protéger le patient de son médecin

Dans ces conditions, c'est-à-dire avec un médecin sadique face à un patient impuissant, quel type d'action doit être envisagé par le régulateur central ? Le patient a besoin du médecin pour améliorer son état de santé mais ce dernier est intéressé à la souffrance du patient. Il peut être contreproductif d'envoyer un patient consulter son médecin. Le régulateur doit alors protéger le patient – qui a besoin de soins – du médecin – qui fournit les soins.

Ce paradoxe est lié à l'incapacité de l'économie *mainstream* à reconnaitre les spécificités des soins de santé. Prendre soin des besoins des autres membres d'une communauté est une caractéristique morale universelle (Joyce, 2006). C'est particulièrement le cas dans le domaine de la santé puisque les problèmes de santé sont involontaires. De ce point de vue, éviter la souffrance des patients constitue une justification morale à l'action médicale. Cela est nettement plus profond que la question de l'altruisme (instrumental ou non) dans la mesure où le médecin lui-même est limité par l'obligation de satisfaire les

besoins du patient. Il ne peut pas manipuler la souffrance humaine dans son intérêt personnel sous peine d'être ensuite noyé par la culpabilité. Ce sens moral provient en partie du besoin de santé du fait de sa dimension objective et de la nécessité d'éviter la souffrance des patients (Batifoulier et al. 2013)

Paradoxe de théorie économique 2 : L'altruisme médical débouche sur un sadisme médical.

Problème de politique économique 2 : Dans cette conception du médecin, il faut protéger le patient d'un danger encore plus grand que la maladie : le médecin lui-même.

2.3.Altruisme médical masochiste

2.3.1. Du biais égocentrique au masochisme médical

L'analyse de l'altruisme médical est également vulnérable au « biais égocentrique », mis en avant par la littérature psycho économique, qui décrit la tendance des individus à croire que les autres feraient la même chose à leur place (Levy Garboua et al., 2006, Petit, 2011). Lors d'un problème de coordination, les individus ont ainsi tendance à surpondérer la proportion des partenaires qui choisiront la même stratégie qu'eux. Celui qui veut coopérer pense plus facilement que les autres coopéreront et celui qui pense faire défection pense plus facilement que les autres feront défection. C'est ce que les psychologues nomment l'effet de faux consensus ou biais égocentrique (Ross et al., 1977).

Les résultats d'une expérimentation peuvent permettre de comprendre la portée de ce biais cognitif (Verlhiac, 2000, p. 154). Plusieurs sujets ont été invités à regarder une vidéo ainsi que les réactions à cette vidéo de quinze personnes. Après visionnage, il était demandé aux sujets de se rappeler le nombre de personnes qui étaient en accord avec les idées présentées dans la vidéo. Les sujets finissaient l'expérience en donnant leur point de vue (« pour » ou « contre »). Lorsque la répartition des opinions était très clairement discriminatoire (par exemple, 10 « pour » et 5 « contre »), les sujets répondent avec justesse quant à la répartition de l'échantillon. Il n'y a pas d'effet consensus. Mais, lorsque les opinions sont plus nuancées (avec un partage moins fort ou avec des options de réponse plus élevées), la justesse des réponses se dégrade fortement. L'expérimentation montre alors

l'existence d'un biais égocentrique : les sujets se prononcent plus probablement en accord avec leurs propres croyances.

Ainsi, à partir de sa propre position, un médecin bienveillant peut inférer la bienveillance du patient. Cette projection peut être dangereuse. Dans la formalisation de l'altruisme sous forme de préférence du médecin, on ne sait rien du patient. La manière dont le patient réagit à l'engagement éthique du médecin est supposée positive. Or, rien ne lui interdit de profiter de la naïveté du médecin. L'altruisme de l'un peut même conduire à doper l'opportunisme de l'autre. Ce paradoxe de l'altruisme unilatéral a été développé par Marciano (1998, 2005) à partir d'une extension du « théorème de l'enfant gâté » de Becker. Si le patient n'est pas supposé avoir les mêmes « valeurs » éthiques que le médecin, il peut se conduire en « mauvais enfant gâté » et le « retour sur investissement éthique » qu'espère le médecin est inexistant. On est alors dans un cadre de « dilemme du Samaritain »[32] où le médecin est gagnant à se montrer altruiste tout en sachant que le patient va exploiter la situation. Le médecin soucieux du bien-être du patient est alors un médecin naïf et la bienveillance est contreproductive. Tout se passe comme si un maximisateur contrarié était considérablement fragilisé par un maximisateur pur (Encadré 2.4.).

[32] Qui est une extension du dilemme du prisonnier.

Encadré 2.4. Le dilemme du Samaritain

Le dilemme du Samaritain (voir Buchanan, 1975, Ballet et al. 2007,) est un jeu célèbre mettant aux prises un individu « parasite » qui cherche à profiter de la charité d'un Samaritain. Le parasite doit choisir entre travailler pour subvenir lui-même à ses besoins ou solliciter la générosité d'autrui. Le dilemme du Samaritain est alors le suivant : soit il décide de ne pas aider le parasite, auquel cas il viole ses principes moraux, soit il aide le parasite ce qui équivaut à l'encourager dans son égoïsme. Il est possible de formuler le dilemme de la façon suivante (Tableau 2.4.) :

Tableau 2.4. Le dilemme du samaritain

		Le Samaritain	
		Aider	Ne pas aider
Le parasite	Travailler	(2, 2)	(0, 3.(1-a))
	Profiter	(3, 3.a)	(1, 1)

Il s'agit de la même matrice des gains que pour une situation de dilemme du prisonnier mis à part le paiement du Samaritain lorsque les deux joueurs ont une stratégie divergente. Lorsque le Samaritain aide et que le parasite profite, le premier des deux ne perd pas tout : selon son degré d'altruisme (a ; 0<a<1), il retire une satisfaction de son comportement vertueux. Mais, s'il n'aide pas alors que le parasite fait l'effort de travailler, selon son degré d'altruisme, le Samaritain éprouve une peine qui réduit son utilité. Notons que le cas où a = 0 est le dilemme du prisonnier. Lorsque 3.a >1 la solution du jeu change et l'équilibre de Nash devient « aider / profiter ». Il y a alors exploitation du Samaritain par le parasite.

2.3.2. Protéger le médecin de son patient

Dans ces conditions, le régulateur central doit protéger le médecin de son altruisme inconditionnel. Celui-ci est coûteux et contreproductif puisqu'il active un comportement opportunisme chez le patient, source d'augmentation des dépenses. La tutelle peut alors être amenée à taxer cet altruisme inconsidéré. Le médecin ne peut en effet réagir face à un patient opportuniste car il a les « mains liées par son altruisme ». Il doit déléguer les sanctions à un tiers (la tutelle) qui va obliger le patient à être altruiste en retour. Ce schéma où l'altruisme médical devient un problème pour le régulateur est imputable à une défaillance de la confiance et de la réciprocité. La formalisation des valeurs par une interdépendance des fonctions d'utilité rend perverse la bienveillance médicale.

Le modèle construit par Liu et Ma (2012) approche l'idée selon laquelle il est nécessaire de protéger le médecin de son propre altruisme. Dans ce modèle, les médecins sont hétérogènes dans leur degré d'altruisme. Dans un jeu en quatre étapes, ils doivent s'engager sur un protocole de soins (dont l'un est peu cher mais inefficace et l'autre cher mais efficace) et appliquer ce protocole lorsque la maladie se déclare chez le patient. De son côté, le régulateur doit trouver l'assurance optimale pour le patient et la méthode de paiement optimale pour le médecin. Leur résultat est très proche de l'idée d'une protection du médecin.

En effet, dans le *first best*, le régulateur doit offrir un contrat de paiement à tous les médecins comme s'ils étaient tous semblables au plus opportuniste d'entre eux. En conséquence, les médecins altruistes sont conduits à renoncer à leurs actions altruistes pour leur propre bien. Lorsque le *first best* n'est pas réalisable, certains médecins, sachant qu'ils seraient incapables de résister à leurs penchants altruistes, renoncent à s'engager dans le processus de soin. En effet, si ces médecins (trop) altruistes s'engagent, ils risquent de procurer à leurs patients des soins trop coûteux étant donné leur état de santé mettant ainsi en danger la santé financière de leur entreprise. Le régulateur doit alors intervenir à la première étape du jeu pour que cela n'arrive pas. Dans un cas comme dans l'autre, tout se passe comme si les médecins devaient être protégés de leur propre altruisme dans l'intérêt du système de santé.

L'économie comportementale (*behavioral economics*) a montré que la coopération est impossible si l'hypothèse d'intéressement n'est pas levée (Camerer et al. 2011, Fehr et Fischbacher, 2003, Gintis et al. 2003, Kahneman, 2011). L'interaction patient/médecin n'est pas une exception à cette règle. Or, dans cette modélisation de l'éthique médicale, il n'y a ni éthique ni altruisme : le besoin du patient est placé au second plan pour lui préférer la figure de l'opportunisme. Encore une fois, il n'y a ni éthique médicale, ni altruisme dans ce type d'action. La conception d'un patient froid et cynique qui exploite l'altruisme médical est équivalente à la conception d'un individu hédoniste réduit à une machine à plaisir – *a pleasure machine* (Hodgson, 2013). Elle ignore complètement la peur et l'angoisse associée à la maladie. Les émotions constituent pourtant un important mécanisme de régulation de la morale individuelle (Joyce, 2006)

Paradoxe de théorie économique 3 : L'altruisme médical débouche sur un masochisme médical.
Problème de politique économique 3 : Dans ces conditions, il faut protéger le médecin de son propre altruisme en ciblant le patient parasite sans quoi aucune coordination n'est possible.

2.4. Prime à l'égoïsme et punition de l'altruisme

2.4.1. Quel contrat pour le médecin altruiste ?

La grande majorité de la théorie des incitations formalise désormais le comportement du médecin en prenant en considération l'existence d'un altruisme (TSE2). Nous avons déjà mis en avant que celui-ci n'est pas justifié autrement que par la volonté de ne pas modéliser le

médecin comme un simple entrepreneur individuel. De la demande induite à la théorie des contrats, l'altruisme médical devient une mention incontournable car il n'apparaît pas sérieux et même contreproductif de le passer sous silence.

Sans changement par rapport à l'altruisme dans la problématique de l'induction, l'altruisme de la théorie des contrats vient tempérer le pouvoir discrétionnaire du médecin mais pas l'annuler. Comme il contredisait l'existence d'un effet d'induction, un altruisme total nie l'utilité d'un contrat incitatif. Il revient à supposer l'existence d'une relation d'agence parfaite, ce qui fait perdre tout son intérêt scientifique à la problématique de l'agence. L'économiste n'aurait nul besoin de construire des outils d'incitations extrinsèques à l'honnêteté si l'existence de motivations intrinsèques joue déjà le même rôle[33].

La solution de facilité consisterait à récuser l'éthique ou à la considérer comme une servitude encombrante. La travailler dans un cadre contractuel est plus difficile. En effet, un médecin purement opportuniste est totalement sensible aux incitations financières et le contrat alors plus facile à écrire. Que faire par contre d'un médecin soucieux du respect d'une éthique professionnelle ?

Une approche possible consiste à moduler le contrat en fonction des caractéristiques du médecin. Un médecin uniquement sensible à son intérêt privé est beaucoup plus facile à orienter vers les choix de la puissance publique. Il suffit de le dédommager pour la désutilité de son effort. Mais si le médecin est altruiste, même à dose réduite, cette politique peut être difficile à mettre en œuvre car l'aléa moral qu'il faut contrecarrer n'est plus certain mais seulement éventuel. On doit alors distinguer les médecins par leur degré d'altruisme (Bardey, 2002, Bardey et Lescur, 2006). Un médecin concerné par l'état de santé de ses patients produit de la « haute » qualité sans incitation :

> *« Indeed, it is very likely that providers vary widely in their altruism towards patients, so it would be necessary to pay different providers differently – the more benevolent having a smaller share of incurred cost reimbursed.» (Jack, 2005, p. 75).*

La politique d'incitation est donc nécessaire pour un médecin opportuniste qu'il faut réorienter mais pas pour un médecin altruiste qui est déjà spontanément intéressé par l'état de santé de ses patients. L'altruisme n'a pas besoin d'incitation et est donc moins coûteux.

Toutefois, la tutelle ne sait pas si le médecin est opportuniste ou non. Le contrat va donc devoir révéler au préalable le type du médecin (Bien, 2001, Jack, 2005 ; Encadré 2.5.). Par

[33] Nous reprendrons plus profondément la distinction entre motivations intrinsèques et motivations extrinsèques au Chapitre 3.

exemple en dirigeant les plus altruistes vers le secteur public, avec des incitations plus faibles, et les plus opportunistes vers le secteur privé, avec de fortes incitations financières. C'est ce que propose le modèle réalisé par Ma (2007) en tentant de prendre au sérieux l'hypothèse altruiste :

> *« It is indeed true that many models predict that maximizing behaviors may actually lead to sincere and altruistic actions. They typically push the 'rational' preference one step higher into the hierarchy of analysis; agents behave sincerely because of their fear of punishment or their expectation of future rewards. I tend to believe that this method obscures rather that enlightens. For policy matters, it is more important to study the practical implications of sincere and altruistic behaviors rather than simply to push the question one step back » (Ma, 2007, p. 80).*

Le degré d'altruisme étant hétérogène selon les médecins, toute la question est de savoir comment la tutelle peut exploiter l'existence de cet altruisme. Dans un contexte où les finances publiques sont en péril, il faut orienter les médecins altruistes à venir travailler dans le public. À l'inverse, les médecins opportunistes doivent être laissés entre les mains du secteur privé. Il est ainsi souhaitable que le secteur public paye moins les médecins de sorte à attirer ce qui se suffisent de leur vertu.

Encadré 2.5. Inciter des médecins à révéler leur altruisme. Le modèle de Jack (2005)

La question que se pose Jack (2005) est de savoir comment acheter des soins de qualité et peu chers. La contrainte de marché suppose de façon assez irréaliste que le patient est capable de juger de la qualité du soin. Pour résoudre ce problème, si l'on suppose qu'à cause de l'éthique médicale le médecin incorpore l'utilité du patient dans sa propre fonction d'utilité, il subsiste un problème : il est très difficile de mesurer le degré d'altruisme spécifique à chaque médecin. Celui-ci est hétérogène et inconnu du planificateur social. Dans ce modèle, un assureur cherchant à maximiser le bien-être social contracte avec les médecins. Les médecins choisissent la qualité des soins (q>0) et l'effort (e>0) le tout affectant ses coûts C(q, e)>0. Cette fonction est supposée convexe dont les dérivées partielles sont :

$$C_1 > 0, C_{11} > 0, C_2 < 0, C_{22} > 0 \; et \; C_{12} < 0$$

L'auteur suppose que le médecin subit une extra-désutilité liée au choix du couple (q, e). On note la désutilité nette du médecin :

$$\gamma(q, e, \theta) = \psi(e) - \theta\phi(q)$$

Où $\theta \in [\theta_0 ; \theta_1]$ est un coefficient mesurant le niveau d'altruisme du médecin. $\gamma(.)$ est la fonction de désutilité nette du médecin, $\psi(.)$ est la fonction du coût de l'effort et $\phi(.)$ est la fonction de bénéfice altruiste du médecin.

L'utilité du consommateur est fonction croissante de la qualité des soins S(q). Le producteur est rémunéré par un montant c de ses coûts et par un transfert t prévu par l'assureur. Le médecin perçoit donc t+c et la patient paye $(1+\lambda)(t+c)$ avec $\lambda>0$ représentant le poids marginal de la fiscalité. Le surplus net du consommateur est donné par :

$$S(q) - (1 + \lambda)(t + c)$$

Et le surplus du producteur est donné par :

$$u = t - \gamma = t - \psi + \theta\phi = \pi + \theta\phi$$

Si l'on passe sous silence dans un premier temps la composante altruiste du surplus du médecin, la fonction objectif du planificateur social est :

$$W = S(q) - (1 + \lambda)(t + c) + t - \psi = S(q) - (1 + \lambda)(\psi + c) - \lambda\pi$$

Dans ces conditions, un médecin de type θ accepte de signer un contrat si et seulement si son utilité nette est supérieure à 0 ($u \geq 0$). Il faut alors s'assurer pour le planificateur que :

$$\pi \geq -\theta\phi$$

En intégrant l'altruisme dans la fonction objectif de bien-être social on a :

$$W_+ = S(q) - (1 + \lambda)(t + c) + t - \gamma = S(q) - (1 + \lambda)(\gamma + c) - \lambda u$$

Lorsqu'il est impossible de contracter sur le montant de l'effort, de la qualité des soins et qu'il est également impossible de connaitre le type de médecin (θ), il est nécessaire de proposer des contrats visant à révéler le niveau d'altruisme du médecin. Il faut proposer un menu de contrat portant sur c et t dans lequel est prévu un paramètre de partage des coûts (*cost-sharing*) différent selon le type de médecin.

2.4.2. Les effets délétères de la prime à l'égoïsme

Selon ce schéma théorique, un médecin déjà altruiste nécessite moins d'incitations que le médecin totalement égoïste. L'altruiste est pénalisé de ce point de vue car il recevra moins de compensation financière de la tutelle. L'incitation financière est une prime à l'égoïsme pour le médecin égoïste. Quel intérêt alors à être altruiste ? Et ce d'autant plus si l'altruisme n'est qu'un instrument au service de l'intérêt privé ?

Ce paradoxe de théorie économique conduit à des problèmes importants pour la politique publique. Comme l'ont montré avec une certaine régularité les travaux de George

Akerlof, si le montant de la rémunération est un élément important pour expliquer l'effort, les normes sociales (Akerlof, 1982), la justice (Akerlof et Yellen, 1990) et l'identité sociale (Akerlof et Kranton, 2000) sont également centrales. Dans ce cadre d'analyse, s'il y a des différences de rémunération significatives entre des individus relativement homogènes, les moins bien payés peuvent réduire leur effort.

Traiter deux professionnels différemment pose un problème de conception de la justice et présuppose l'existence d'un sens moral. Ainsi, il est possible qu'un médecin altruiste réagisse violemment s'il existe des différences de rémunération importantes et cela d'autant plus que ces différences constituent une prime à l'égoïsme. Ce problème est particulièrement pertinent dans le système de santé français où certains médecins peuvent imposer des prix de marché alors que d'autres sont astreints à un prix administré (Batifoulier et al. 2011, Delattre et Dormont, 2003). En effet, les médecins de secteur 1 doivent pratiquer un prix administré tandis que les médecins de secteur 2 peuvent librement fixer leurs tarifs. Deux catégories de médecins sont alors créées avec différents niveaux de rémunération. Cette situation, initialement considérée non problématique, est devenue un problème majeur de politique publique depuis que les médecins les moins rémunérés cherchent à réduire les différences de rémunérations considérées illégitimes et injustes.

Paradoxe de théorie économique 4 : L'altruisme médical justifie et rend encore plus rentable l'égoïsme.

Problème de politique économique 4 : Il faut distribuer des primes à l'égoïsme qui risquent d'activer des réactions hostiles aux altruistes.

La formalisation de l'éthique médicale dans un cadre utilitariste pose de redoutables problèmes de théorie comme de politique publique. Il nous faut maintenant chercher à reconstruire une approche de l'éthique. Nous nous appuyons sur la notion de besoin en santé.

3. Besoins, éthique en santé et éthique médicale

Après avoir discuté les travaux de la TSE2 lorsqu'elle cherche à modéliser l'éthique médicale, il apparait un double problème. D'une part, ses formalisations engendrent des paradoxes de théorie économique et des impasses de politique économique. D'autre part, contrairement aux ambitions initiales de son programme de recherche, la TSE2 semble

incapable de montrer en quoi le champ de la santé est original par rapport au reste de l'économie – et donc, en quoi le médecin est un acteur atypique dans une économie. En réalité, le médecin de la théorie standard n'a pas d'épaisseur propre car il affiche un altruisme instrumental générique.

Dans cette section, nous proposons de repartir des critiques faites à l'hypothèse éthique de la TSE2 afin de proposer une autre définition de l'éthique médicale. Pour ce faire nous souhaitons réhabiliter un concept malmené dans la théorie économique : le besoin. Le patient n'est pas un consommateur comme un autre. La demande de soins médicaux ne dépend pas tant des préférences ou des désirs du patient que de ses besoins de santé. Les premiers sont volontaires et subjectifs. Les seconds sont largement involontaires et objectifs. Or, le fait pour le médecin de faire face à des désirs ou à des besoins n'a pas les mêmes implications éthiques.

Notre conviction est qu'il existe un lien entre cette originalité du bien santé – c'est un besoin – et l'existence d'une éthique médicale chez les professionnels de santé. Le besoin de santé est le point de départ de l'éthique médicale. En effet, le mauvais état de santé – la maladie – créé un besoin de services médicaux qu'il est d'autant plus impératif de satisfaire que l'empathie vis-à-vis du patient est grande. C'est donc l'empathie partagée à la vue du malade qui crée les conditions de construction d'une éthique en santé. La souffrance crée une obligation morale. Mais, cette obligation morale n'est pas propre au médecin, elle est partagée. C'est en cela que l'on parle d'éthique en santé et non d'éthique médicale. L'éthique médicale – c'est à dire l'éthique des médecins, et plus largement l'éthique des professionnels de santé – est dérivée de l'éthique en santé. La première se nourrit de la seconde pour construire les principes déontologiques propres à l'activité médicale. En conséquence, le besoin de santé favorise la création et le développement d'une conception originale du comportement médical – à mille lieues de l'*homo economicus*.

La suite de cette section propose une définition plus poussée du besoin en santé – concept au fondement de l'éthique en santé et de l'éthique médicale. Notre définition du besoin santé s'inspire de trois conceptions complémentaires qui forment le plan de la section : le besoin vital (3.1), le besoin objectif (3.2.) et le besoin social (3.3.). La dernière section (3.4.) dresse des conclusions sur les liens entre les différentes conceptions du besoin de santé, l'éthique en santé et – ce qui nous intéresse plus particulièrement – l'éthique médicale.

3.1 Le besoin vital

Dans la mesure où l'on cherche à appréhender la dimension éthique de la santé – et, par répercussion, l'éthique médicale –, une première approche pertinente pour définir la notion de besoin de santé peut être à trouver dans la philosophie morale (Batifoulier et al., 2010).

D'après le philosophe David Wiggins (2005), le besoin prend ses racines dans des considérations bien plus profondes que les désirs (ou les préférences). Il y a une discontinuité entre les désirs et les besoins. Cette rupture repose dans le rapport qu'entretiennent ces notions avec la réalité. Contrairement au désir, le besoin a pour caractéristique principale sa dimension objective. Les désirs s'appuient plutôt sur des croyances que les individus ont sur le monde – celles-ci n'ayant pas nécessairement de rapport avec la réalité : il est envisageable que l'objet du désir ne possède pas les attributs pour lesquels il est désiré. Par opposition, le besoin est quant à lui ancré puissamment dans le monde objectif.

Il peut cependant sembler parfois que le besoin ne soit doté que d'une dimension instrumentale. En effet, en se posant la question de ce que satisfait le besoin – l'objet du besoin, on aperçoit l'instrumentalité du besoin. Il est même souvent l'instrument de l'objet désiré, si bien que le besoin ne serait qu'un moyen pour une fin donnée. Par exemple, dans le cas de la santé, le patient a besoin de médicaments pour se soigner. Ou encore, les individus ont besoin d'argent pour partir en vacances. En en restant là, la tentation est grande de rejeter la notion de besoin : puisque le besoin n'est qu'un substitut à l'objet désiré, le besoin n'aurait pas d'intérêt en soi.

Au lieu de se débarrasser de la notion de besoin, Wiggins associe à sa dimension instrumentale le caractère de nécessité. C'est la définition substantielle du besoin. Avoir besoin de quelque chose, c'est qu'il est nécessaire de détenir cette chose sous peine d'éprouver une nuisance. Le besoin est donc défini par la recherche de l'évitement de la souffrance. Contrairement aux cas du besoin instrumental, ici, l'objet du besoin est l'évitement d'un mal rendu possible par la vulnérabilité universelle des êtres humains. Pour reprendre notre exemple, les médicaments permettent d'éviter un mal, l'argent sert à partir en vacances… et assez peu probablement à éviter un mal. Le besoin substantiel ne désire rien. Parmi l'ensemble des besoins (Encadré 2.6.), Wiggins nomme besoin vital (*vital need*) le type de besoin dont les individus ne peuvent être privés, même à court terme, sans quoi ils risquent de perdre leur pouvoir à vivre leur vie (*agentive power*). L'absence de satisfaction d'un besoin

de santé peut conduire à l'incapacité de mobiliser son corps ou son esprit et à se trouver dans un état de dépendance.

Encadré 2.6. Les différents types de besoin chez Wiggins (2005)

La notion de souffrance est source d'une importante variabilité et il reste difficile de définir un degré de souffrance acceptable. La question du temps de restriction est également essentielle. Peut-on se passer de l'objet du besoin un jour, deux jours, six mois, etc. ? Vaut-il-mieux souffrir un peu aujourd'hui et peu demain ou pas du tout aujourd'hui et beaucoup demain ? Afin de délimiter l'acceptable de l'inacceptable, Wiggins distingue trois différents seuils de besoins associés à des niveaux de souffrance différents.

- Besoins enracinés (*entrenched needs*) : ces besoins sont insensibles aux modifications de seuils. La souffrance de la personne privée de l'objet de son besoin est moralement inacceptable. À long terme, l'évolution de l'état des choses peut cependant avoir pour effet de réduire le degré de souffrance.

- Besoins basiques (*basic needs*) : pour des raisons environnementales ou relatives à la condition humaine, il n'est pas possible, même à long terme, d'envisager la privation de ces besoins. Il n'y a pas d'alternative à leur satisfaction bien qu'ils soient aisément substituables.

- Besoins vitaux (*vital needs*) : les besoin vitaux partagent les mêmes caractéristiques que les besoins basiques mais, en plus, ils ont la particularité d'être peu substituables.

Du point de vue de l'économiste, le besoin ainsi défini à l'avantage d'être moins fréquent que le concept de préférence. Cette restriction est en fait impérative pour la construction d'une justification morale à l'éthique en santé. Il est en effet plus facile et plus opératoire de construire la justification morale à partir du besoin substantiel – qui se réduit aux situations de souffrance – plutôt qu'avec le concept de préférence – potentiellement sans autre limite que l'imagination des individus.

L'originalité du bien santé peut alors être éclairé par la notion de besoin vital qui donne un fondement possible – il y en a d'autres – à la réflexion sur les besoins en économie de la santé (Batifoulier et al. 2010). Il ne s'agit donc pas de dire que tous les besoins de santé sont vitaux – sinon seuls les patients à l'article de la mort pourraient être présentés comme ayant un besoin de santé. Plus largement, il est possible de développer une conception étendue du besoin de santé, en définissant le besoin vital comme permettant d'éviter non seulement la souffrance, mais aussi l'incapacité de vivre sa vie de façon indépendante (Batifoulier et al. 2014).

Dans ce cadre d'analyse, l'ensemble des membres de la société se créent des obligations morales réciproques. Les besoins vitaux de santé étant des conditions de possibilité d'une vie libre et digne, il est impératif de les satisfaire. Encore une fois, ce dont il est question n'est pas seulement du médecin, mais de l'ensemble des composants de la société. Le besoin vital – au sens large – est alors le fondement d'une éthique dans le domaine de la santé.

3.3. Le besoin objectif

En complément de la notion de besoin vital, le concept de besoin en santé, que nous pensons être à l'origine de l'éthique médicale, peut se nourrir de la théorie du besoin objectif développée par l'économiste institutionnaliste Geoffrey Hodgson (2013). Dans son récent ouvrage *From pleasure machines to moral communities. An evolutionnary economics without homo economicus*, l'auteur défend que, même lorsque le *mainstream* reconnait les limites de l'application du cadre standard à l'analyse de la santé (demande induite, externalités, asymétries d'information, incertitude, etc.), il passe à côté de ce qui constitue l'originalité fondamentale du « bien » santé : c'est un besoin objectif. Le besoin de santé est objectif au sens du besoin vital car sa non-satisfaction entraine une souffrance mentale et/ou physique et une perte de liberté d'action – pouvant aller (mais ne se résumant pas) jusqu'à la mort. Mais, l'apport principal d'Hodgson par rapport à la philosophie morale est autre. Ce que montre son analyse est que le besoin de santé est un besoin objectif particulier car, contrairement à d'autres besoins objectifs (comme l'alimentation ou l'éducation, par exemple), il présente trois caractéristiques particulières qui exacerbent la commune humanité face aux problèmes de santé (Tableau 2.5.) : les besoins de santé sont inégalement distribués et involontaires ; les besoins de santé sont variés et idiosyncratiques ; les besoins de santé ont une considérable propension à évoluer.

Tableau 2.5. Comparaison des besoins de santé, d'éducation et d'alimentation

Besoin	Degré de distribution involontaire et inégale	Degré de variation et d'idiosyncrasie	Taux d'évolution du besoin
Santé	Fort	Fort	Fort
Education	Faible	Moyen	Moyen
Alimentation	Faible	Faible	Faible

Source : adapté d'Hodgson (2013, p. 195)

3.2.1. Les besoins de santé sont inégalement distribués et involontaires

Les besoins de santé sont aléatoires au sens où, dans une large mesure, l'occurrence de la maladie est aléatoire[34]. Il est impossible de savoir qui sera touché par la maladie. Ce fait – ou du moins cette croyance partagée – produit une forme d'universalité devant la maladie. Chacun peut être touché par la maladie et en souffrir. Les besoins de santé sont également

[34] Il existe néanmoins des déterminants sociaux de santé sur lesquels nous revenons plus loin.

involontaires. L'idée défendue ici énonce que personne ne décide intentionnellement d'être ou de ne pas être malade. Bien qu'il existe des comportements à risque, il est éthiquement difficile de porter un jugement sur la responsabilité du malade. Par exemple, il est peu probable d'entendre dire à une personne fumeuse dont on vient de diagnostiquer un cancer de la gorge, « c'est mérité ». L'une des raisons qui explique l'inconvenance liée à cette proposition provient du fait que l'occurrence de la maladie reste, quoi qu'il arrive, aléatoire : tous les fumeurs n'ont pas un cancer de la gorge et les non-fumeurs peuvent en être atteints.

On peut, nous dit Hodgson, comparer le besoin de santé à un autre besoin afin de prendre la mesure de son originalité. Si on s'intéresse, par exemple, au besoin d'éducation, on constate d'emblée que les exigences d'éducation sont mieux distribuées : tout le monde a besoin d'un niveau similaire[35] d'éducation. Pour prendre un autre exemple, les besoins d'alimentation sont distribués de façon relativement égalitaire entre les individus. Mais, dans le cas de la santé, la distribution est inégale. Tout le monde n'a pas les même besoins de santé et à cette inégalité ne correspond aucune responsabilité. Un individu sans problèmes de santé pourra toujours se dire : « cela aurait pu m'arriver ». Ces particularités du besoin de santé sont alors à l'origine de l'empathie et de la sympathie que l'on éprouve pour le malade.

3.2.2. Les besoins de santé sont variés et idiosyncratiques

Les besoins de santé sont variés dans la mesure où les besoins de santé sont très divers en fonction des individus. Encore une fois, le plus simple est de comparer avec d'autres besoins. Concernant l'éducation, les besoins des élèves sont hétérogènes notamment du fait des inégalités sociales mais, la variété entre élèves n'est pas de si grande ampleur qu'en santé. De façon analogue, même lorsque certains individus ont des besoins d'alimentation particuliers (à cause de leur corpulence, d'allergies, etc.), la variabilité des situations est finalement très faible comparativement au besoin de santé.

La variété des besoins de santé est si grande que ceux-ci sont idiosyncratiques. En effet, en matière de santé, les individus reçoivent d'une façon qui leur est particulière les influences des divers agents qui les affectent. La réponse au traitement est singulière parce qu'il existe toujours des différences physiques et mentales entre les individus. Le génome, l'environnement présent et passé, etc. sont des sources expliquant qu'à état de santé et

[35] Similaire et non identique dans la mesure où les inégalités sociales ont d'important effet en matière de réussite scolaire.

traitement de santé identiques, les résultats peuvent être divergents. En d'autres mots, il semble impossible de traiter tous les patients de la même façon[36].

3.2.3. Les besoins de santé ont une considérable propension à évoluer

Les besoins de santé ont la propriété d'évoluer de façon relativement rapide. Pour illustrer ce point, il suffit d'étudier les croyances des réformateurs sociaux ayant introduit les politiques de santé en France et en Angleterre dans la première moitié du XXème siècle. Ils pensaient assez unanimement que les besoins de santé n'évolueraient pas dans le temps. Mais au contraire, l'histoire des systèmes de santé montre que le vieillissement de la population et l'innovation technologique ont considérablement élevé le seuil de souffrance tolérable. Autrement dit, ce sont les besoins de santé qui se sont élevés.

Cette caractéristique n'est pas commune à toutes les formes de besoins. Le besoin d'alimentation est globalement le même aujourd'hui qu'hier. Les recherches et découvertes sur la question n'impliquent pas autant d'évolution que lorsqu'une maladie autrefois incurable trouve un traitement. Les besoins d'éducation sont eux plus évolutifs. Par exemple, la dernière décennie a vu l'apprentissage des outils informatiques modifier le contenu et la forme de certains apprentissages. Mais, globalement, les enseignements changent peu ou en tout cas moins vite que ce que l'on observe pour la santé.

L'approche du besoin de santé d'Hodgson est complémentaire à celle venant de la philosophie morale. Si dans le premier cas, le besoin vital insiste davantage sur la commune humanité vis-à-vis de la souffrance du malade et de l'impératif d'une bonne santé pour déployer ses capabilités, avec Hodgson, l'argument porte plus spécifiquement sur la commune humanité face à l'occurrence et la sévérité de la maladie. Chacun d'entre nous peut avoir des problèmes de santé d'une ampleur très variable sans que notre responsabilité individuelle soit en cause. On peut toujours dire d'un cancre qu'il ne travaille pas, il est éthiquement plus douteux de dire qu'un malade est responsable de sa situation. Selon nous, cette vulnérabilité commune est une autre source des considérations éthiques en santé.

Lorsque Hodgson discute de la forte propension à évoluer des besoins de santé, il nous amène à considérer la dernière dimension du besoin dont nous traiterons dans cette section : le besoin social.

[36] Nous reviendrons dans la troisième partie sur cette question.

3.4. Le besoin social

Le besoin de santé est vital (au sens large) et objectif. Partant de cette proposition, on pourrait postuler que le besoin est naturel, qu'il n'obéit pas à d'autres lois que celles de la physique et de la biologie. Penser cela serait une profonde erreur. En effet, la morbidité comporte une part de ressentie de la part des individus qui obéit à des normes sociales (Gadreau, 1978, Batifoulier, 2014). L'état de santé « malade » *versus* « non malade » est défini différemment selon les populations et les classes sociales. En conséquence, le besoin de santé est exprimé différemment selon la position sociale de celui qui juge l'état de santé.

L'idée selon laquelle les individus sont les seuls capables de juger leur propre état de santé et donc de leur besoin n'est pas exempte de critique. La plus importante d'entre elles est que, dans cette conception, tout ce passe comme si l'état de santé déclaré était strictement équivalent à l'état de santé réel. Cette hypothèse est loin d'être acquise, bien au contraire. Les travaux d'Etilé et Milcent (2006) montrent par exemple qu'à état de santé comparable on observe des écarts considérables dans les déclarations d'état de santé. La perception de l'état de santé et l'aspiration à une meilleure santé divergent fortement selon les données sociodémographiques des individus : par exemple, les personnes à haut revenu déclarent plus de maladies chroniques que les autres alors qu'ils sont ceux qui sont statistiquement le moins atteints. À l'inverse, les moins diplômés déclarent un meilleur état de santé perçu.

Ces données factuelles laissent penser que les préférences individuelles sont loin d'être un indicateur parfait des besoins – la définition de ceux-ci étant liée à des conditions sociales. Ces conclusions rejoignent celles mises en évidence par les travaux de Sen (1987) ou d'Elster (1985) sur les préférences adaptatives (Encadré 2.7.). Selon Sen, la formation des préférences a de grandes chances d'être biaisée par l'acceptation de la situation. Dans le cas des inégalités de genre qu'il étudie, l'auteur nous prévient :

> *« L'opprimé apprend si bien à supporter son fardeau qu'il ou elle finit par le négliger. Le mécontentement est remplacé par l'acceptation, la révolte sans espoir par le conformisme silencieux, [...] la souffrance et la colère par une endurance joyeuse. Comme les gens apprennent à s'ajuster aux horreurs existantes par l'absolue nécessité d'une survie sans surprise, les horreurs ont l'air moins terribles dans le système de mesure des utilités ». (Sen, 1984, p. 309 cité par Gilardone, 2009, p. 364)*

Le même raisonnement peut être tenu pour le besoin de santé. Les classes sociales dominées apprennent et intériorisent un état de santé qui leur parait naturel et inébranlable – du fait de leurs conditions d'existence entre contingence et déterminisme. En conséquence, si les personnes aux revenus les plus faibles déclarent moins de maladies chroniques, ce n'est

pas parce qu'elles en subissent moins, c'est parce qu'elles s'ajustent à leurs précarité sociale – conséquence directe des inégalités sociales.

Encadré 2.7. Les préférences adaptatives. Elster (1985) : le renard et les raisins

Selon fable d'Esope *Le renard et les raisins*, après avoir désiré fortement manger des raisins malheureusement trop hauts pour être attrapés, le renard s'en va en se disant que finalement les raisins n'étaient pas mûrs : les raisins sont trop verts. La psychologie sociale parle de dissonance cognitive c'est-à-dire de la simultanéité de deux propositions inconciliables provoquant une souffrance mentale. Les individus créent alors des stratégies de réduction de la dissonance dans le but de réduire cette souffrance. Dans notre fable, le renard veut quelque chose qu'il ne peut obtenir, c'est pourquoi il critique l'objet de son désir. Appliqué aux besoins de santé, un patient atteint d'une maladie incurable peut adapter ses préférences en critiquant la situation de non maladie ou en réduisant la peine qu'il éprouve. Dans les deux cas, il réduit la dissonance cognitive entre son désir de santé et sa situation de maladie. Par exemple, un patient atteint d'acouphènes pourra s'estimer heureux de ne pas être atteint de surdité ou trouver sa situation antérieure pas si désirable qu'on pourrait le croire *a priori*.

Prenant en compte le besoin social, il faut nécessairement réévaluer notre conception du besoin de santé. Dire que le besoin de santé est universel, c'est dire qu'il échappe aux données sociales et culturelles au sens où il n'est pas spécifique à certaines communautés, à certaines périodes ou à certains lieux. L'arrêt cardiaque est universel en ce sens : il peut s'attaquer à tous les êtres humains. On ne peut pas en induire pour autant que le besoin universel de santé ne reconnait pas l'existence de traits culturels, sociaux et historiques. Il y a une commune aspiration des humains à satisfaire les besoins de santé, indépendamment des cultures et des idéologies. Le besoin de santé ne peut donc pas être réduit à un besoin biologique, mais ce besoin biologique existe : il participe de notre commune humanité. La reconnaissance d'une commune humanité est alors un point de départ essentiel pour la construction – sociale et historique – d'un monde commun. Avec la notion d'universalisme du besoin de santé, nous cherchons simplement à défendre l'idée qu'il existe des besoins fondamentaux qui ne sont pas réductibles dans des particularismes sociaux.

La composante sociale du besoin produit une nouvelle source pour l'éthique en santé. Si l'on admet que les positions sociales importent dans la définition du besoin de santé, alors accorder plus d'attention et plus de ressources aux plus défavorisés devient une exigence morale. Cette inégalité de traitement ne peut pas être considérée comme une injustice car elle ne fait que corriger les inégalités sociales. Le besoin de santé tout comme sa variabilité doivent alors être extraits de l'illusion d'une scientificité parfaite se déclarant neutre et objective. Ce n'est pas uniquement l'expert ou le patient qui décident ce que sont les besoins de santé. Cette définition est à rechercher dans le commun – dans le politique.

3.5. Besoin de santé et éthique médicale

Notre définition du besoin de santé se nourrit de trois approches principales : le besoin est un besoin vital (au sens large) ; le besoin est un besoin objectif ; le besoin est un besoin social. Chacune d'entre elle fonde une exigence morale particulière qui pénètre l'éthique en santé (respectivement) : éviter la souffrance de l'autre et développer ses capabilités ; penser pouvoir être l'autre qui souffre et combattre la fatalité de la maladie ; reconsidérer les souffrances apprises et corriger les inégalités. Cette éthique en santé est d'après nous l'un des fondements – mais pas le seul[37] – de l'éthique médicale (Tableau 2.6.).

Tableau 2.6. Besoins de santé, éthique en santé et éthique médicale

	Besoin vital	Besoin objectif	Besoin social
Définition	La non satisfaction entraine la souffrance et/ou la perte de capabilité	La distribution des cas et leur sévérité est aléatoire dans la population totale	Il répond à des contingences culturelles sociales et historiques
Impact sur les exigences morales de l'éthique en santé	Eviter la souffrance de l'autre - Développer les capabilités	Penser pouvoir être l'autre qui souffre - Combattre la fatalité de la maladie	Reconsidérer les souffrances apprises - Corriger les inégalités
Impact sur l'éthique médicale	Le médecin doit faire preuve de bienfaisance et de non-malfaisance	Le médecin doit traiter tous les patients de manière identique (justice)	Le médecin doit corriger les inégalités sociales (justice)

L'ensemble des caractéristiques du besoin de santé que nous venons de décrire impliquent une atmosphère particulière dans le champ de la santé. Parce que les individus ne sont pas responsables de leur état de santé, la maladie créé un sentiment moral de sympathie universelle vis-à-vis du malade. Elle n'est en rien spécifique au médecin et/ou aux autres professionnels de santé, elle est commune à tous ceux qui partagent la condition humaine. Cette sympathie provient non seulement de la conscience de l'extrême précarité de l'humanité face à l'occurrence de la maladie mais aussi – et surtout – de la souffrance du malade. C'est parce le malade souffre que sa condition est inacceptable. Le traitement étant nécessaire à l'évitement de la situation de mal, une éthique commune se forge autour du besoin de santé.

[37] L'analyse ici réalisée est en effet assez lointaine de la médecine libérale traditionnelle – les principes humanistes dont il est question n'étant pas les seuls à fonder les préceptes éthiques de la médecine libérale (Batifoulier, 2011). Nous y reviendrons dans la troisième partie de la thèse.

Rappelons que l'éthique dont il est ici question n'est pas l'éthique des médecins mais, l'éthique en santé. C'est évidemment en partie d'elle que découle l'éthique des médecins, c'est-à-dire l'éthique médicale. Mais, la reconnaissance de la souffrance de l'autre – et donc du besoin de santé – est commune à tous.

Pour en arriver à l'éthique médicale, il est encore nécessaire de prendre en considération sa situation particulière dans le monde de la santé. Le médecin est un professionnel qui détient une expertise sur la maladie. Cette compétence est centrale puisqu'elle peut lui permettre de réduire la souffrance du malade. Or, la réduction de la souffrance – autrement dit la satisfaction du besoin – est impérative pour que la condition inacceptable que vit le patient cesse de s'offrir aux yeux de la société.

L'expertise du médecin lui créé donc une double obligation morale tant vis-à-vis du patient que du reste de la société : un grand pouvoir implique de grandes responsabilités. D'une part, le médecin a l'obligation de satisfaire le besoin du patient. En plus de son rôle d'expertise, dans une perspective de *care*, le médecin est la seule légitimité capable de reconnaitre la vulnérabilité du patient face à la maladie (Léonard, 2012). En reconnaissant l'état de souffrance, le médecin libère le patient de l'injonction à la bonne santé. Contrairement à ce que défend la TSE, le médecin ne profite pas de la souffrance du patient : il l'endosse.

Mais ce n'est pas tout. D'autre part, si le médecin doit prendre soin du patient c'est également pour réduire le niveau de souffrance de la société elle-même qui, par empathie, partage la situation de détresse du malade. Les caractéristiques particulières de la santé confèrent ainsi à l'expert de ce domaine – le médecin – une puissante responsabilité morale (avant même d'être juridique).

Selon nous, l'éthique médicale dérive de l'éthique en santé. C'est parce que la santé est un besoin universel que s'y créent des obligations morales. L'éthique médicale est une forme spécifique aux médecins qui, à partir de ce cadre général, s'est adapté pour encadrer les règles qui doivent s'appliquer à l'expert du domaine. C'est ainsi, par exemple, que l'éthique médicale a pour coutume de s'organiser autour des quatre principes/obligations fondateurs que sont la bienfaisance, la non-malfaisance, la justice et l'autonomie (Guillon, 1994).

L'autonomie est l'obligation consistant à respecter les choix du patient concernant sa propre maladie. Il s'agit de donner du poids à ses opinions et à ses croyances dans le

processus de décision médicale. Tant que les décisions du patient n'entrent pas en conflit évident avec la situation d'autres individus, le patient autonome doit rester le dernier juge. La justice est un principe obligeant le médecin à traiter de façon équitable les revendications des patients (dans la répartition des ressources rares, par exemple). La bienfaisance et la non malfaisance (ne pas nuire) impliquent plus directement l'obligation, respectivement, d'agir pour le bien-être du malade et de ne pas faire de mal au malade.

Ces obligations nous semblent découler assez directement de l'éthique en santé. C'est parce que le patient souffre et qu'il a un besoin de santé qu'il est nécessaire de lui apporter un bien (principe de bienfaisance) sans toutefois aggraver sa souffrance (principe de non malfaisance). Le respect de l'autonomie implique que, si chacun partage par empathie la souffrance du patient, c'est lui en dernier ressort qui doit être juge de sa situation. Enfin, le principe de justice vient reconnaitre à nouveau l'égalité des patients devant la maladie et donc leur commune dignité. Mais, en raison de la condition sociale diverses des patients, l'exigence de justice peut nécessiter une inégalité de traitement[38].

Conclusion

Quelles conclusions peuvent être tirées de l'introduction de l'hypothèse éthique dans la TSE ? Qu'est-ce que l'hypothèse d'éthique médicale apporte à l'hypothèse plus traditionnelle de maximisation des profits (TSE1) ?

Si l'on suit l'instrumentalisme méthodologique de Milton Friedman, la théorie économique ne doit pas être jugée sur le réalisme des hypothèses mais sur leur capacité à prédire les comportements. Par exemple, nous dit l'auteur, il peut être tout à fait pertinent de poser l'hypothèse que des joueurs de billard calculent avant chaque coup les angles et les trajectoires de la boule bien que l'on sache que ce n'est pas le cas. Les hypothèses n'ont pas besoin d'être réalistes et elles peuvent même être tout à fait contraires à la réalité tant qu'elles permettent de prédire. Cette position est dite instrumentale – instrumentalisme méthodologique – dans la mesure où les théories ne sont que des instruments pour faire des prédictions.

[38] Une illustration possible de cette inégalité de traitement est la pratique de discrimination par les prix (cf. Chapitre 1).

Dans ce cadre, l'hypothèse de maximisation des profits pour le médecin n'est pas surprenante. Mais, les économistes de la santé ont rapidement démontré l'incapacité de cette hypothèse à expliquer le comportement médical : si le médecin est un opportuniste pur, quelle limite trouve-t-on à l'induction de la demande ? L'introduction d'une éthique médicale permettant de ne pas faire du médecin un entrepreneur individuel comme un autre est non seulement plus réaliste mais aussi indispensable à la formalisation.

Cependant, cette éthique reste formalisée dans le langage de la théorie du choix rationnel et est assimilée à un altruisme instrumental. Dans ce chapitre nous avons évalué l'hypothèse d'éthique médicale formulée par la TSE2. Selon nous, en substituant l'hypothèse d'altruisme instrumental (TSE2) à l'hypothèse de maximisation du profit (TSE1) l'économie de la santé (de la TSE) perd en cohérence ce qu'elle gagne en réalisme sans toutefois produire des recommandations de politique publique viables.

En effet, le détour par l'histoire de l'économie de la santé nous montre que l'hypothèse d'un médecin uniquement motivé par son intérêt égoïste est désormais considérée comme une hypothèse irréaliste (tel ne fut toujours pas le cas[39]). Mais, l'hypothèse de maximisation du profit a l'avantage de ne pas provoquer de paradoxes de théorie économique ainsi que de produire des recommandations de politique économique très claires. Si les médecins ne sont motivés que par la maximisation de leur profit, ils doivent être financièrement incités par un mixte de capitation et de paiement à la performance[40]. Bien que cette position ne souffre pas de problème de cohérence, les recommandations de politique économiques sont généralement inefficaces ou, au mieux, partiellement efficace[41].

Le seul intérêt du concept d'altruisme médical est de « sauver » la théorie économique avec une hypothèse plus réaliste. Cependant, bien que cette hypothèse soit plus réaliste, elle induit des paradoxes de théorie économique. Si les problèmes théoriques n'intéressent que les théoriciens, la difficulté provient du fait que ces paradoxes théoriques sont à l'origine de problèmes de politiques économiques – eux très concrets.

Nous pensons que les paradoxes de théorie économique ainsi que les problèmes de politique économique sont étroitement liés dans la mesure où dans la formulation de la question posée par la TSE repose déjà sur des réponses prêtes à l'emploi. En conséquence,

[39] On pourra se reporter au chapitre 1 – notamment aux travaux de Kessel et Pauly.

[40] Cf. chapitre 1.

[41] Cf. chapitre 4 en ce qui concerne le paiement à la performance médicale.

une de nos conclusions est que la résolution des problèmes de politique économique requiert la sortie du cadre de la TSE. Ainsi, il devient crucial de reconsidérer en profondeur l'altruisme médical et l'éthique médicale. Le calcul utilitaire interdit toute considération éthique. En plaçant les motivations éthiques et les motivations personnelles au même niveau, il devient impossible de considérer l'éthique comme une norme sociale selon laquelle le médecin donne quelque chose « pour rien » et sans attendre de retour.

Comme l'a déjà montré Elster (1998)[42], si les normes sociales étaient comptées comme des utilités économiques, elles n'auraient aucun effet. Avec la plupart des normes, l'effet désiré est obtenu en recherchant un autre objectif. Lorsqu'il est appliqué au cas du médecin, ce raisonnement montre que, paradoxalement, ce n'est qu'en étant insensible à la gratitude du patient ou à l'approbation sociale, et plus généralement en adoptant un comportement éthique de façon désintéressée, que le médecin obtiendra l'approbation sociale et l'engament réciproque de son patient. En développant une éthique professionnelle, le médecin s'attend à un retour du patient. Il n'y a pas d'éthique gratuite. Mais la qualité de cette éthique réside dans le fait qu'elle ne requiert aucun retour. L'éthique n'est payante qu'à la seule condition de ne pas être faite dans ce but, comme l'enseigne le langage et la sociologie du don (Caillé, 2005 ; Chanial, 2008) ou l'économie de la santé hétérodoxe (Davis et Mc Master, 2013, Hodgson, 2008). Une source possible pour discuter des considérations éthiques dans le domaine de la santé réside alors dans la nature universelle du besoin de santé.

[42] L'exemple de la pilule de non culpabilité d'Elster peut être adapté au cas du médecin pour montrer que personne ne peut échapper à sa « mauvaise conscience » par l'achat de cette pilule, puisque vouloir être immoral est être immoral : pour penser à la pilule, il faut déjà avoir un sentiment de culpabilité (Batifoulier et Thévenon, 2003). Le problème moral n'est donc pas strictement réductible à un problème de coût, qui mettrait en équivalence des comportements aux qualités morales fortement opposées.

Chapitre 3 :

Payer plus pour motiver moins ? Performance médicale et *crowding out effect*

Introduction

L'économie *mainstream* peine à intégrer l'éthique médicale dans ses modèles. Les recommandations de politique publique émanant de ces travaux deviennent alors problématiques et le champ est laissé libre pour les préconisations traditionnelles – le paiement à la performance. Pourtant, il subsiste chez les économistes l'idée que le médecin n'est pas un acteur comme les autres. Issu des travaux de psychologie sociale (Deci et Ryan, 2002), le succès récent de la thématique de l'effet d'éviction (*crowding out effect* ; COE) des motivations intrinsèques (MI) par les motivations extrinsèques (ME) en est un témoin. L'idée est relativement simple : lorsqu'un individu intrinsèquement motivé – par l'intérêt de l'activité elle-même – se voit soumis des ME – la motivation est instrumentale : on recherche l'argent, la réputation, etc. –, les motivations extrinsèques peuvent évincer les MI. Au final, contrairement à ce qu'apprend la théorie économique, payer plus peut motiver moins.

La thématique d'un effet d'éviction des motivations intrinsèques a été récemment développée en économie de la santé pour réexaminer l'efficacité des dispositifs incitatifs. Quand les médecins sont sensibles à des motivations intrinsèques, il n'est pas impossible qu'une récompense financière altère la signification de l'acte médical aux yeux du médecin comme l'ont souligné un nombre grandissant de travaux en France (Videau, 2010, Ventelou et Videau, 2011)[43]. Ainsi, le paiement de tâches jusqu'alors effectuées gratuitement peut être interprété comme un défaut de confiance par le médecin et engendrer un certain désengagement, paradoxalement source de dépenses supplémentaires (Batifoulier et Ventelou, 2003, Batifoulier et al., 2009, Lievaut, 2012). L'étude des motivations médicales dans le domaine de la prévention, n'a pas permis d'établir empiriquement l'existence d'un *crowding*

[43] Le chapitre 4 propose une revue de littérature empirique du *crowding out* médical.

out effet. Mais, elle a révélé que l'efficacité d'un dispositif incitatif n'est pas garanti alors même que le paiement à la performance se développe (Videau et al. 2010, 2012).

Cette question de la motivation au travail est au cœur de nombreux développements au sein de la Gestion des Ressources Humaines dans la lignée des travaux de Mayo (1945) et visant à prendre en compte la hiérarchie des besoins (Maslow, 1943), la responsabilisation et l'implication des salariés (McGrégor, 1969) ou l'enrichissement du travail (Herzberg, 1971)[44]. Dans le cas particulier du médecin, elle prend toutefois un sens particulier car il peut être considéré comme l'archétype du professionnel intrinsèquement motivé puisque son activité paraît être d'un grand intérêt pour lui comme pour le patient qui lui confie sa santé. Dès lors, n'y-a-t-il pas un risque à mettre en place le paiement à la performance ? Inciter le médecin engendre-t-il une réduction de ses motivations non instrumentales (MI) ? L'originalité du champ justifierait de penser d'autres modes de rémunération moins ostensiblement instrumentaux – tel le salariat par exemple.

Dans une certaine mesure, la littérature sur les motivations médicales rejoint celle sur l'éthique médicale : il s'agit de ne pas faire du médecin un *homo economicus* comme un autre. Il est possible de questionner la pertinence d'un nouveau chapitre sur l'intégration de « préférences sociales » dans la théorie du choix rationnel. Les apports de cette nouvelle littérature peuvent sembler limités. Quelle différence faire entre l'analyse de l'éthique et celle des motivations ? En quoi la théorie des motivations justifie-t-elle un traitement spécifique ? Deux raisons principales nous conduisent à faire le choix d'un chapitre à part entière.

D'une part, l'entrée par les motivations permet d'introduire l'hypothèse de substituabilité entre « motivations sociales » et « motivations financières ». Quand bien même il existerait une éthique médicale, celle-ci n'est pas contradictoire avec l'intéressement de la théorie du choix rationnel. L'incitation financière ne peut qu'accroître la motivation des médecins mus par l'éthique. Inciter de façon indifférenciée les altruistes et les égoïstes permet de hausser la motivation de tous[45]. Avec l'hypothèse de COE, la menace d'effet pervers du paiement à la performance – et de l'incitation en général – est bien plus sérieuse : les altruistes peuvent être démotivés par l'incitation.

Il existe une deuxième différence d'importance entre motivation et éthique. L'éthique renvoie à des jugements moraux qui impliquent nécessairement le social. Or, comme on l'a vu

[44] Pour une analyse de ces différentes théories de la motivation, voir notamment Levy Leboyer (1994, 2006) et Roussel (1996).

[45] Une première critique de ce point de vue a été donnée à la section 2.2.4. (prime à l'égoïsme).

au chapitre précédent, la TSE2 est mal à l'aise avec le social d'où son intérêt pour la thématique des motivations et son inclinaison, *a priori*, individualiste. La théorie économique *mainstream* a davantage d'affinité avec la psychologie qu'avec la sociologie. La psychologie de la motivation est une échappatoire plus naturelle pour sauver la TSE2 et la thématique incitative. Il nous faudra cependant faire attention à la différence entre ce que le *mainstream* voit dans la psychologie et la source d'inspiration légitime qu'elle peut être. Ce chapitre propose de montrer comment, appliquée au cas du médecin, la théorie des motivations est utilisée par le *mainstream*.

Ces deux arguments nous semblent suffisants pour justifier un chapitre supplémentaire dédié à la question des motivations. Notons avant de poursuivre que ce champ de la littérature est relativement récent et qu'il porte avant tout sur la motivation du salarié et peu de travaux s'intéressent explicitement aux motivations du médecin. Nous prenons toutefois ici la liberté de considérer que le médecin ne représente qu'un cas particulier et nous appliquerons donc la problématique de la motivation au travail au cas de la motivation médicale[46].

La difficulté majeure dans ce chapitre provient du fait que l'intégration des motivations dans le cadre la théorie économique s'est faite sous deux formes qu'il importe de distinguer.

(i) La première méthode est attachée à l'auteur principal ayant introduit la thématique des motivations dans les débats économiques : Bruno Frey. Dans ses travaux précurseurs sur les hypothèses de MI et de COE, l'économiste présente sa méthode comme relevant de l'inspiration plutôt que de l'impérialisme (Frey et Benz, 2002).

(ii) Comme pour l'éthique médicale, le *mainstream* s'est peu à peu approprié les questionnements sur les motivations en les incorporant à l'hypothèse d'optimisation. Les MI remettent en cause l'intéressement et ces économistes vont essayer de prendre en compte les comportements désintéressés avec des modèles comme ceux de Bénabou et Tirole (2003, 2006). L'origine du *crowding out effect* serait alors à rechercher dans les problèmes d'information.

[46] Depuis le premier chapitre nous constatons qu'en dépit de la reconnaissance d'une originalité du médecin l'économie de la santé *mainstream* n'est qu'une application d'un cadre général à un problème particulier. Le médecin est mis au centre d'une relation d'agence (chapitre 1) et on lui prête des préférences altruistes qui n'ont finalement rien de particulier (chapitre 2). En ce qui concerne la motivation, les rares textes existant démontrent que la même procédure est à l'œuvre.

Afin de prendre au sérieux l'influence des MI médicales dans la compréhension du champ, la première section de ce chapitre présente les travaux de psychologie sociale d'où est issue l'hypothèse de COE. Cette dernière provient des travaux anti-utilitaristes de la Théorie de l'Auto-Détermination et, plus précisément, d'un sous-ensemble nommé Théorie de l'évaluation cognitive (1.). La section suivante présente l'originalité des travaux de « passeur » de Frey (2.). Le rôle des motivations intrinsèques est à même de remettre en cause les résultats les plus solidement acquis de l'économie du travail. Payer à la performance peut démotiver. La dernière section fait la critique de la TSE2 lorsqu'elle cherche à traduire les motivations médicales dans le langage des asymétries d'information (3.). Malgré le potentiel subversif de l'hypothèse d'éviction des motivations intrinsèques par les motivations extrinsèques, la réduction du COE à un problème d'information génère des paradoxes de théorie économique et des problèmes de politique publique.

1. Aux racines du *crowding out effect* : la théorie de l'autodétermination

Cette thèse n'est pas une thèse de psychologie de la motivation, ni même une thèse d'économie de la motivation. Pour cette raison, il n'est pas d'intérêt majeur de réaliser une revue de la littérature sur les théories de la motivation en psychologie sociale qui permettrait de replacer plus complètement la Théorie de l'Auto-Détermination (TAD) dans les enjeux théoriques qui lui sont propres. Nous nous contenterons ici de présenter les éléments que nous jugeons les plus pertinents de la TAD – l'objectif étant d'évaluer la façon dont le *mainstream* a inséré cette théorie dans son corpus théorique.

Pour définir le concept de motivation, la littérature en psychologie de langue française s'appuie généralement sur la définition de Vallerand et Thill (1993). Nous proposons de donner ici celle de Fenouillet (2012) qui est l'actualisation de la définition précédente :

> *« La motivation désigne une hypothétique force intra-individuelle protéiforme, qui peut avoir des déterminants internes et/ou externes multiples, et qui permet d'expliquer la direction, le déclenchement, la persistance et l'intensité du comportement », (Fenouillet, 2012, p.9).*

Pour ce qui nous concerne, il est intéressant de noter à quel point, à première vue, la définition de la motivation est éloignée de la notion d'éthique. Alors que celle-ci suppose des jugements moraux portant sur le collectif, donc un impact normatif du social (Joyce, 2006), la motivation se conçoit comme un comportement individuel qui, éventuellement, est marqué

par des déterminants externes. Le questionnement porte sur les causes – exclusivement individuelles ? – qui produisent des effets sur le comportement (direction, déclenchement, persistance et intensité). Les théories de la motivation produisent ces réponses (Fenouillet, 2012).

Nous allons voir que la TAD est non seulement une théorie de la motivation mais aussi une théorie du bonheur (1.1.). Le bonheur y est conçu comme provenant de la satisfaction de trois besoins psychologiques fondamentaux eux-mêmes déterminant la MI (1.2.). Pour en venir au *crowding out effect*, il faut faire un pas supplémentaire et comprendre que la TAD est un ensemble de mini-théories dont l'une d'entre elles – la théorie de l'évaluation cognitive – énonce les conditions sous lesquelles les ME peuvent évincer les MI (1.3.).

1.1. La TAD comme théorie du bien-être

Historiquement, la grande majorité des recherches en psychologie avait pour objet la définition chez les individus de ce qui ne va pas psychologiquement (Cowen, 1994). Les études se donnaient un triple objectif : repérer les psychopathologies, comprendre les processus sous-jacents au déclenchement de la maladie et trouver des remèdes pour soigner le patient. Par la suite, l'intérêt des chercheurs s'est porté sur les déterminants de la bonne santé mentale. Désormais, l'hypothèse principale est que l'étude des lois d'une bonne santé mentale est l'étape nécessaire pour comprendre les états psychopathologiques. De l'étude de la maladie, le centre d'intérêt des psychologues s'est déplacé vers celle de la santé[47]. Bien que les recherches s'accordent sur cet objet commun, les avis divergent quant au chemin à emprunter pour atteindre l'état de santé – de bien-être psychologique. Il est ainsi courant de différentier deux conceptions distinctes des moyens disponibles aux individus pour parvenir à la situation de bien-être : l'hédonisme et l'eudémonisme (Ryan et Deci, 2001).

La tradition hédoniste du bonheur est l'approche dominante en psychologie. Elle suppose que le bien-être subjectif consiste à vivre beaucoup de situations agréables et peu de situations désagréables. Selon Diener (2000) la bonne vie peut être recherchée par le biais du bien-être subjectif (*Subjective Well-Being* ; SWB). La mesure du SWB se fait par des questionnaires où l'on demande aux individus s'ils vivent des situations de bonheur, à quelle fréquence et pour quelles raisons. Cette mesure subjective serait justifiée par un impératif

[47] Notons que cette évolution propre à la psychologie reprend de façon assez analogue la définition de la santé de l'Organisation mondiale de la santé : « la santé est un état de complet bien-être physique, mental et social, et ne consiste pas seulement en une absence de maladie ou d'infirmité », (OMS, 1946).

démocratique : qui mieux que les individus eux-mêmes peuvent décider de ce qu'est leur bien-être ?

Cependant, le courant développé par Diener et ses collègues pose une série de problèmes irrésolus. Cette position philosophique en psychologie conduirait notamment à la substitution du politique par le psychologique. En effet, si le bien suprême est la recherche de plaisir, alors seule doit compter la science qui dit savoir mesurer les plaisirs : la psychologie[48]. Au lieu de discuter collectivement du bien par le politique, il suffit de déléguer au psychologue la recherche du plaisir. La psychologie hédoniste ferait du même coup table rase du passé et nierait l'essence même de la vie humaine, la quête de sens (Laguardia et Ryan, 2000).

Contre cette approche, il est apparu nécessaire aux psychologues de la TAD de chercher une autre conception philosophique sur laquelle bâtir une nouvelle théorie du bien-être : c'est la tradition eudémoniste en psychologie (Deci et Ryan, 2000). L'eudémonisme est la doctrine philosophique qui postule que le souverain bien, le but ultime de toute activité humaine, est la recherche du bonheur. La version aristotélicienne de l'eudémonisme affirme cependant que l'erreur fondamentale des hédonistes est de définir le bonheur par l'intermédiaire des plaisirs (van Bosch, 1997). Le problème provient du fait que la satisfaction des plaisirs et l'évitement des déplaisirs sont un état instantané. Or, comme il serait impossible de vivre toute sa vie une succession d'états positifs, les insatisfactions qui ne manquent pas d'advenir interdisent un état de bonheur durable.

Le bonheur, selon Aristote relu par les psychologues de la TAD, ne peut pas être un état passif mais une activité. La question de la recherche du plaisir devient annexe, et, ce qui compte est de trouver l'activité menant au bonheur, celle qui est la fin de l'homme. Une telle activité ne peut-être le désir puisque même les animaux en sont capables. L'activité qui distingue l'homme des autres êtres vivants est la pensée. Dès lors, le bonheur eudémonique correspond à une situation où les individus font concorder leurs actions et leurs aspirations (*daimon*).

Ainsi, on peut distinguer l'eudémonisme de l'hédonisme dans la mesure où ce dernier ne requiert pas la poursuite d'activités « qui développent la croissance personnelle et la réalisation du soi » (Laguardia et Ryan, 2000, p.283). En d'autres mots, certaines actions, bien

[48] La section 2 de ce chapitre revient largement sur cette affirmation à partir d'un développement à propos de l'économie comportementale qui se veut être un rapprochement entre psychologie et économie. En ce sens, l'économiste peut également postuler au titre de conseiller du prince en bonheur du peuple.

qu'elles soient plaisantes augmentent le bonheur instantané – le plaisir – sans contribuer au bien-être psychologique – et donc au « vrai » bonheur. Dans cette conception anti-utilitariste, le « bonheur », entendu comme la somme des plaisirs et déplaisirs, ne conduit pas au bien-être psychologique – même s'il peut y amener en tant que résultat secondaire. Reste alors à définir pour la TAD ce qui permet de développer la « réalisation du soi ».

1.2.Besoins psychologiques fondamentaux et motivation intrinsèque

D'après les travaux de la TAD, la réalisation de soi est le critère essentiel du bien-être et ce sont les Besoins psychologiques fondamentaux qui permettent de fournir les conditions minimales de ce bien-être psychologique (Baumister et Leary, 1995 ; Reis et al., 2000) – par la concordance au *daimon*. Les auteurs entendent par besoin ce qui est essentiel à la vie au sens où nous sommes construits pour répondre à ces besoins. La TAD soutient que les individus, d'une façon universelle, cherchent à satisfaire trois besoins psychologiques fondamentaux (Deci et Ryan, 2002)[49] :

- Besoin de compétence : le besoin de compétence se matérialise par le fait que les individus cherchent à se sentir utiles dans les interactions sociales et cherchent des opportunités de l'être. Il conduit les individus à trouver des objectifs compatibles avec leurs compétences afin d'entretenir ou d'améliorer leurs capacités. Il ne s'agit pas de qualifier les capacités elles-mêmes mais plus tôt de juger le sentiment de confiance dans la réalisation de l'action. C'est un sentiment d'efficacité sur son environnement qui autorise la stimulation de la curiosité et donc la croissance et le développement psychologique optimal.

- Besoin de connexion (*relatedness*) : il s'agit du sentiment d'être connecté, en lien, avec les autres, de s'en occuper et d'être l'objet de l'attention. Les individus doivent alors avoir le sentiment d'appartenir à une même communauté. Ce besoin n'est pas instrumental au sens où il n'est pas recherché comme une récompense (comme le statut social par exemple) mais plutôt comme un sentiment d'unité sociale. C'est un besoin de relation avec autrui d'après lequel il existe un sentiment d'appartenance à des personnes qui sont importantes pour soi.

- Besoin d'autonomie : le besoin d'autonomie recouvre l'idée que l'action doit provenir originellement de la volonté individuelle. Elle doit être perçue comme l'expression du soi et non comme une contrainte extérieure. Il ne faut pas confondre autonomie et indépendance.

[49] Bien qu'essentiels pour le bien-être, les besoins psychologiques fondamentaux n'épuisent pas – loin s'en faut – l'éventail des besoins et motivations humaines. Leur place est centrale dans la mesure où leur satisfaction est nécessaire pour parvenir à une situation de bien-être.

L'indépendance est définie par l'absence de lien ou d'influence extérieure alors qu'une action autonome peut être guidée par une autre source que le soi. L'individu décide volontairement de son action : l'action autonome est celle qui émane du « vrai soi » et dont la causalité est perçue comme venant « de l'intérieur ».

Dans une perspective eudémoniste, la congruence avec le « vrai soi » est définie par l'intermédiaire des besoins psychologiques fondamentaux. Pour que l'individu ait une chance de parvenir au bien-être psychologique, il doit satisfaire ses besoins psychologiques fondamentaux et c'est l'atteinte de ce bien-être qui favorise la situation de bonheur. Si ces besoins sont négligés alors il s'en suivra un impact négatif sur la croissance, l'intégrité et le bien-être (Deci et Ryan, 2000).

La notion de besoin psychologique fondamental permet à la TAD de relier les buts de l'action au bien-être. Il ne reste plus alors qu'à trouver pourquoi les individus poursuivent tel ou tel but : qu'est ce qui sous-tend la motivation ? Le « pourquoi » du comportement porte sur ce locus de causalité de l'action et distingue la motivation intrinsèque de la motivation extrinsèque (Deci et Ryan, 2000) :

- Motivation intrinsèque (MI) : un individu est intrinsèquement motivé lorsqu'il s'engage dans une activité pour le plaisir et la satisfaction inhérente à l'activité elle-même. Un médecin motivé intrinsèquement va travailler en premier lieu pour le plaisir que lui procure son activité. Il aime soigner les malades, il apprécie le défi intellectuel, etc.

- Motivation extrinsèque (ME) : un individu est extrinsèquement motivé lorsqu'il s'engage dans une activité de façon instrumentale. L'action est entreprise pour des raisons différentes de l'activité elle-même. Un médecin extrinsèquement motivé va travailler avant tout pour les honoraires, le prestige social associé à son activité, etc. La médecine est dans ce cas un moyen pour autre chose.

- Amotivation (AM) : il s'agit d'une situation où l'individu n'attend de l'action ni récompense ni satisfaction directe. L'individu se sent alors incompétent et sans contrôle. Souvent l'amotivation précède l'abandon de l'activité en question. Le médecin ne ressent rien pour son activité : il fait médecine comme il pourrait faire n'importe que autre métier.

Selon Deci et Ryan (2002), la MI est celle qui est en accord avec le « vrai soi » (le *daimon*). Il existerait dans cette optique une importante relation entre les besoins psychologiques fondamentaux, la MI et le phénomène de bien-être psychologique. On peut alors imaginer des désirs, des buts, qui, même s'ils sont satisfaits, ne répondent pas aux

exigences des besoins psychologiques fondamentaux et n'entrainent pas la situation de bien-être. La satisfaction des désirs, lorsqu'ils s'opposent aux besoins psychologiques fondamentaux, peut entrainer, paradoxalement, une situation d'insatisfaction. En effet, l'une des conséquences de l'eudémonisme est que tous les buts ne sont pas également bons (Sheldon et Kasser, 1998). Les individus ne peuvent plus faire le raccourcit plaisir-bonheur, il est nécessaire de se poser la question du sens de l'action.

Les psychologues de la TAD distinguent alors deux types de buts : les buts fondamentaux sont des buts où les actions entreprises viennent du « moi intérieur » et sont, en conséquence, celles qui satisfont le plus les besoins psychologiques fondamentaux. On dit, dans ce cas, que la motivation est intrinsèque : l'action est une émanation du « soi intérieur », elle est motivée par la satisfaction inhérente à son accomplissement. À l'inverse, les buts compensatoires sont extrinsèquement motivés et n'apportent pas ou peu de satisfaction. Ils n'aident pas à la satisfaction des besoins psychologiques fondamentaux et au bien-être – même s'ils peuvent être une source momentanée de plaisir. En résumé, les buts fondamentaux sont la seule solution pour améliorer le bien-être psychologique alors qu'un accent mis sur des buts hétéronomes et extrinsèques est associé à une baisse de la santé (Deci et Ryan, 1985, 2002).

Après ces développements préliminaires, il reste maintenant à en venir au *crowding out effect* – par l'intermédiaire de la théorie de l'évaluation cognitive.

1.3. La théorie de l'évaluation cognitive

Le concept de besoin psychologique fondamental est le fondement sur lequel repose toutes les travaux se réclamants de la Théorie de l'auto-détermination. En ce sens, la TAD est constituée d'une multitude de mini-théories qui partagent ces principes communs (Deci et Ryan, 2002). À titre d'exemple, en plus de la Théorie de l'évaluation cognitive que nous allons maintenant présenter, la TAD comprend également la théorie de l'intégration organismique (Encadré 3.1.), le modèle hiérarchique de la motivation (Vallerand et Miquelon, 2008), etc.

Encadré 3.1. La théorie de l'intégration organismique. Deci et Ryan (2002)

Selon Deci et Ryan (2002), la dichotomie MI/ME est trop manichéenne c'est pourquoi ils introduisent dans leur théorie la possibilité de rendre autonome une action au départ extrinsèquement motivée (et inversement). La théorie de l'intégration organique est basée sur l'hypothèse que les individus sont naturellement enclins à intégrer leurs expériences. La distinction MI et ME n'est pas une dichotomie mais un continuum. Il existe donc différents types de régulation du comportement selon le degré d'autonomie perçu de l'action.

Régulation externe : il s'agit de la forme la moins autonome de régulation externe où l'acteur va simplement chercher à éviter une punition ou à obtenir une récompense. C'est la forme que l'on sous-entend lorsque l'on parle sans précisions de ME.

Régulation introjectée : c'est le premier niveau d'internalisation où l'individu essaye d'éviter les sentiments comparables à la honte, la culpabilité etc. Il n'est question que de l'estime de soi par rapport au regard d'autrui, de la réputation.

Régulation identifiée : d'un niveau d'internalisation supérieur, l'individu est conscient des valeurs en jeu dans l'action et juge celles-ci comme personnellement importante. On franchit un pas important vers l'autodétermination puisque le besoin d'autonomie est alors fortement comblé.

Régulation intégrée : elle relève de la régulation extrinsèque la plus autonome. L'individu met en accord ses valeurs et ses attentes personnelles avec l'activité qu'il entreprend. Mais, à la différence de l'action totalement autorégulée (MI), l'action est tout de même accomplit en vue d'une récompense extérieure.

La théorie de l'évaluation cognitive est la seule chose que la littérature économique *mainstream* retient des travaux autour de la TAD : c'est d'elle que vient l'hypothèse de *crowding ou effect*. La théorie de l'évaluation cognitive a pour objet l'analyse des MI lorsque sont introduites des ME. Si l'on imagine un médecin agissant de manière intrinsèque, quelle sera sa réaction lorsqu'une récompense (ou une punition) est introduite ? Les MI étant fondées sur l'existence de besoins psychologiques fondamentaux, l'introduction de ME peut altérer la satisfaction de ces besoins. Plus particulièrement, selon la théorie de l'évaluation cognitive les besoins de compétence et d'autonomie peuvent être affectés par l'introduction de ME. La mise en place par l'environnement de ME va être interprétée par l'individu. Deux situations sont possibles :

- Le mécanisme de *perceived competence* est lié au besoin de compétence. La mise en place de ME signifie pour l'individu contrôlé que le régulateur ne fait pas confiance en ses capacités. Il y a remise en cause de la compétence de celui qui agit. L'acteur voyant son action dévalorisée risque très probablement de baisser son niveau de MI. Au final, la ME aura évincé la MI ceci ayant pour résultat final probable une dégradation de la motivation totale.

- Un second mécanisme d'éviction existe qui a pour objet central le besoin d'autonomie. Il s'agit du mécanisme de *perceived locus of causality*. Encore une fois, l'introduction de régulation extérieure (ME) va engendrer une interprétation de l'acteur. Cette fois-ci il va penser que le régulateur ne reconnaît pas son implication dans l'activité (sa MI). L'action ne serait effectuée que du fait de la ME et non pas de la volonté autonome de l'acteur. En d'autres mots, l'individu ne serait pas la cause originelle de l'activité. De la même manière cette

interprétation, qu'elle soit justifiée ou non par l'intention du régulateur, implique une modification du niveau MI.

Ces deux mécanismes d'interprétation (*perceived*) sont au cœur de la théorie de l'évaluation cognitive. L'extérieur peut influer de deux manières sur les MI. Soit les ME ont un caractère de contrôle et elles sont interprétées comme aliénant l'autonomie et la compétence. Soit les ME prennent la forme d'information et peuvent alors encourager les MI. Or, il n'existe pas de règle permettant de qualifier telle ou telle pratique de contrôlante ou non. La même régulation peut, selon le contexte et les individus en question, être interprétée comme une forme de contrôle ou comme une information amicale.

Concernant la vérification empirique de la théorie de l'évaluation cognitive[50], il existe une très volumineuse littérature depuis l'article fondateur de Deci en 1971. Il est courant de distinguer trois types de travaux empiriques. D'abord, toute une série de travaux ont pour objet de démontrer que les récompenses extrinsèques et autres récompenses environnementales ne remettent en rien en cause les MI (par exemple, Flora, 1990). Lorsque c'est le cas, les effets négatifs des récompenses ne seraient pas gênantes parce que transitoires (Dickinson, 1989). Ces travaux se rattachent aux théories traditionnelles de la motivation et sont les descendants directes du behaviorisme (Skinner, 1938, Vroom, 1964, Atkinson, 1964). Le deuxième type de travaux provient des auteurs de la TAD (Deci, 1971, Deci, 1975, Deci et Ryan, 1985). Pour eux, les ME peuvent nuire aux MI. Enfin, un troisième type de travaux portent sur la théorie de l'attribution (Lepper et al., 1973, Lepper et Greene, 1978). À partir d'expériences empiriques, les auteurs concluent que rémunérer un individu intrinsèquement motivé revient à lui fournir une sur-justification à son comportement. L'*overjustification effect* stipule alors que les ME peuvent avoir un effet négatif sur les MI. Dans ce cadre, l'individu perd le sens qu'il attribuait à la tâche : il ne la réalise plus parce qu'il est intrinsèquement motivé mais parce qu'il est incité pour le faire.

Si cette controverse a occupé de longues années les psychologues, il semble que depuis la revue de littérature réalisée par Deci et al. en 1999, un consensus semble se faire autour de l'idée que les récompenses ne sont pas nécessairement motivantes. Elles peuvent avoir des effets très variables selon les contextes et la modalité de ME. Les récompenses intangibles, comme les récompenses verbales et autres retours (*feedbacks*) positifs, ont un effet positif sur

[50] Il faut distinguer celle-ci de l'évaluation de l'hypothèse de *crowding out effect*. Bien que les deux hypothèses empiriques à tester soient similaires, nous nous intéressons ici toujours aux travaux issus de psychologie de la motivation. Dans la section 2, nous proposons une présentation des résultats empiriques de l'hypothèse d'éviction – cette fois-ci à partir de travaux réalisés par des économistes.

la MI dans la mesure où ces récompenses stimulent le besoin de compétence de l'individu. Si on dit à un doctorant qu'il fait bien son travail, il est probable que sa MI en soit bénéficiaire parce qu'il se sent compétent dans son activité.

Les récompenses tangibles non anticipées sont des récompenses distribuées sans que les individus n'en aient été informés préalablement. C'est par exemple une prime exceptionnelle récompensant l'atteinte d'un objectif. La plupart des travaux concluent que ce type de récompense n'affecte pas la MI. Seuls les travaux de Lepper et al. (1973) sur la théorie de l'attribution trouvent que les récompenses tangibles non anticipées peuvent avoir un effet sur les MI. Cet effet dépendrait du niveau de MI : si celui-ci est bas, les MI peuvent croitre ; si celui-ci est haut, les MI peuvent chuter.

Les récompenses tangibles anticipées et contingentes à la réalisation d'une tâche sont proposées avant que ne commence la tâche. Les individus savent donc qu'ils peuvent prétendre à une récompense complémentaire. Lorsque ces récompenses sont contingentes à la réalisation d'une tâche – c'est-à-dire qu'elles sont liées à un objectif uniquement quantitatif comme le paiement à la pièce – un consensus émerge dans la littérature : les effets sur les MI sont plus probablement négatifs. Produire des systèmes d'incitation anticipés et contingents réduit la MI.

Dans le cas des récompenses tangibles anticipées contingentes à un niveau de performance, les objectifs dont il est question ici comportent une dimension quantitative et qualitative. Il s'agit de paiement à la performance. L'évaluation de ce type de récompenses pose un problème d'évaluation important dans la mesure où il implique deux étapes : (i) une procédure d'évaluation comprenant notamment la mise en place de retours sur le taux de réalisation des indicateurs (ii) le versement de la récompense monétaire. Or, il est impossible d'isoler statistiquement l'effet du versement (le paiement à la performance) du reste de la procédure. En conséquence, les psychologues ont recours à des expériences avec deux groupes : l'un soumis au paiement à la performance et l'autre ne recevant uniquement les retours. La méta-analyse de Deci et al. (1999) conclue à un impact négatif sur la MI d'autant plus fort que les retours sont négatifs.

En résumé, un large accord émerge de la littérature sur la question de l'articulation entre MI et ME. La théorie de l'évaluation cognitive semble vérifiée. Notons cependant qu'elle prédit que l'introduction de ME peut réduire les MI ; rien n'est dit de l'effet global sur la motivation. Il est possible que la réduction de la MI soit compensée par l'ampleur de la ME.

Pour arriver à une réduction de la motivation totale, il faut que l'effet à la baisse de la MI soit relativement supérieur à l'effet positif de la ME.

Enfin, notons que contrairement à ce que l'on pouvait penser de prime abord, la TAD n'est pas exempte de considérations sociales. Le besoin de relation à autrui, par exemple, suppose que l'individu doit être lié à une communauté. La TAD est une psychologie sociale. Cependant, il subsiste une différence importante entre les besoins psychologiques fondamentaux de la TAD et le concept d'éthique : l'absence de jugement sur le bien et le mal. Si ce qui compte est uniquement la satisfaction des besoins psychologiques fondamentaux, rien n'est pourtant dit sur la façon de les satisfaire. Par exemple, appartenir à une communauté de médecins criminels peut suffire à satisfaire le besoin psychologique fondamental de relation à autrui.

Maintenant que nous avons parcouru les éléments qui nous semblent essentiels à la compréhension de la TAD, nous sommes en mesure d'évaluer la façon dont le *mainstream* l'a incorporé en économie. La TAD porte un potentiel subversif important. Elle parle de besoin – en opposition aux préférences. Elle parle d'eudémonisme comme conception anti-utilitariste de l'individu – en opposition à l'utilitarisme de la TS. Elle dit que les individus peuvent travailler et faire des efforts pour rien (MI) – en opposition à l'hypothèse d'intéressement et de désutilité de l'effort. Elle dit que payer peut avoir un impact négatif sur la motivation – en opposition avec la loi d'une augmentation de l'offre de travail pour toute variation positive du prix du travail. Elle dit que les individus sont sujets à l'interprétation de leur environnement social et des politiques d'incitation qui en émanent – en opposition au calcul généralisé et systématique de l'homme économique. Que reste-t-il de tout ça une fois la TAD retraduite dans le langage *mainstream* ?

2. Au-delà de l'impérialisme traditionnel de la science économique : le travail de passeur de Bruno Frey

Les travaux de la TAD ne sont pas neufs, bien au contraire. Mais, ce n'est qu'au tout début des années 1990 que les économistes vont s'emparer de la problématique. L'économiste qui aura le rôle de « passeur » de la psychologie à l'économie est Bruno Frey. Son rôle étant central dans l'analyse de ce que l'on appelle aujourd'hui le *crowding out effect* (COE), il nous semble évident de nous attarder sur son travail. C'est en effet lui le premier à révéler

l'ensemble des problèmes et des perspectives relatifs au COE. Nous verrons que si la TAD est vaste et puissante, Frey n'en importe que la théorie de l'évaluation cognitive à laquelle il rattache d'autres types de littératures en psychologie. Son intérêt majeur porte sur l'effet d'éviction potentiel des motivations intrinsèques par les motivations extrinsèques. Cette hypothèse étant en contradiction totale avec la théorie économique *mainstream*, pour qui une hausse du prix relatif ne peut avoir pour effet que la hausse de l'offre, toute une partie du travail de l'auteur sera de faire accepter aux économistes son hypothèse novatrice.

Notre objectif est d'évaluer la façon dont Frey a introduit la Théorie de l'évaluation cognitive en économie et les conséquences que cela a sur le *mainstream*. Dans un texte paru en 2002 (Frey et M. Benz, 2002) Frey précise son positionnement par rapport à la Théorie Standard Etendue (TSE). Selon lui, ces dernières années ont vu une évolution favorable du rapport entre économie et psychologie. Alors que la TSE avait l'habitude d'intégrer tous les aspects de la vie humaine dans le cadre de la théorie du choix rationnel – à la suite de prestigieux auteurs comme Garry Becker –, désormais les économistes seraient plus attentifs aux limites de leur cadre analytique et adapteraient leurs théories aux résultats de la psychologie – dans une conception moins hostile des relations entre sciences sociales. La première méthodologie est qualifiée d'impérialisme et la seconde d'inspiration.

Il ne fait aucun doute que les travaux de Bruno Frey relèvent de l'inspiration. En s'instruisant des résultats de la psychologie de la motivation, l'auteur amende largement le cadre standard. Ses travaux constituent une percée importante dans la théorie économique : la théorie de l'évaluation cognitive va devenir la *Motivation Crowding out Theory* (2.1.). En opposition à la tradition *mainstream*, la MCT va peu à peu acquérir sa légitimité grâce à l'accumulation d'éléments de preuves empiriques (2.2.). De cette double construction (théorique et empirique) ressortent de nouvelles recommandations pour la politique de ressources humaines (2.3.).

2.1. De l'hypothèse de crowding out effect à la Motivation crowding theory

Le premier article important écrit par Bruno Frey sur la question des motivations date de 1992 (Frey, 1992). D'après l'auteur, la TSE s'attarde presque exclusivement sur l'opportunisme des individus pour expliquer leurs changements de comportement. Si un

individu change de décision, c'est que les prix relatifs (implicites[51] ou explicites) ont été modifiés. Cette modification ne peut être due qu'à une cause externe à l'individu, les préférences étant stables. À l'inverse, les études en psychologie suggèrent qu'il peut exister des causes non externes du comportement. La théorie de la MI de Deci et Ryan est un exemple de ce type de conceptions.

Cependant, l'existence de MI n'est pas un problème pour la TSE : il suffit d'associer les MI à des préférences. Là où Frey rompt avec la conception traditionnelle de l'homme économique c'est lorsqu'il énonce que la MI peut se modifier. Car si la MI peut se modifier, c'est que les préférences changent. Les MI peuvent alors être évincées par les ME. Autrement dit, l'effet prix relatif peut modifier les préférences de telle sorte que, plutôt que d'observer la traditionnelle augmentation de l'offre suite à l'augmentation des prix relatifs, celle-ci diminue. Le comportement change parce que les préférences changent. L'individu de la théorie économique ne serait alors plus, tel le chien Pavlovien, l'esclave de ses préférences programmées. Il est désormais nécessaire d'étudier les coûts cachés des récompenses (Lepper et Greene, 1978).

Dans un texte ne faisant pas explicitement référence au COE, Frey (1993) donne une épaisseur considérable à sa pensée par le recours à la notion de travail moral (*work moral*). L'angle de la critique est l'hypothèse de tire-au-flanc (*shirking*) mobilisée dans la TSE pour justifier le contrôle externe de la main d'œuvre. Selon lui, l'effort produit par un individu est non seulement fonction de la compensation monétaire (ce sur quoi la TSE insiste) mais aussi d'un travail moral[52]. Ce travail moral est une donnée importante dans différentes situations : contrôle externe difficile à mettre en place ; difficulté de formuler une régulation claire et de contrôler son application ; difficulté d'évaluer la contribution individuelle d'une action collective ; difficulté de mesurer la qualité.

Ce raisonnement, même s'il n'y fait pas directement référence, rappelle les situations de contrats incomplets. On peut alors dire que Frey anticipe ici l'argument de Kreps, selon lequel il est tout à fait rationnel de s'appuyer sur le MI lorsque les contrats sont incomplets. Dans ces cas-là, il faut relativiser l'hypothèse de désutilité du travail qui peut, dans certains cas, être entièrement fausse (Kreps, 1997, p.361). En effet, d'après Kreps, dans certains métiers des ambiguïtés fortes sur ce qu'il convient de faire sont possibles pour ne pas dire fréquentes.

[51] Une variation de prix implicite est liée à l'évolution d'éléments non monétaires comme, par exemple, les coûts d'opportunité.

[52] S'il ne parle pas de MI, l'auteur décrit le travail moral comme un *intrinsic factor* (Frey, 1993, p. 1224).

Comme dans la médecine, le travail est difficile à mesurer et à quantifier objectivement ce qui peut engendrer deux types de problèmes. D'une part, les employés peuvent être incités à tricher dans leurs outils de mesure, et, d'autre part, la rationalité limitée des agents peut les obliger à évaluer leurs performances *ex post*. Etant donné le vague de la mesure du travail (qualité, intensité, etc.), on ne peut pas *ex ante* s'accorder sur une rémunération *ex post*. En conséquence, des réponses « pragmatiques » aux contingences sont certainement préférables à la maximisation avant l'accomplissement de la tâche. Ce constat repose sur les limites cognitives humaines : il est en effet impossible de tout calculer et de tout prévoir. Ainsi, dans ce type de situation, il est préférable de laisser faire les MI, l'introduction d'incitations étant incapable de résoudre le problème de coordination.

Selon Frey (1993), dans ces situations, à défaut de calcul il existe un contrat psychologique ou un contrat implicite entre l'employé et l'employeur. Ce dernier accorde confiance et respect au premier en reconnaissant son travail moral. L'agent ne tire pas au flanc tant que le principal ne lui attribue pas un niveau d'effort moral plus bas que celui qu'il pense avoir effectivement. Dans cette optique, la surveillance, le contrôle ou les incitations que pourrait mettre en place le principal signent l'aveu de défiance du principal. Ce faisant, le principal attribue une morale particulièrement dégradante à l'agent et le contrat implicite est rompu.

Dans ses différents textes, afin de fonder théoriquement ces conjectures, Bruno Frey mobilise une large palette de théories en psychologie sociale :

(i) Autonomie : lorsque les individus se sentent contrôlés, il y a un risque d'éviction des MI. La source théorique est clairement la théorie de l'évaluation cognitive de Deci et Ryan (2002). L'éviction est une façon pour les individus de reprendre le contrôle (Frey, 1992, 1997a).

(ii) Surjustification : dans le même esprit que Deci et Ryan, Frey mobilise la notion d'*Overjustification* (Frey, 1993) issue des théories de l'attribution (Lepper et al., 1973). Les récompenses externes sont perçues par les individus comme une justification supplémentaire à leur comportement initialement intrinsèquement motivé. Or, cette sur-justification modifie le sens de l'action et la rend moins attrayante.

(iii) Auto-évaluation : la ME signale aux yeux des individus que le régulateur ne reconnait pas leur MI. Dans d'autres textes, Frey parle d'estime de soi (*self-*

esteem), la personne affectée ressent que sa compétence n'est pas appréciée à sa juste valeur (Frey, 1992, 1997b).

(iv) Norme de réciprocité : le raisonnement développé suit exactement celui des différents travaux d'Akerlof (1982). La motivation extrinsèque peut être interprétée comme le signe de la rupture d'une relation de don / contre don entre l'employeur et les employés. L'incitation supplémentaire ne permet pas de maintenir le niveau de sur-travail issu de la norme de réciprocité.

(v) Colère altruiste : Lorsqu'une motivation externe est appliquée, l'individu perd toute possibilité de signaler sa MI. En réponse, il évince sa MI en exhibant cette fois-ci une colère altruiste (*altruism anger*, Frey, 1997b).

Dans un texte paru dans l'*American Economic Review* (Frey et Oberholzer-Gee, 1997), Frey précise sa position par rapport à la TSE. Il s'agit pour lui de rouvrir le débat opposant Richard Titmuss aux deux prix Nobel Arrow (1972) et Solow (1971) : les MI et les ME ne sont pas nécessairement complémentaires, elles peuvent être substituables. Du point de vue de la TSE, la MI n'est pas une découverte : les individus ont des goûts qu'ils expriment lors de leurs interactions sur les marchés. La nouveauté c'est l'hypothèse de substituabilité des registres de motivation. Or, ce débat, ouvert par Titmuss en 1970, a été clos dès 1972.

Dans son livre *The gift relationship. From human blood to social policy*, Titmuss (1970) s'interroge sur le mode de transfert de sang optimal. Dans un contexte de pénurie de sang, l'auteur explique que l'introduction d'un prix pour un bien qui autrefois relevait du don aurait pour conséquence de réduire significativement la quantité et la qualité du sang échangé. La MI à donner son sang serait ainsi réduite par l'existence de « donneurs » extrinsèquement motivés, c'est-à-dire payés[53]. Pour ces raisons Titmuss s'oppose radicalement à l'ouverture d'un « marché du sang ». Les deux prix Nobel d'économie sont en revanche plus mesurés et ne voient pas pour quelles raisons l'existence de « donneurs » extrinsèquement motivés aurait un effet sur les donneurs intrinsèquement motivés[54]. Depuis lors jusqu'à ces dernières années, la thèse de complémentarité des registres de motivation est restée dominante.

Comment explique-t-on alors que le débat se soit rouvert avec plus de succès aujourd'hui ? Deux hypothèses semblent probables. D'une part, les travaux en psychologie sociale sur l'effet d'éviction n'en étaient qu'à leur début. Deci écrit le premier article majeur

[53] L'auteur n'utilise pas le cadre conceptuel de la TAD. Nous l'utilisons ici afin de faciliter la lecture.

[54] On pourra se reporter à Steiner (2001) pour une discussion de ce problème.

sur la question en 1971. D'autre part, les travaux de Bruno Frey connaissent un meilleur sort parce que les économistes connaissent aujourd'hui les limites empiriques de la théorie du choix rationnel. Ainsi, l'hypothèse de *crowding out effect* est solidement étayée par ce qu'il manquait à Richard Titmuss : une accumulation de preuves empiriques.

2.2. Acquérir une légitimité à partir de l'accumulation de preuves empiriques

Après avoir mis au point les grandes lignes de sa théorie au début des années 1990, Bruno Frey aura ensuite pour projet principal de la diffuser, de l'affiner et de la prouver. Le texte publié en 2001, en collaboration avec Reto Jegen dans le *Journal of economic surveys* (Frey et Jegen, 2001), est de ce triple point de vue une consécration. Du point de vue de la diffusion, l'hypothèse de *crowding out effect* est désormais si bien connue des économistes qu'elle mérite une place dans cette prestigieuse revue – et dans de nombreuses autres. Du point de vue théorique, l'article de 2001 semble opérer une montée en généralité déjà existante auparavant mais pour la première fois aussi clairement énoncée : le *crowding out effect* est un cas particulier d'une théorie plus générale, la *Motivation Crowding Theory* (MCT).

La théorie *mainstream* s'intéresse à l'effet prix relatif et considère la régulation externe comme ayant uniquement un effet disciplinant (*disciplining effect*)[55]. C'est ce que prédit la théorie de *l'agence* la plus standard (TS et TSE). L'individu réagit mécaniquement à toute variation des prix relatifs. La MCT introduit dans le cadre économique standard la notion de MI comme elle a été travaillée par les psychologues. En plus de l'effet d'éviction, la régulation externe peut avoir pour effet d'entrainer la MI. Au total deux types de relations entre motivations sont envisageables. (a) La régulation externe produit un entrainement de la MI : l'effet disciplinant et le *crowding effect* s'additionnent. C'est le *crowding in effect*. (b) La régulation externe produit une réduction de la MI plus puissante que l'effet disciplinant. La motivation totale baisse au lieu de croitre : c'est le *crowding out effect* (Encadré 3.2.).

[55] Cf. chapitre 1 pour un commentaire de cette notion.

Encadré 3.2. La Motivation crowding out theory

Le modèle de Frey (1997)

Le but de Frey est d'intégrer dans un modèle simple – cherchant uniquement à donner une intuition à l'idée de *crowding out effect* – deux effets contradictoires : l'effet prix (*price effect*) et l'effet d'éviction (*crowding out effect*). Il s'agit, dans le cadre d'une relation d'agence, d'étudier l'impact d'une régulation externe (E) sur le comportement de celui qui la reçoit. Le but du principal est, grâce à la régulation, d'améliorer la performance (P) de l'agent. L'agent décide de son niveau d'effort de performance après avoir étudié la balance coût (C) / bénéfice (B) avec :

$$\frac{\partial B}{\partial P} = B_P > 0 \; et \; \frac{\partial C}{\partial P} = C_P > 0$$

Frey suppose également des bénéfices marginaux décroissants (B_{PP}<0) et des coûts marginaux croissants (C_{PP}>0). Bien entendu les coûts et les bénéfices sont corrélés à la régulation externe. On a :

$$B = B(P, E); B_P > 0, B_{PP} < 0$$

$$C(P, E); C_P > 0, C_{PP} > 0$$

L'agent maximise son bénéfice net (B-C) en choisissant son niveau de performance optimal P^*, d'où :

$$B_P = C_P$$

L'agent va alors rationnellement augmenter son niveau de performance tant que le bénéfice marginal est égal au coût marginal. On trouve en dérivant la condition d'optimalité l'effet de la variation de l'intervention (E).

$$B_{PE} + B_{PP}\frac{dP^*}{dE} = C_{PE} + C_{PP}\frac{dP^*}{dE}$$

En isolant l'effet de la régulation :

$$\frac{dP^*}{dE} = \frac{B_{PE} - C_{PE}}{C_{PP} - B_{pp}} \gtreqless 0$$

Dès lors, Frey distingue trois cas selon le signe de la fraction.

Suivant la théorie principal-agent habituelle, la régulation externe améliore la performance en imposant une hausse des coûts marginaux (grâce à la surveillance des comportements tire-au-flanc) et/ou une baisse du bénéfice marginal : C_{PE}<0. C'est l'effet prix relatif (*relative price effect*) ou effet disciplinant (*diciplinning effect*). L'effet d'éviction est nul dans la mesure où le contrôle n'affecte pas le niveau des bénéfices (B_{PE}=0). Il n'y a pas de motivation intrinsèque et l'effet de l'intervention externe est toujours positif :

$$\frac{dP^*}{dE} > 0$$

Si l'intervention externe accroit la motivation intrinsèque, alors la performance s'améliore également par l'intermédiaire de la hausse du bénéfice marginal (B_{PE}>0). Dans ce cas, l'effet disciplinant et l'effet motivation vont dans le même sens.

A l'inverse, si la régulation externe réduit la motivation intrinsèque alors le bénéfice marginal décroit (B_{PE}) et si l'on suppose que l'effet disciplinant est nul (C_{PE}), on aboutit, de façon certaine, à la situation de *crowding out effect* :

$$\frac{dP^*}{dE} < 0$$

En général l'effet disciplinant et l'effet d'éviction jouent ensemble. L'effet sur la performance dépend alors de leur force relative.

De son côté, le principal cherche à maximiser son utilité (U) en anticipant la réaction de l'agent au contrôle externe. Le profit est croissant de l'output (X), lui-même dépendant de la performance de l'agent, avec des rendements décroissants. D'où :

$$X = X(P); X_P > 0, X_{PP} < 0$$

La régulation implique un coût (K) de type :

$$K = K(E);\; K_E > 0 K_{EE} > 0$$

Le principal cherche également à maximiser son utilité (U=X-K). On obtient les conditions du premier ordre :

$$X_P \frac{dP^*}{dE} = K_E$$

Cette dernière équation permet de donner le niveau optimal d'intervention (E^*). Celui-ci dépend de l'effet de l'intervention sur le niveau de performance optimal. Il est alors à nouveau possible de distinguer trois cas :

(i) Lorsque l'intervention n'affecte pas le niveau de performance ($\frac{dP^*}{dE} = 0$), il est rationnel pour le principal de ne pas intervenir. La régulation ne modifie pas la performance.

(ii) Lorsque l'effet prix domine l'effet motivation, et donc que la régulation a un effet positif ($\frac{dP^*}{dE} > 0$), le principal a tout intérêt à mettre en place une régulation externe. Les conclusions sont les mêmes lorsqu'il y a un *crowding in effect*.

(iii) Lorsque l'effet motivation domine l'effet prix ($\frac{dP^*}{dE} < 0$), la régulation est à proscrire car elle réduit la performance.

La prise en compte de la motivation de l'agent incite désormais le principal à étudier à nouveau la stratégie de fixation de la régulation externe.

Cette clarification théorique étant posée, il reste à analyser la question de la preuve du COE. Dès 1992, l'auteur proposait un grand nombre d'applications possibles de sa théorie : politique environnementale, prévention du crime, politique sociale, politique de main d'œuvre, politique fiscale, etc. (Frey, 1992).

Un exemple de démonstration empirique de l'existence de COE peut être décrit, avant de mener un repérage plus systématique de cette littérature. Le texte « A constitution for Knaves crowds out civic virtues » (Frey, 1997a) s'intéresse à la relation entre le type d'institutions politiques et l'évasion fiscale. L'idée de départ de l'auteur est de faire la critique de la théorie du *Public Choice* et, plus précisément, de la position de Brennan et Buchanan (1983) sur la relation entre l'opportunisme des acteurs (citoyens, personnages politiques et bureaucrates) et le type d'institutions efficaces dans une « économie de marché libre ». Selon Frey, cette façon de poser le problème des institutions politiques n'est pas la bonne. Les vertus civiques existent. La question n'est pas tant de comprendre pourquoi il y a des criminels et des délinquants (de l'opportunisme) mais, pourquoi y-a-t-il tant d'« innocents » ? La thèse de l'auteur est dans la lignée de ses travaux sur le COE : penser une constitution à partir de l'hypothèse que les individus sont opportunistes génère un effet d'éviction… rendant les individus opportunistes.

D'après l'auteur, deux conditions sont importantes pour expliquer le COE dans le cas des institutions politiques : (i) Les possibilités individuelles de participation institutionnelle doivent être importantes. Si l'individu ne prend pas part à la décision, il peut se sentir contrôlé (ia). Si l'individu ne peut pas prendre une décision parce qu'il est jugé incompétent, c'est son estime de soi (*self-esteem*) qui est remise en cause (ib). (ii) Plus proches sont les relations

entre l'administrant et l'administré plus puissant sera le COE. Ainsi, de manière plus générale, pour éviter le COE, il faut que la constitution face, au moins en partie, confiance aux citoyens (Frey, 1997a, p. 1048).

Eloigner les institutions politiques des citoyens serait une marque de défiance qui peut avoir pour effet d'évincer les vertus morales. Afin de démontrer ces assertions théoriques, Frey mène une analyse économétrique sur la question de l'évasion fiscale et de la taxe morale (*tax moral*) – composante essentielle de la vertu morale. Dans cette optique, il est impossible d'expliquer les comportements de paiement d'impôts sans faire référence à la morale. La hausse de la probabilité de détection et de punition de l'évasion fiscale devrait théoriquement mener à une réduction de ce type de comportement (hausse du coût moyen de la triche). Pourtant, empiriquement cela ne fonctionne pas. D'après les résultats obtenus sur la démocratie Suisse, c'est à cause de la taxe morale : l'effet de hausse de l'anticipation de la punition est contrebalancé par la baisse de la taxe morale (qui encourage à payer ses impôts).

Plus largement, la littérature empirique sur le COE s'étant très fortement développée depuis le début des années 2000, nous synthétisons dans le tableau 3.1. un échantillon de différentes études portant sur les champs les plus étudiés – dont l'économie de la santé.

Tableau 3.1. L'étude empirique du COE dans différents champs de la littérature : un échantillon

Domaine de recherche				
Marché du travail	Comportement pro-sociaux	Bien commun (environnement)	Institutions politiques	Économie de la santé
Gneezy et Rustichini (2000a) ; Fehr et Gächter (2000) ; Bewley (1995) ; Frey et Götte (1999) ; Austin et Gittell (1999) ; etc.	Upton (1973) ; Gneezy et Rustitchini (2000b) ; Ariely et al. (2009) ; Fiorillo (2011) ; Mellström et Johannesson (2008) ; etc.	Frohlich et Oppenheimer (1998) ; Frey et Oberholzer-Gee (1997) ; Kunreuther et Easterling (1990) ; Cardenas et al (2000) ; etc.	Bohnet et al. (2001) ; Frey (1997a) ; Kelman (1992) ; Degli Antoni (2009) ; etc.	Holmas et al. (2010) ; Forest (2008) ; Videau et al. (2010) ; Videau et al. (2011) ; Sicsic et al. (2012) ; Lievaut (2010) ; etc.

Source : Frey et Jegen (2001), complété par nous

2.3. Une théorie des ressources humaines et de la politique publique

La MCT est désormais théoriquement limpide et empiriquement validée. Il ne reste plus qu'à faire un inventaire des préconisations issues de ses développements. Qu'a-t-on appris sur la gestion des ressources humaines et sur les politiques publiques grâce aux travaux de Bruno

Frey ? Sous quelles conditions peut-on s'attendre à observer un COE ? Voici une synthèse possible des résultats théoriques auxquels l'auteur aboutit :

- Incomplétude contractuelle et MI (Frey, 1993, Frey, 1997c) :

Toutes les tâches qui ne peuvent faire l'objet d'un contrat complet sont dans ce cas. Il est moins coûteux de s'appuyer sur les MI que de rédiger un contrat complet.

- L'uniformité de la régulation (Frey, 1993, Frey, 1997b, Frey, 1997c) :

Plus la régulation est uniforme, plus les effets contreproductifs sont probables. En effet, il serait théoriquement optimal de distinguer les tire-au-flanc des autres pour bien choisir qui surveiller. Comme cela n'est pas possible, soit il n'y a pas de régulation du tout (et les non tire-au-flanc se sentent lésés) soit il y a une même régulation pour tous (et les non tire-au-flanc sont sujet au COE). La difficulté est plus grande encore lorsque l'on s'intéresse au secteur public dans la mesure où la rationalité administrative produit des régulations uniformes.

- Le besoin et l'absence d'alternative (Frey et al., 1997) :

Lorsqu'un projet du type *Not In My Back Yard* (NIMY) est nécessaire, il y a moins de chance que le COE se produise. Les projets NIMBY sont des projets dits d'intérêt collectif mais dont les conséquences peuvent être particulièrement néfastes dans la zone géographique de leur implantation. C'est par exemple le cas des sites d'enfouissement de déchets nucléaires, des usines chimiques, des prisons, des parcs éoliens, des autoroutes, etc.

- Réglementation et incitation (Frey et al., 1997, Frey, 1997c) :

La règlementation est toujours plus néfaste pour les MI que les incitations financières. Dans ce derniers cas, les individus ont au moins la liberté de refuser alors que dans le premier ils sont contraints.

- Degré de démocratie (Frey et al. 1997, Frey, 1997a, Frey, 1997b, Frey, 1997c) :

Les vertus morales ont le plus de chances d'être sauvegardées quand la démocratie participative est fortement développée (référendums classiques, référendums d'initiative populaire). Par ailleurs, dans le cadre des projets NIMBY, lorsque le processus de décision du lieu d'installation a été jugé juste (légitime), le COE est moins probable.

- Proximité régulateur-régulé (Frey, 1997a, Frey, 1997b, Frey, 1997c) :

Plus la relation est personnelle entre le régulateur et le régulé – c'est-à-dire plus elle implique reconnaissance, confiance et loyauté –, plus le COE est probable.

- Contingence de la régulation (Frey, 1997b, Frey, 1997c) :

Plus la régulation externe est contingente à une performance spécifique – comme le paiement à la performance –, au lieu d'être définie à un niveau général, plus le COE est probable. La monétisation de la régulation a tendance à souligner la contingence de la régulation.

Dans un article plus récent, Frey pousse sa théorie de la motivation dans un terrain *a priori* en dehors de son champ : la rémunération des dirigeants d'entreprise (Frey et Osterloh, 2005). Traditionnellement, l'outil le plus adéquat pour optimiser la valeur pour l'entreprise est le paiement à la performance. Mais, nous disent les auteurs, ce mode de rémunération est à l'origine des scandales d'entreprise (par exemple, WorldCom, Xerox et Enron). Cela s'explique par l'inattention accordée aux MI des individus. Malgré le suivi des préconisations de la théorie de l'agence, l'introduction de mécanismes de paiement à la performance (P4P)n'a pas ainsi eu d'effet substantiel sur la performance des entreprises. Aujourd'hui, même Jensen reconnait les difficultés liées à ce mode de rémunération qui constituera une sorte de drogue. Mais selon ce dernier, il suffirait d'améliorer le *design* de ses mécanismes pour en augmenter le rendement. La thèse de Frey et Osterloh est inverse :

> *« [We argue] that high-powered incentive compensation, even if it could be optimally designed, does not solve the problems in the corporate sector identified but aggravates it » (Frey et Osterloh, 2005, p. 97).*

Le paiement à la performance produit une série d'effets contreproductifs importants dans le cadre de l'entreprise : il donne le signal au directeur que faire son devoir sans sur-rémunération est socialement inapproprié ; il rapproche la firme des conditions de marché concurrentiel dans lesquelles les comportements pro-sociaux sont inadéquats ; il change le contrat relationnel en contrat transactionnel (avec moins d'éléments socio-émotionnels) ; il fausse la neutralité des superviseurs réduisant ainsi la justice perçue des procédures etc. En conclusion, et cela en opposition avec la TSE, il faut prendre en compte les MI et ne pas s'effrayer de ses résultats : « Yes, Managers should be paied like burecrats » (Frey et Osterloh, 2005). Autrement dit, même pour les dirigeants – que dire alors du médecin –, il peut être avantageux de préférer un salaire fixe à au paiement à la performance !

3 Comment lire les motivations intrinsèques avec la théorie économique de l'information ?

Les travaux de Bruno Frey et ses collègues behavioristes sur le COE ont depuis longtemps fait l'objet d'une traduction dans le cadre de la TSE2. Dans cette section, notre travail cherche à montrer que la notion de MI n'est pas compatible avec la conception standard de la rationalité médicale mobilisée par la théorie de l'agence. L'hypothèse de MI, parce qu'elle relativise l'importance des comportements purement instrumentaux, doit nécessairement déboucher sur la prise en compte d'une certaine forme de désintéressement. Ainsi, à rebours de certaines conceptions de la MI qui cherchent à la ré-encastrer dans la théorie du choix rationnel, nous cherchons au contraire à montrer que la compréhension des motivations médicales exige de prendre ses distances avec ce cadre d'analyse standard.

Le raisonnement est mené en deux temps. Dans un premier temps, nous montrons que le COE, quand il est ramené à une relation d'agence, est le résultat d'un simple problème d'information modifiant la stratégie des individus (3.1.). Dans un second temps, nous verrons que l'hypothèse de MI, en posant la question du sens de l'activité médicale, est difficilement incorporable au cadre standard d'analyse économique basé sur la désutilité de l'effort et l'intéressement (3.2.). Surgissent alors à nouveau des paradoxes de théorie économique et des problèmes pour la politique publique.

3.1. MI, COE et TSE2 : un problème d'information

Et si les questions posées par la TAD n'étaient rien d'autre que des problèmes d'information ? C'est la position de la TSE2 pour qui il est hors de question d'accepter que le COE produise un changement de préférence ou une remise en cause de la théorie du choix rationnel. La problématique des registres de motivation est alors retraduite dans le cadre de la relation d'agence. On peut classifier cette littérature en cinq grand cas de figure.

Dans le modèle du principal informé, l'asymétrie d'information est l'avantage du principal et le COE provient d'une mauvaise interprétation de l'agent sur l'information qu'a le principal sur (3.1.1.). Lorsque l'agent dispose d'un avantage informationnel, il est possible que l'existence de MI implique un problème de sélection adverse pour le principal qui cherche à recruter les agents les plus motivés (3.1.2.). L'existence de MI peut aussi être problématique lorsque l'on pose l'hypothèse qu'une partie des agents sont conformistes. Le

principal doit alors chercher à inciter les conformistes à se comporter comme des intrinsèquement motivés (3.1.3.). Enfin, l'effet d'éviction peut provenir d'un phénomène de la recherche d'estime sociale (3.1.4.) ou de réputation (3.1.5). Le tableau 3.2. synthétise l'ensemble des modèles.

Tableau 3.2. Le crowding out effect dans la TSE2

Section	Nom du modèle	Problème posé	Cause probable du COE	Référence typique
3.1.1.1.	Modèle du principal informé	Le principal est informé et l'agent interprète l'incitation comme un signal de la difficulté de la tâche	L'incitation peut révéler la difficulté de la tâche et le défaut de confiance du principal	Bénabou et Tirole (2003)
3.1.1.2.	Modèle de recrutement des dévoués	Les agents sont hétérogènes dans leur motivation : les dévoués produisent de la bonne qualité et les intéressés la moindre qualité	L'incitation a pour effet d'attirer les plus intéressés ce qui réduit la proportion totale de dévoués	Taylor (2007)
3.1.1.3.	Modèle des conformistes	En plus des intéressés et des intrinsèquement motivés, le régulateur doit gérer le comportement des conformistes	Le niveau de l'incitation est un signal qui peut inciter les conformistes à agir comme des intéressés	Sliwka (2007)
3.1.1.4.	Modèle à estime de soi	L'agent et le principal cherchent à se séduire mais peinent à se faire confiance	L'incitation peut signaler le manque d'estime du principal envers l'agent	Ellingson et Johannesson (2008)
3.1.1.5.	Modèle réputationnel	L'acception de l'incitation à un impact sur la réputation	L'acception de l'incitation peut signaler aux autres (en dehors de la relation d'agence) une déviance morale	Bénabou et Tirole (2008)

3.1.1. Les ME signalent la difficulté de l'activité médicale

Les économistes considèrent habituellement que le médecin dispose d'une position avantageuse sur le marché du soin puisqu'il est détenteur d'un stock d'informations plus important que le principal (assureur et/ou patient). Une telle conception peut être remise en cause en supposant que c'est le principal qui dispose d'une information privée sur le médecin. Celle-ci porte sur un élément déterminant le paiement de l'agent, c'est à dire soit sur ses caractéristiques (compétence, expérience, formation...) soit sur la tâche elle-même (difficulté, intérêt...). L'agent ignore cette information mais en reçoit tout de même un signal imparfait qu'il pourra utiliser par la suite. Ces hypothèses sont au cœur du modèle de principal informé (Bénabou et Tirole, 2003) dans lequel cela pose un problème de confiance qui provoque le COE.

Le principal, ayant intérêt à ce que l'agent essaie et réussisse la tâche proposée, peut chercher à l'inciter financièrement. En plus des effets habituels de l'incitation sur les contractants, Bénabou et Tirole (2003) mettent en évidence l'existence d'un *confidence-management motive*. Il s'agit du pouvoir de gestion que détient le principal sur la confiance en soi de l'agent. Notons qu'il s'agit d'un type de confiance tout à fait original, bien différent du sens commun. C'est, ici, une probabilité subjective de réussite de l'agent (évaluée par lui) qui se trouve dépendante de la stratégie de manipulation d'information du principal. Il y a alors un lien entre la confiance en soi de l'agent et la confiance que le principal accorde à l'agent, cette dernière étant entendue comme la probabilité subjective de réussite de l'agent calculée par le principal. La confiance du principal sur la probabilité de réussite de l'agent est ainsi directement liée, à travers l'incitation, à la confiance en soi de l'agent.

L'agent sait que le principal est informé et que celui-ci peut utiliser cette information à son désavantage. L'incitation est un signal de l'information cachée qu'il convient d'interpréter. L'agent ajuste alors ses croyances en fonction de l'interprétation qu'il se fait de la régulation. L'effet d'éviction émerge lorsque l'agent voit dans l'incitation du principal un signal négatif sur l'information cachée. Selon les auteurs, c'est le cas lorsque l'incitation est forte parce qu'elle est interprétée comme une mauvaise nouvelle sur la difficulté de la tâche. L'incitation décourage l'effort en réduisant la confiance en soi de l'agent : c'est le COE.

Dans le système de santé, pour qu'il y ait COE sur la base d'un problème de confiance, le régulateur doit détenir une information sur le médecin suffisamment importante pour que

celle-ci ait une influence sur sa rémunération. Nous proposons plusieurs illustrations du type d'informations susceptibles d'avoir ces caractéristiques :

- Capacité des médecins à atteindre les objectifs. Les médecins peuvent être incertains quant à la possibilité d'atteindre les objectifs. Par exemple, depuis 1996, le parlement vote l'Objectif national des dépenses d'assurance maladie qui n'est que rarement respecté. L'objectif étant peu crédible, l'incitation peut être interprétée par les médecins comme un manque de confiance de la tutelle sur leur capacité à réussir la tâche.

- Difficultés administratives liées à l'enregistrement des objectifs. N'étant pas suffisamment dotés en outils d'information (Bras, 2011), les médecins peuvent interpréter les incitations comme une preuve de la lourdeur administrative. Encore une fois, la mise en place d'incitation financière peut être perçue comme un signe de difficulté de la tâche.

Ces informations semblent répondre aux exigences imposées : elles sont en possession des pouvoirs publics mais pas nécessairement des médecins et elles influent sur la rémunération de ces derniers. Le médecin peut ainsi interpréter l'incitation comme une information sur la difficulté de la tâche (un déficit de confiance de la part de la tutelle). Et, en conséquence, il va, malgré un revenu potentiellement croissant, réduire son niveau de motivation et donc son effort. La MI à l'atteinte d'un objectif est évincée par la ME. Cependant ce n'est pas la seule explication au COE. L'incitation peut aussi poser un problème de sélection adverse.

3.1.2. Un problème de sélection adverse : recruter les médecins les plus motivés

Le modèle de principal informé est assez peu fréquent : le plus souvent c'est le médecin-agent qui est informé. Dans une première série de modèles, l'hypothèse de MI trouve une reformulation dans un problème de sélection adverse. Le principal – la tutelle assurant les soins (puissance publique ou entreprise privée) – cherche à recruter les agents les plus dévoués, les plus engagés – ceux motivés intrinsèquement. La MI renouvelle alors la question du recrutement : payer plus attirerait les moins dévoués (Heyes, 2007, Taylor, 2007).

Le modèle de Taylor (2007) s'intéresse par exemple au marché des infirmières – que l'on peut étendre au marché des professionnels de santé. L'auteur considère que les médecins se répartissent en deux catégories : (i) des médecins pour qui la médecine n'est qu'un simple travail et qui produisent un niveau de qualité minimal et (ii) des médecins pour qui la médecine est aussi une vocation et qui produisent un niveau de qualité non observable

supérieur au niveau minimal. Si l'on suppose un niveau de rémunération identique pour les deux types de médecins, la proportion de médecins dévoués décroit avec l'augmentation du salaire. En effet, la hausse de la rémunération attire plus de médecins intéressés alors que les médecins dévoués étaient déjà attirés – malgré la faiblesse de la rémunération. Par ailleurs, étant donné que les médecins sans vocation produisent une qualité inférieure, plus le salaire est élevé moins la qualité des soins est bonne. Il est alors rationnel pour la tutelle connaissant cet effet de sélection de proposer un niveau de salaire inférieur au niveau de *first best*.

Ce type de modèle illustre l'idée que les politiques de rémunération peuvent affecter le niveau de MI des personnes recrutées. Le fait de distribuer des hauts salaires réduit le niveau moyen de MI dans la main d'œuvre ainsi que le niveau de qualité moyen. C'est, selon l'auteur, une explication possible du COE.

Dans le modèle de Delfgaaw et Dur (2007), un monopoleur fait face au même problème de fixation de salaire que précédemment. Plus le salaire est haut, plus l'offre de travail sera abondante mais moins elle sera motivée – d'où un problème d'incitation à l'effort une fois les individus recrutés. L'employeur doit alors chercher à révéler l'information privée du salarié sur son niveau de motivation. D'autres modèles (Besley et Ghatak, 2005, Delfgaaw et Dur, 2008), s'intéressent plus spécifiquement à l'emploi dans le secteur public en posant l'hypothèse que certains agents peuvent être motivés par l'exercice dans le secteur public (*public service motivation*). Par exemple, Prendergast (2007) suppose qu'il existe une forte variation dans le degré dont les agents se soucient des actions qu'ils entreprennent. Le dévouement n'est pas alors uniforme comme peut l'illustrer la variabilité des agents de fonctions publiques face aux usagers.

Si dans ces modèles il est possible que la motivation soit affectée par l'intention du principal – par exemple, les religieux peuvent se démotiver s'ils se rendent compte qu'ils travaillent pour une église corrompue (Besley et Ghatak, 2005) –, la motivation des individus n'est pas affectée par les individus qu'ils côtoient. La section suivante s'intéresse à cette question en montrant que le dévouement des uns peut entrainer le dévouement des autres.

3.1.3. Les médecins conformistes

L'asymétrie d'information sur le type des individus peut poser un autre problème au principal si l'on considère que les préférences des uns peuvent dépendre de celle des autres (Ichino et Maggi, 2000, Mas et Moretti, 2009, Sliwka, 2007). Dans cette optique, la

motivation des uns est déterminée en partie par le comportement des autres. Ces modèles supposent comme les précédents qu'il existe une hétérogénéité de motivations entre les agents et que cette diversité d'inclinaison motivationnelle est cachée au principal et aux autres agents.

Dans le modèle de Sliwka (2007), parmi les trois types d'agent dont il sera question, deux types d'entre eux sont caractérisés par une motivation inébranlable : les médecins égoïstes (*strictly selfish*) – qui ne s'intéressent qu'à eux-mêmes – et les médecins justes (*fair*) – qui se soucient du bien-être de leur principal. Ce souci de l'autre est formalisé dans le langage de l'altruisme – leur utilité est croissante avec l'utilité du principal. L'innovation majeure de ce type de modèle par rapport à la section précédente est l'existence d'un troisième groupe : les médecins conformistes dont l'amour de la justice ou d'eux-mêmes dépend de leur entourage – c'est-à-dire du type majoritaire de médecin qui les entoure. Un médecin ne connait pas le type d'un autre médecin et la politique de recrutement de la tutelle est une information permettant d'évaluer le type majoritaire des médecins. Le problème de la puissance publique est alors de faire en sorte que les médecins conformistes agissent comme des justes. Même si ce n'est pas le cas, il faut que la politique d'incitation du principal soit telle que les conformistes croient être entourés de médecins justes.

Le point clef de cette formalisation est l'émergence d'un équilibre dans lequel les médecins conformistes sont convaincus que les plus nombreux sont les médecins de type « juste ». Supposons que la puissance publique sache quelle est la proportion de justes et des égoïstes. Dans le cas où elle dispose d'une majorité d'égoïstes, elle doit payer de fortes rémunérations variables destinées à inciter à l'effort – comme le P4P – et de faibles salaires – n'incitant pas à l'effort. À l'inverse, s'il existe un nombre majoritaire de médecins justes, il est optimal pour l'entreprise de payer de faibles rémunérations variables et un salaire relativement haut. En effet, la rémunération variable n'a pas lieu d'être puisqu'il est plus avantageux de se fier à la motivation des justes qui ne coûte rien.

Le fait d'augmenter le salaire fixe renforce alors le signal que la population totale est en majorité composée de médecins justes. La réponse des conformistes à ce signal crédible est de se comporter comme des justes et – paradoxalement – le profit (ou les économies) réalisé par la puissance publique est plus élevé que si elle payait de fortes primes à la performance. Au total, il y a un coût à ces faibles incitations (on n'est pas au *first best*) mais le bénéfice est plus grand dans la mesure où la puissance publique a convaincu les conformistes de faire un effort gratuit (lié à la MI des médecins justes et à la MI des médecins conformistes).

Ce modèle conçoit la confiance comme un signal d'une norme sociale (Sliwka, 2007). La tutelle peut choisir de mettre en place une rémunération à la performance. Mais, en ne le faisant pas elle donne un signal de confiance et les médecins travaillent plus pour moins – du fait des conformistes. Les médecins sont plus généreux dans leur effort que s'il y avait une autre norme (la norme des médecins égoïstes). Le COE médical est alors lié à une mauvaise politique d'incitation de la puissance publique. Si celle-ci préfère ne pas encourager les conformistes à être justes – en mettant en place une forte rémunération à la performance – si elle ne fait pas confiance, la défiance sape la motivation des conformistes. Plutôt que d'agir comme des justes, ils agissent comme des égoïstes. Le COE médical est alors un effet de composition des effectifs : les conformistes changent de norme d'effort.

3.1.4. Le médecin et l'estime de soi

Une autre possibilité pour introduire le concept de MI dans le cadre de la TSE2 est de le rapporter à l'estime de soi. Dans ce type de formalisation, les individus ont le désir d'être bien vu par les autres (Ellingson et Johannesson, 2007 et 2008) et cette « préférence sociale » peut expliquer le COE. Les individus veulent que les autres les estiment, les apprécient mais, ils n'ont qu'un signal imparfait de ce que les autres pensent réellement d'eux. L'utilité des deux participants en est alors affectée.

L'identité de l'audience (du public) est importante dans ce modèle. Les individus ne veulent pas être estimés par n'importe qui. Dans le modèle d'Ellingson et Johannesson (2008), par exemple, le bon public pour l'agent est le principal. Dans le cas de la santé qui nous importe, le médecin rechercherait l'approbation des pouvoirs publics – ou plus largement de la tutelle (privée ou publique). L'estime est alors définie comme une forme de respect. Plongée dans le monde des asymétries d'information, l'estime ne peut cependant être que ce que le médecin (agent) croit que la tutelle (principal) croit.

On suppose l'existence de deux types de médecins, certains talentueux et d'autres non talentueux. Cette information est non observable par le principal et l'utilité du médecin dépend, en plus des arguments habituels, de l'estime que lui accorde le principal – celle-ci dépendant notamment de la probabilité subjective que le principal pense que le médecin en question soit talentueux. L'estime des pouvoirs publique dépend du niveau d'effort choisi par le médecin. Cependant, pour les médecins moins talentueux, l'effort exercé a des effets positifs sur l'estime proportionnellement moins importants que pour les médecins talentueux.

Les moins talentueux sont ainsi incités à faire moins d'efforts parce qu'ils savent que ceux-ci auront un faible impact sur l'estime que leur portera la puissance publique.

L'intérêt du modèle est alors de proposer une explication originale du COE médical. Pour ce faire, il faut remarquer que l'estime portée aux autres est d'autant plus forte que l'on croit (principal et agent) que l'autre pense de même. Ainsi, lorsque la puissance publique offre une rémunération, cela a un impact sur les jeux sociaux qui vont suivre. De ce point de départ dépendra le niveau d'estime de soi que les uns vont attendre et accorder aux autres. Tout le problème de la puissance publique est alors de faire une proposition de rémunération crédible signalant son niveau d'estime. Si cette proposition est suffisamment pro-sociale, le médecin répondra par une action bénéficiant au principal : l'estime répond à l'estime. Sinon, une incitation signalant la défiance engendre un COE[56].

3.1.5. Un problème de réputation médicale

L'exercice de la médecine est une source de prestige qu'un comportement purement et ouvertement opportuniste remettrait en cause. Pour prendre en compte ces effets, il est possible de se référer au modèle à extraction de signal de Bénabou et Tirole (2006) reposant sur l'hypothèse d'*overjustification effect*. L'idée est que rémunérer une action en change le sens et sa justification, ce qui, en conséquence, rend l'action moins attrayante et réduit la motivation (voir également Brekke et al. 2003). Selon cette approche, en plus de la MI et de la ME, les individus sont mus par une motivation réputationnelle (MR). Celle-ci est associée aux aspirations qu'ont les individus à être reconnus socialement. C'est une mesure de leur sensibilité au prestige social. La motivation d'un individu est alors le résultat d'un calcul entre les différentes utilités qu'il retire compte tenu de ses préférences, des politiques d'incitations (ME) et de la double influence de la société (MR) et de son « être intérieur » (MI). Deux effets peuvent être soulignés.

Premièrement, on observe un effet direct des récompenses sur la motivation : les ME compromettent la valeur réputationnelle de la bonne action. Accepter une incitation pour une tâche quelconque revient à « sur-justifier » son comportement et ainsi à en modifier le sens. L'individu en question, afin de ne pas paraître avide, préfère refuser une rémunération supplémentaire (ME) plutôt que de risquer l'opprobre (MR). La ME change la signification de

[56] Ce type de modèle ressemble fortement à ce que peut proposer la théorie des jeux avec le jeu de la confiance (McCabe et al. 2003, Falk et Kosfeld, 2006) : or, si la confiance est rentable, elle n'est pas rationnelle.

l'acte ce qui le rend moins attrayant : l'utilité retirée de la ME ne suffit pas à compenser la désutilité engendrée par la mauvaise opinion que se font les autres de moi (MR).

Du point de vue de l'économie de la santé, cette construction se rapproche de l'idée selon laquelle la déontologie professionnelle est une condition essentielle des libertés dont bénéficie la profession médicale. Les médecins, dans leur quête de pouvoir face à l'État et aux caisses d'assurance maladie, mettent en avant le prestige social de la profession. Cette dimension éthique leur permet aujourd'hui de défendre en contrepartie un ensemble de libertés coûteuses pour la collectivité (libre prescription, paiement à l'acte, liberté d'installation, dépassements d'honoraires, etc.). Dès lors, il est important pour la corporation que le médecin ne soit pas assimilé à un simple marchand ce qui peut se traduire par le refus rationnel des incitations.

Une application du modèle à *overjustification effect* au comportement du médecin est donnée par Siciliani (2009). Selon lui, les dispositifs incitatifs sont interprétés par les médecins[57] comme un signal d'avidité qui est sanctionné par le stigmate social. Accepter une incitation revient, du point de vue du médecin, à admettre la croyance selon laquelle il faut impérativement le payer plus pour qu'il fasse son travail. La tâche, de prévention par exemple, n'est plus justifiée en elle-même ; elle est sur-justifiée par l'incitation. Dans ce modèle, la mise en place d'une politique d'incitation a pour effet de remodeler l'arbitrage des médecins. Ceux-ci doivent procéder à un arbitrage entre effort, gains financiers (provenant de l'incitation) et prestige social. Etant entendu que le fait pour un médecin d'accepter l'incitation est le signal pour les patients d'une forme d'avidité, l'auteur montre que l'arbitrage entre incitations financières et prestige peut engendrer un COE.

On peut par ailleurs mettre en avant l'effet des normes sociales sur la motivation. L'action individuelle dépend non seulement des préférences mais aussi des normes sociales, de ce qui est le comportement usuel (approprié). Dans ce contexte, une action qui est considérée normale prend force de règle-contraignante. En effet, si une telle action existe alors elle est suivie collectivement, et, le fait de refuser de suivre la norme engendre une stigmatisation. À l'inverse, si une action est jugée collectivement utile mais très peu suivie alors le fait de ne pas la suivre n'aura pas d'effets négatifs (ce n'est pas une norme). La décision stratégique d'accomplir ou pas une action repose sur la complémentarité ou la substituabilité de l'action individuelle avec les normes sociales.

[57] Pour être plus précis, le modèle de Siciliani suppose une hétérogénéité des médecins dans leur niveau de MI (ce dernier étant formalisé par une préférence altruiste plus ou moins prononcée pour la « bonne réputation »).

Concernant l'activité médicale il semble bien que l'altruisme, l'attention du médecin à l'égard de son patient, n'est pas un acte « héroïque » mais plutôt une attente sociale. Ainsi, le médecin se doit de refuser les incitations pour conserver un certain prestige social. Être un médecin attentif est un comportement normal qui n'est pas valorisé. À l'inverse, l'opportunisme est sanctionné par le discrédit. De ce point de vue, l'incitation n'est pas nécessaire et moins efficace que la publicité des médecins ouvertement opportunistes. Ici, la pression sociale est tellement forte (pairs, patients...) qu'il est impossible d'accepter une rétribution au risque d'être stigmatisé par la société.

3.2. La dissolution des MI dans le calcul rationnel

Comme pour le cas de l'éthique médicale, nous pensons que la tentative de faire rentrer le concept de MI et de COE dans le cadre de la TSE2 porte en elle de nombreuses limites. La première d'entre elle est que, paradoxalement, dans plusieurs modèles censés discuter du COE, on ne discute pas de COE. Il ne s'agit même pas de dire que la formalisation est incapable de capter le sens de ce phénomène complexe – la suite de la section s'attarde sur cette critique. Plus modestement, dans ces modèles, il n'est même pas question de COE. Par exemple, dans le modèle à sélection adverse de Taylor (2007), la réduction de la MI (vocation) totale suite à la hausse des salaires n'est en rien due au COE : l'éviction est un phénomène individuel. Or, dans ce modèle, l'éviction est liée à une modification de la composition des médecins travaillant pour la tutelle (plus d'égoïstes et moins de dévoués). Aucun médecin ne voit se modifier son niveau de motivation – ou même d'utilité – suite à la hausse des salaires. C'est un effet de composition global qui n'a rien à voir avec le COE.

Plus substantiellement, il est possible de faire deux critiques sur les modèles à MI et COE. La MI et le COE supposent, d'une part, que le médecin ne soit pas uniquement sujet à la désutilité de l'effort (3.2.1.) et, d'autre part, qu'il soit sujet au désintéressement (3.2.2). En cherchant à discuter du COE sans remettre en cause authentiquement ces deux hypothèses, la théorie économique des motivations peine à comprendre les motivations médicales.

3.2.1. MI et désutilité de l'effort

La première difficulté des modèles de la TSE2 dédiés à l'hypothèse de COE est liée à l'hypothèse de désutilité de l'effort. L'introduction de la MI devait pouvoir assumer l'hypothèse d'un médecin naturellement volontaire à l'accomplissement d'une tâche (MI) ce

que ne permettent pas les modèles précédents. De façon traditionnelle, l'agent est considéré comme un tire-au-flanc et le COE présenté n'a pas de ressorts psychologiques : c'est une logique d'intérêt (ME) qui pousse l'agent à refuser l'effort. Dans le modèle de Bénabou et Tirole (2003), par exemple, l'agent ne fait que réviser la probabilité qu'il a de réussir la tâche qu'on lui a assigné grâce à l'information contenue dans la ME. Où se trouve alors la MI à la tâche ? On reste dans l'interaction stratégique. Si avant l'introduction de la ME, sa motivation était plus élevée ce n'est pas parce qu'il était intrinsèquement motivé mais parce qu'il avait une meilleure probabilité subjective de réussite (et donc une meilleure espérance de gain).

Au mieux, l'hypothèse de substituabilité des motivations est traduite par une fonction d'effort discontinue (James, 2005, Murdock, 2002). James (2005), par exemple, montre que la MI, lorsqu'elle existe, vient compenser la désutilité de l'effort inhérente à tout travail. Mais on peut alors se demander comment l'accomplissement d'une tâche peut à la fois engendrer une désutilité de l'effort (hypothèse standard de désutilité de l'effort) et une utilité de l'effort (hypothèse de MI) ? Comment peut-on en même temps valoriser une action et la déprécier ? Il semble bien que dans ce type de modèles, la MI n'a d'intérêt que parce qu'elle permet de limiter la désutilité initiale de l'effort.

Nous considérons à l'inverse que ce qui est renfermé dans l'hypothèse de MI est une critique du principe de désutilité de l'effort sur lequel la TS est construite. Être motivé intrinsèquement revient à ne pas compter son effort. De ce point de vue Berthe (2001) montre de façon pertinente les insuffisances de cette conception économique où le travail est uniquement désagréable. La seule façon cohérente de considérer une tâche à la fois comme une utilité et une désutilité serait d'avoir une conception physicaliste de l'effort. Dans cette perspective, développée notamment par Charles-Augustin Coulomb (1736-1806) et Jules Amar (1879-1935), il faudrait pouvoir mesurer le rendement de la machine humaine et donc le travail (au sens de la science physique) maximal mis en œuvre par les individus compte tenu de leurs capacités physiques. Le problème de ce type de thèse est double. D'une part, la mesure énergétique du travail est incertaine puisqu'elle dépend de la musculature et de l'intensité générée par l'individu singulier qui la met en marche et, d'autre part, rien ne permet de transposer ce modèle au travail intellectuel, même si celui-ci se traduit également par de la fatigue (Bidet et Vatin, 2009).

C'est ainsi que certains auteurs reposent la question de la désutilité de l'effort comme par exemple Liebenstein (1969) qui construit une fonction d'utilité de l'effort. Ici le comportement salarié n'est plus à la minimisation de l'effort mais à une maximisation de son

utilité de l'effort parce que l'individu n'est pas heureux quand il n'a rien à faire. L'effort n'est alors plus qu'une souffrance, il peut être confortable. Dans le cas du médecin, celui-ci est chaque jour confronté à une multitude de patients avec des besoins et des attentes aussi diverses que complexes. Certains travaux ont mis en avant cette dimension intellectuelle du travail du médecin pour expliquer la pratique des actes gratuits[58] (Batifoulier et al., 2009). L'hypothèse d'un médecin naturellement réticent à l'effort devient problématique.

Malgré ces développements, force est de constater que tous les modèles classiques actuels, du modèle canonique d'arbitrage travail-loisir aux modèles plus développés d'agence, sont fondés sur l'effort-souffrance. Il n'est à aucun moment question d'une MI du médecin. Ce modèle ne peut prendre en compte l'hypothèse de MI puisqu'il réduit les éléments déclencheurs de l'action au seul intérêt (ME) au lieu de les étendre (à la MI). Cette stratégie est d'autant plus problématique qu'elle implique une nécessaire réduction des orientations du comportement.

Paradoxe de théorie économique 5 : La MI médicale n'échappe pas à la désutitlité de l'effort.

Problème de politique économique 5 : On ne sait pas s'il faut inciter ou non le médecin pour qu'il produise l'effort.

3.2.2. MI et désintéressement

La seconde difficulté des modèles à MI et COE de la TSE2 porte sur l'hypothèse d'intéressement limité – comme pour le cas de l'éthique médicale, la MI étant purement instrumentale, elle reste purement intéressée. Prétendre qu'il existe une force intérieure permettant de motiver les individus en plus de la force extérieure revient à postuler l'existence de deux déclencheurs du comportement (MI et ME). Bien que cette proposition relève de la tautologie, elle nous paraît essentielle dans la mesure où les conséquences logiques de l'hypothèse de dichotomie des motivations ne sont pas prises au sérieux par la TSE2. Mais, on ne peut en rester là puisqu'il reste à déterminer ce qui oriente l'action, son contenu concret. Dans le cas d'une action déclenchée par la ME il semble que ce soit l'intérêt lucratif, égoïste et instrumental. Prendre en compte l'hypothèse de MI veut dire que les individus peuvent avoir

[58] La sévérité de la maladie permet de sélectionner les patients par une certaine forme d'attraction intellectuelle.

des comportements orientés spontanément vers autre chose que cette forme d'intérêt. À ce stade de l'argumentation un éclaircissement sur la notion d'intérêt doit être donné.

Selon Caillé (2009), il est possible de distinguer deux formes d'intérêts qui, bien que recouvrant des conceptions très différentes, tendent à être systématiquement confondues dans l'intérêt bien compris. D'une part, « l'intérêt à » est un intérêt instrumental au sens où l'action est accomplie par stratégie ou par contrainte. D'autre part, « l'intérêt pour » est lui passionnel au sens où l'action est plaisante en soi. Dans le premier cas ce qui est recherché est le gain utilitaire dans le second c'est la réalisation de la tâche elle-même. De ce point de vue, l'action désintéressée n'est pas une action sans intérêt mais, une action relevant de « l'intérêt pour ».

À ce titre, la théorie du prestige présentée précédemment relève de l'« intérêt à ». En effet, on pourrait croire le médecin désintéressé puisqu'il refuse une incitation qui modifie le sens de son action. Mais, en réalité, s'il refuse cette récompense c'est pour ne pas subir une perte en termes de prestige social (Siciliani, 2009). On en revient au simple bilan coût-avantage. Dans le modèle à « sur-justification » ce qui motive (et entraîne un COE) est la recherche intéressée d'une utilité maximale obligeant un arbitrage entre profit monétaire et prestige social. Ici les MI et MR sont conditionnelles à la maximisation de l'utilité. Le médecin refuse d'augmenter son niveau d'effort, suite à l'incitation, pour ne pas apparaître aux autres (ou à lui-même) comme un opportuniste, qu'il sait être en réalité. Ce type d'argumentation a déjà fait l'objet d'une discussion critique en économie de la santé (Batifoulier et Gadreau, 2006).

De la même façon, dans le modèle à principal informé (Bénabou et Tirole, 2003), et, par extension, dans les modèles où l'incitation est une information stratégique, la question de la confiance doit prendre une toute autre tournure. Rappelons que dans ce type de modèle la confiance est réduite à une probabilité subjective de réussite de la tâche. Si celle-ci est suffisamment forte alors le principal peut avoir « confiance » ce qui se traduit par une politique d'incitation faible supposée donner « confiance en soi » à l'agent. Ce dernier interprète l'incitation comme un signal positif et pense pouvoir être capable de réaliser la tâche.

Or, il semble bien que cette conception de la confiance en termes de calcul intéressé est particulièrement peu adaptée au monde médical car elle est contradictoire et paradoxale (Batifoulier et Gadreau, 2006). Elle est contradictoire au sens où s'en remettre à son médecin pour le patient relève d'une autre nature que le calcul intéressé des coûts et des avantages

(Karpik, 2003). Mais cette conception est également paradoxale parce que si la confiance est basée sur le calcul et que les partenaires le découvrent, la confiance se vide de son sens. Diluée dans le calcul, la confiance n'est plus utile à la coordination. Donner un rôle à la confiance exige ainsi d'admettre l'existence et la puissance coordinatrice de la notion d'« intérêt pour ».

Les difficultés de la théorie standard à intégrer des concepts nomades comme la MI proviennent de l'hypothèse même d'individu intéressé. Tout désintéressement, affiché ou non, est impossible puisqu'en dernière analyse il s'agit pour l'économiste d'intéressement déguisé. Cette tendance à négliger les actions désintéressées vient de ce qu'Elster (2009) appelle, à la suite de Ricœur, l'herméneutique du soupçon. L'économiste, par définition septique et cynique vis-à-vis du comportement humain, passerait à côté de ce qui constitue une très grande partie de ce comportement, le désintéressement. Pour expliquer cette posture méthodologique Elster explique, en paraphrasant Tolstoï, que « tous les acteurs intéressés le sont de la même façon, tandis que tous les acteurs désintéressés le sont chacun à leur façon » (p. 14).

De ce point de vue s'il y a un COE, ce n'est pas du fait d'un arbitrage entre prestige et revenus financiers. Prendre en compte l'hypothèse de MI et de désintéressement oblige le théoricien à questionner le sens des actions, ce qui change radicalement le registre d'analyse. Ce qui est en cause dans le cas du COE c'est le sens même de l'activité des individus (à l'opposé d'une logique de calcul selon laquelle il existe théoriquement toujours un prix suffisamment fort pour infléchir l'effort des individus). Dès lors, plutôt que de s'attarder à débusquer puis à réduire tous les intérêts au seul « intérêt à », il convient de faire place à l'analyse des différentes dimensions qui structurent et qui motivent l'activité médicale.

À l'évidence, le modèle de l'*overjustification effect* est à l'opposé de ce genre de conception. En effet, c'est par « intérêt à » que les individus s'engagent dans une action désintéressée. On entrevoit tout de suite le paradoxe qui ne peut être résolu qu'en substituant le désintéressement (MI) par un désintéressement instrumental, et, *in fine*, par un pur intéressement (ME). Or, l'idée même de MI est de postuler l'existence de comportements totalement désintéressés.

Paradoxe de théorie économique 6 : La MI et l'effet d'éviction sont intéressés.

Problème de politique économique 6 : On ne sait pas s'il faut inciter le médecin pour stimuler l'effort ou ne pas inciter pour éviter l'éviction.

Conclusion

Ce chapitre a montré que le *mainstream* rencontre de fortes difficultés à introduire les travaux issus de la psychologie sociale sur les motivations dans son corpus théorique sans les dénaturer. En effet, « la psychologie prend le pas sur l'ingénierie mécanique et les mathématiques comme source de nouveaux savoirs, et la notion de personne socialisée vient remplacer celle d'*homo economicus* chère à la théorie classique » (Palpacuer et al., 2010, p.18)

Dans le cadre de la TS2, le travail du médecin reste, comme n'importe quelle autre tâche, une activité dominée par la recherche instrumentale du lucre et du prestige social. Ainsi, le COE n'est qu'une conséquence d'un transfert d'informations qu'occasionne la mise en place d'incitations. Cependant, c'est l'essence même de la MI qui est niée par l'arrière-plan utilitariste des modèles d'agence. Il devient difficile de pousser l'analyse à son terme puisque l'on omet alors une grande partie de ce qu'est l'activité médicale. Elle n'est pas une peine que l'on vend sur un marché comme un autre. Il s'agit plutôt d'une activité qui a des aspects stimulants et intrinsèquement motivants au cours de laquelle plus qu'un simple service, c'est une forme de don de soi qui est vendu.

Le retour aux intuitions de Frey (1997d) sur la question des motivations et d'Arrow (1963) sur la nature de l'activité médicale semble être particulièrement judicieux. En effet, la question posée par l'hypothèse de substituabilité des motivations ouvre de nouvelles perspectives quant à la notion même de MI et de travail. Être intrinsèquement motivé, c'est voir dans son travail autre chose qu'un simple gagne-pain : c'est lui donner du sens. Dès lors, l'introduction, dans une théorie de l'action, de considérations authentiquement éthiques paraît inévitable.

Du point de vue de la politique économique de santé, il semble de plus en plus clair que la discussion sur la rémunération des médecins ne doit plus porter strictement sur les aspects techniques du paiement à la performance mais, sur les déterminants (non financiers) de la motivation médicale (Mannion et Davies, 2008). À la manière d'Andersen (2009) par exemple, en plus des déterminants financiers, il convient de s'intéresser aux comportements influencés à la fois par les normes professionnelles et par l'esprit de service public. Ce dernier point est particulièrement bien documenté par le travail de sciences de gestion de Forest (2008) dans son analyse de la motivation au travail des médecins au sein de l'hôpital public.

L'analyse des sources de la motivation débarrassée du monisme de l'intérêt permettrait ainsi de poser la question essentielle des valeurs, c'est-à-dire celle de la signification de l'activité médicale. Dès lors le COE pourrait s'interpréter comme le résultat de la confrontation entre différentes valeurs associées à l'exercice de l'activité médicale.

Mais, pour en arriver, là il faut se doter d'une théorie robuste permettant d'expliquer le *crowding out effect* sans dénaturer le message anti-utilitariste de la TAD. Comme le veut la théorie de l'évaluation cognitive, les acteurs doivent être dotés de la capacité d'interprétation et de jugement dans les situations où ils sont confrontés à des motivations extrinsèques. Les besoins psychologiques fondamentaux peuvent être une nouvelle source de l'éthique médicale et des besoins de santé (lié notamment à l'autonomie du patient). Une dernière piste de recherche doit mener à une théorie de l'action laissant place aux comportements désintéressés – comme ce que peut proposer le paradigme du Mouvement anti-utilitariste en science sociale (MAUSS) ou de l'Économie des Conventions.

Deuxième partie :

Une évaluation du paiement à la performance médicale

Le difficile passage de la théorie à la pratique

Introduction de la deuxième partie

Malgré les défaillances théoriques du *mainstream* mises en évidence dans la première partie, celui-ci a un impact dans la réalité. Aujourd'hui, le paiement à la performance médicale n'est plus uniquement une question théorique : il se diffuse à travers le monde. Dans cette deuxième partie nous proposons une évaluation de ce dispositif. Nous montrons que ses vertus théoriques ne se retrouvent pas dans l'analyse empirique. Par ailleurs, nous mettons en évidence une partie de la « face cachée » du paiement à la performance médicale. Comme la lune a une face cachée qu'on ne voit jamais, le paiement à la performance produit des résultats inattendus et rarement soulignés, susceptibles de remettre en cause sa pertinence.

Après avoir présenté la généralisation du paiement à la performance dans les systèmes de santé, le chapitre 4 propose une revue de littérature critique sur l'évaluation de ce dispositif. La « mondialisation » du recours au paiement à la performance médical témoigne de sa prétention à être une réponse universelle aux difficultés rencontrées dans systèmes de santé – indépendamment des caractéristiques propres à chaque pays. Pourtant, l'évaluation du paiement à la performance médicale ne produit pas de résultats très encourageants. Bien que de nombreuses études aient été faites, l'efficacité du dispositif n'est pas démontrée. Il est impossible d'attribuer au paiement à la performance médicale l'amélioration de la qualité des soins ou la réduction des dépenses de santé. On doit alors se demander si d'autres politiques publiques moins onéreuses n'auraient pas pu être entreprises pour les mêmes résultats. Enfin, l'évaluation empirique du paiement à la performance met en évidence l'existence d'effets pervers. Par exemple, ce dispositif incite le médecin à sélectionner les patients les plus susceptibles d'améliorer sa performance. Il incite également à sélectionner les tâches rémunératrices au détriment des autres. Plus globalement, l'incitation incite les médecins à se jouer des indicateurs.

Ces travaux insistent sur les difficultés méthodologiques liées à l'évaluation du paiement à la performance médicale. Il est ainsi particulièrement difficile de séparer ce qui relève du paiement (à la performance) de ce qui relève à l'information des médecins. Parfois, les sommes distribuées sont si importantes qu'il n'est pas possible d'avoir un groupe de contrôle à comparer au groupe test. Ces difficultés méthodologiques sont alors le socle pour une interprétation relativement étrange des résultats : plutôt que de remettre en cause le principe de paiement à la performance médicale, même marginalement, la faiblesse des

résultats s'expliquerait avant tout par des problèmes de contexte et de conception du dispositif dans les différents systèmes de santé nationaux.

Le chapitre 5 s'inscrit dans ces travaux sur l'évaluation du paiement à la performance médicale. Nous proposons la première étude empirique sur le dispositif français de Rémunération sur objectif de santé publique (ROSP). En coopération avec le Service médical de l'assurance maladie d'Ile-de-France, nous avons pu analyser les résultats des médecins franciliens. Cependant, ce chapitre ne propose pas une évaluation de l'efficacité du paiement à la performance – au sens où nous ne cherchons pas à déterminer si oui ou non ce dispositif incitatif permet d'améliorer les scores obtenus dans divers indicateurs de qualité des soins ou d'efficience des prescriptions. Notre travail s'oriente vers la détection d'effet pervers ou d'effets inattendus du ROSP.

Deux questions principales traversent notre perspective empirique. D'une part, nous cherchons à évaluer dans quelle mesure l'augmentation de la performance mesurée peut effectivement s'interpréter comme une augmentation de la performance réelle. Autrement dit, est-ce que l'augmentation de la performance signifie nécessairement un « mieux » ? Pour ce faire, nous consacrons une large partie de ce chapitre à l'analyse de la formule de rémunération du ROSP et à la discussion critique des indicateurs retenus par l'assurance maladie. Empiriquement, nous montrons un effet « taille de la patientèle » selon lequel il est possible pour un médecin de faire partie des mieux rémunérés (à la performance) sans pour autant faire partie des médecins ayant les meilleurs scores (de performance). D'autre part, nous nous interrogeons sur la responsabilité ou le mérite qu'ont les médecins dans leurs résultats ROSP. N'est-il pas possible que leurs résultats soient affectés (à la baisse ou à la hausse) par des causes indépendantes de leur action ? A côté de l'effet taille de la patientèle, nous montrons un effet « composition de la patientèle » selon lequel les médecins ayant les meilleures rémunérations à la performance sont aussi ceux qui ont les patients en meilleure santé et les patients les plus pauvres.

Au total, le paiement à la performance médicale n'est pas forcément synonyme d'amélioration de la performance des médecins pour trois raisons principales. D'abord, son efficacité à atteindre les objectifs chiffrés n'est pas démontrée. Ensuite, il n'est pas certain que l'amélioration de la performance mesurée soit désirable. Enfin, il n'est pas certain que la variation des scores soit le fait de la responsabilité ou du mérite du médecin en question.

Chapitre 4

La performance du paiement à la performance médicale. Une revue de littérature

Introduction

La théorie *mainstream* en économie de la santé préconise l'introduction de mécanismes de paiement à la performance (P4P) médicale pour compléter les modes de rémunération traditionnels des médecins. Présumés réticents à l'effort et intéressés, l'idée est d'inciter les médecins à la performance – définie par des indicateurs chiffrés de productivité – dans une situation où les défaillances de marché interdisent de laisser libre cours au principe concurrentiel. L'intérêt théorique de ce type de dispositif est d'améliorer la qualité des soins et de réduire les dépenses en organisant une forme nouvelle de contrôle de la production médicale.

Loin de n'être que des constructions théoriques éloignées de toute réalité, depuis au moins une trentaine d'années, ces idées ont contribué à justifier la réforme de nombreux systèmes de santé dans le monde. L'ambition de ce chapitre est de présenter et d'évaluer ces différents dispositifs de P4P médicale. Les qualités théorisées par les économistes se retrouvent-elles lors des évaluations empiriques ? Parvient-on à réduire les dépenses de santé et à améliorer la qualité des soins par l'utilisation de ce type de dispositif ? Quelle est la performance du paiement à la performance ?

La première section présente un échantillon des multiples programmes de P4P mis en place à travers le monde (1.). Nous nous attarderons plus longuement sur les cas des États-Unis et du Royaume-Uni. Avec les États-Unis nous tenons le berceau du P4P médicale. Cet instrument de régulation de la médecine y est né pour résoudre un problème singulier du système de santé américain : comme nous le verrons plus loin, il est à la fois le système le plus cher sans pour autant se signaler par une qualité des soins élevée. Nous étudierons ensuite plus précisément le système britannique : le *Quality Outcome Framework*. La problématique est ici différente puisqu'à qualité comparable avec les États-Unis, le souci

majeur du système de santé britannique est son sous-financement chronique. Fort de la comparaison de ces deux systèmes de P4P de référence, nous serons en mesure de voir comment à des problématiques de politiques publiques différentes s'est imposée la même solution : le P4P médicale.

La seconde section propose une évaluation du P4P médicale à partir d'une revue de littérature critique (2.). Nous montrons comment l'ensemble de la littérature s'accorde sur le fait que le P4P médicale peine à prouver une quelconque forme d'efficacité – que cela soit pour améliorer la qualité des soins ou pour économiser les ressources rares. De plus, le P4P médicale provoque des effets pervers potentiellement importants (sélection des tâches, sélection des patients, etc.) qui remettent radicalement en cause son utilisation. Pourtant, le principe de P4P n'est pas remis en cause et il se développe. La défaillance apparente du dispositif serait à rechercher dans le contexte qui freine son efficacité. En d'autres termes, l'outil défini théoriquement serait toujours pertinent mais les conditions de son utilisation et l'environnement institutionnel en briderait l'efficacité.

La troisième section fait une revue de la littérature des effets du P4P sur la motivation des médecins (3.). Cette littérature émergente, vise à se doter d'hypothèses nouvelles permettant d'expliquer les faibles résultats du P4P. Nous verrons que les problèmes mis en évidence lorsque l'on s'intéresse à la motivation pour le P4P sont similaires aux précédents : en plus de la revendication d'autonomie de la part de la profession médicale, le contexte apparaît encore comme facteur explicatif des piètres résultats du P4P médicale.

1. Le dispositif de paiement à la performance médicale dans les faits : justifications et internationalisation

Cette section poursuit un triple objectif. D'une part, il s'agit de montrer les causes ayant historiquement justifié la mise en place du P4P médicale – notamment aux États-Unis et au Royaume-Uni. D'autre part, nous voulons décrire précisément ce à quoi ressemble ce type de dispositif. Enfin, nous cherchons à montrer que, quels que soit les contextes nationaux et régionaux, on observe une internationalisation du recours au P4P. La première section présente le cas des États-Unis (1.1), la deuxième celui de la Grande-Bretagne (1.2.), et la troisième section conclue brièvement sur l'internationalisation du P4P médicale (1.3.).

1.1. Les États-Unis : le berceau du paiement à la performance

Nous commençons par présenter le paradoxe du système de santé américain (1.1.1.) afin de comprendre le raisonnement ayant conduit à la mise en place du P4P (1.1.2.).

1.1.1 Le paradoxe américain : une qualité des soins médiocre pour le système de santé le plus cher du monde

Il n'y a pas un système de santé unifié aux États-Unis mais des systèmes de santé. Suivant en cela la tradition politique libérale, l'organisation des soins y est fondée sur le principe de liberté de choix des individus – la liberté étant entendue comme une liberté de marché. L'assurance santé est facultative et repose en majeure partie sur le secteur privé (Galvis-Narinos et Montélimard, 2009). Il en résulte un système de soin très fragmenté où l'intervention publique est résiduelle (Encadré 4.1.). Cette organisation engendre de puissantes inégalités d'accès aux soins. En 2012, 48 millions de personnes n'avaient pas d'assurance santé publique ou privée (soit 15 % de la population totale et 310 millions d'habitants). Parmi elles, on dénombrait 7.2 millions d'enfants entre 0 et 18 ans. Les inégalités sont également variables entre États. Si le taux de non-couverture aux États-Unis est de 15 % en moyenne, il connait une variation comprise entre 4 % pour le Massachusetts et 24 % pour le Texas. Néanmoins, le Massachusetts constitue un cas exceptionnel puisque le deuxième État du classement double le taux de personnes non assurées pour le porter à 8 %.

Encadré 4.1. La fragmentation du système de santé américain

Aux États-Unis, il existe trois niveaux de couverture santé qui s'enchevêtrent :

(i) Les entreprises. En 2012, 48 % de la population était couverte par un contrat d'entreprise, soit près de 149.5 millions de personnes (Kaiser Family Foundation, 2012). Dans ces conditions, les inégalités d'accès à des soins de qualité sont en grande partie expliquées par la qualité de l'emploi. On estime ainsi que 50 % des personnes travaillant dans des entreprises de moins de 10 salariés sont couvertes contre 80 % de celles qui travaillent dans des entreprises de 1 000 salariés (Blavier et Montialoux, 2009). En outre, le degré de couverture est proportionnel à la qualité de l'emploi : les contrats les plus couvrants sont réservés à l'élite salariale tandis que ce sont statistiquement les plus démunis qui en ont le plus besoin. La crise ouverte en 2007 n'arrange rien : si aujourd'hui ce type de couverture représente 48 % de la population, en 2000, il représentait 64 % de la population.

(ii) La couverture fédérale. Celle-ci prévoit quelques filets de sécurité destinés aux plus nécessiteux, aux plus âgés et aux enfants (Cohu et Lequet-Slama, 2007). En 2012, 31 % de la population américaine relevait d'une couverture publique quelle qu'elle soit. Trois politiques publiques d'assurance santé peuvent être distinguées. D'une part, le programme *Medicare*, créé en 1965, est un système universel pour les personnes de plus de 65 ans et les handicapés dans l'impossibilité de travailler. D'autre part, le programme *Medicaid*, créé également en 1965, est lui destiné aux personnes les plus démunies. Le seuil d'éligibilité est fixé à tous les revenus inférieurs au seuil de pauvreté fédéral et à 133 % de ce seuil pour les femmes enceintes et enfants. Enfin, le S*tate Children's Health Insurance Program* (SCHIP) est une politique publique créée en 1997 pour assurer les enfants vivant dans des ménages dont les revenus se situent entre le seuil *Medicaid* et 200 % du seuil de pauvreté fédéral.

(iii) L'État. En plus de l'État fédéral, les États peuvent d'eux-mêmes prendre des initiatives concernant l'assurance santé. Par exemple, le Massachusetts a récemment introduit pour ces habitants un plan d'assurance santé quasi-universel : le *Commonwealth Care* (Weissman et Bigby, 2009). Celui-ci oblige tous les individus à souscrire à une assurance santé et les entreprises à en proposer une. Il étend les seuils d'éligibilité au programme *Medicaid* et prévoit une régulation du marché à partir de standards de couverture et d'accessibilité. En 2012, le Massachusetts est l'État dans lequel la couverture santé est la meilleure des États-Unis : alors qu'un État comme le Nevada compte 23 % de personnes non assurées et 25 % de personnes assurées par l'intermédiaires de programmes publiques (fédéraux ou non), le Massachusetts compte « seulement » 4 % de personnes non assurées pour 35 % de personnes couvertes par des assurances publiques (Kaiser Family Foundation, 2012). Si l'exemple du Massachusetts laisse entrevoir une forte marge de manœuvre pour les États, il arrive souvent que l'État fédéral entrave ce type de projet. Ce fut le cas en 2008 lorsque le gouverneur de Californie Schwarzenegger essaya d'étendre la couverture santé de ses administrés (Isaacs et Schroeder, 2008).

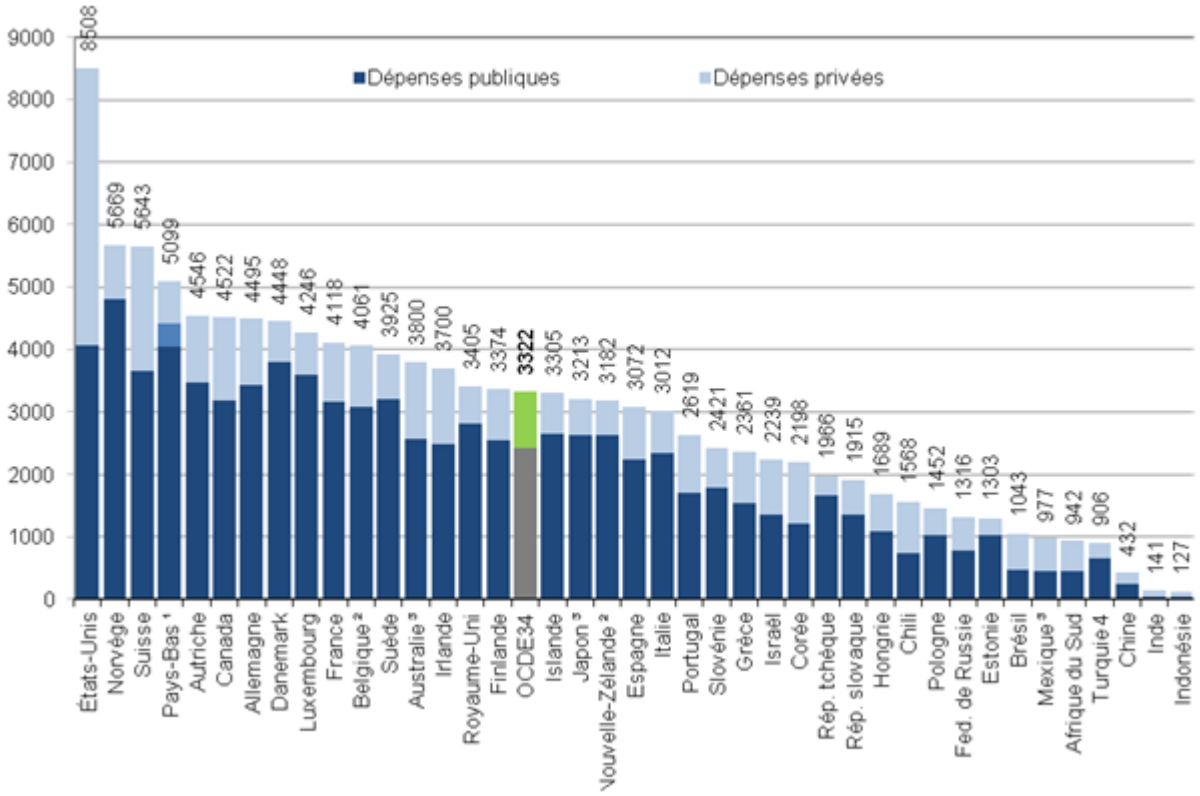
Non seulement, le système de santé américain engendre de fortes inégalités d'accès aux soins mais en plus il se trouve être le système de santé le plus cher du monde. Comme en témoignent les difficultés créées par la récente réforme Obama[59], le coût des soins est une question très sensible aux États-Unis (Brown, 2011). En 2011, les dépenses de santé (publiques et privées) s'élevaient à 17.7 % du PIB contre 9.3 % de moyenne dans les différents pays de l'OCDE (11.6 % pour la France). Le niveau des dépenses de santé

[59] Signée en mars 2010, la loi américaine sur la protection des malades et l'accessibilité des soins rend obligatoire une assurance maladie à compter du 25 décembre 2013, sous peine d'amende. L'amélioration de couverture se fait soit par une extension de *Medicaid* avec subvention de l'État fédéral (25 États républicains sur 50 ont refusé une telle opportunité) soit par la souscription d'une assurance privée subventionnée (plan en bronze, argent, or ou platine selon le prix et le degré de couverture).

américain est ainsi supérieur de plus de 6 points au niveau français (11.6 %) et de plus de 8 points au niveau britannique (9.4 %).

La dépense de santé par habitant des États-Unis s'élève à 8 508 dollars soit deux fois plus que la France (4 118 $) et l'Allemagne (4 495 $) et plus de deux fois et demi la dépense moyenne par habitant des pays de l'OCDE (3 322 $). Une autre caractéristique majeure du système américain porte sur la répartition des financeurs : la majorité des dépenses sont des dépenses privées (52.2 % ; Graphique 4.1.). Comparativement, en France, à peine 23.2 % des dépenses de santé sont privées, le reste relevant du domaine public. La moyenne des dépenses privées des pays de l'OCDE est de 27.3 %. En définitive, sans préjuger de la pertinence des dépenses et du principe même de la dépense, le système de santé américain est particulièrement coûteux et financé majoritairement par le secteur privé.

Graphique 4.1. Dépenses de santé en dollars par habitant selon le financeur (2011)



Source : OCDE, Panorama de la santé 2013 (p. 155)

De façon assez étrange, malgré son coût, le système de santé des États-Unis est également connu pour son niveau moyen de qualité des soins relativement aux autres pays de l'OCDE – comme le montre le tableau 4.1., qui propose quelques indicateurs de qualité des soins issus du *Panorama de la santé* (OCDE, 2013). Les cases sont grisées lorsque le pays en

question est classé à un rang inférieur que celui la moyenne des pays de l'OCDE sur l'indicateur de qualité des soins en question. Par exemple, on observe que les États-Unis ont des taux d'admission à l'hôpital pour l'asthme et les maladies pulmonaires obstructives chroniques sensiblement supérieurs aux taux Européens. Or, d'après les études mentionnées par le *Panorama de la santé*, ces types d'admission sont évitables si le système de soin de première ligne – les soins de ville – est suffisamment performant. Selon ce critère, les États-Unis doivent améliorer leur offre de soin primaire afin d'éviter des admissions inappropriées à l'hôpital. Il existe ainsi des marges de performance. De façon plus générale, si les États-Unis semblent plus performants sur certains domaines – notamment le dépistage du col de l'utérus (1er) –, ils accusent un retard important sur un nombre important de critères de qualité.

Ce tableau recense des critères de qualité des soins mais aussi des indicateurs d'état de santé – c'est-à-dire des indicateurs non liés uniquement à l'action des professionnels de santé. L'état de santé est en effet dépendant de nombreux déterminants tels que l'environnent, l'héritage biologique, les habitudes alimentaires, la pratique sportive, etc. – autant d'éléments en dehors de la portée des professionnels de santé. S'ils ne dépendent pas uniquement de la « performance » des professionnels de santé, ces indicateurs permettent de juger de l'état de santé d'un pays. Or, l'espérance de vie aux États-Unis est de 78.7 ans, soit au 26ème rang des pays de l'OCDE. Dans le même registre, de nombreux indicateurs d'état de santé sont relativement mauvais pour les États-Unis. Ce constat est largement imputable aux inégalités sociales de santé et au défaut de couverture assurantielle. Mais il peut être instrumentalisé pour réformer des conditions d'exercice de la profession et plaider pour la mise en place d'une rémunération au résultat et au paiement à la qualité des soins.

Tableau 4.1.Quelques indicateurs d'état de santé et de qualité des soins, 2011

Critère	États-Unis (classement)	Royaume-Uni (classement)	France (classement)	OCDE
État de Santé				
Espérance de vie à la naissance	78.7 (26ème)	81.1 (15ème)	82.2 (6ème)	80.1
Mortalité due aux maladies cardiovasculaire (/100 000)	127 (22ème)	113 (19ème)	48 (3ème)	122
Mortalité infantile (/1 000)	6.1 (31ème)	4.3 (25ème)	3.5 (18ème)	4.1
Santé du nourrisson (% de nourrissons pesant moins de 2.5 kg la naissance)	8.1 (32ème)	7 (25ème)	6.8 (18ème)	6.8
Prévalence et incidence du diabète pour la population de 20 à 79 ans (/100) (1)	9.6 (6ème)	5.4 (30ème)	5.6 (27ème)	6.9
Incidence du diabète de type 1 parmi les enfants âgés de 0 à 14 ans (/100 000) (1)	23.7 (5ème)	24.5 (4ème)	12.2 (21ème)	17.2
Qualité des soins				
Admission à l'hôpital pour asthme parmi la population adulte (/100 000)	117 (30ème)	60.5 (22ème)	36.6 (14ème)	45.8
Admission à l'hôpital pour MPOC (2) parmi la population adulte (/100 000)	226.1 (22ème)	226.5 (23ème)	101.9 (5ème)	203
Volume total d'antibiotique (dose quotidienne définie / 1 000 habitants)	24.9 (23ème)	18.7 (13ème)	28.2 (26ème)	20.5
Mortalité après un accident vasculaire cérébral (/ 100 admissions)	4.3 (4ème)	10.4 (24ème)	8.5 (16ème)	8.5
Corps étranger laissé dans l'organisme pendant l'opération parmi la population adulte (/100 000)	4.1 (9ème)	5.5 (13ème)	6.2 (16ème)	5.0
Traumatisme obstétrical, accouchement par voie basse avec instrument (/100)	11.1 (18ème)	7.1 (14ème)	2.6 (7ème)	5.4
Traumatisme obstétrical, accouchement par voie basse sans instrument (/100)	1.5 (13ème)	2.5 (17ème)	0.6 (9ème)	1.6
Dépistage du cancer du col de l'utérus (/100)	85 (1er)	68.5 (10ème)	71.1 (14ème)	59.6
Mortalité du cancer du col de l'utérus	2.4 (8ème)	2.7 (14ème)	2.1 (18ème)	3.7
Mortalité du cancer du sein	24.5 (13ème)	30.4 (26ème)	28.8 (24ème)	26.3
Vaccination antigrippale des personnes de plus de 65 ans (/100)	67 (6ème)	74 % (5ème)	55 % (17ème)	50

Source : à partir du Panorama de la santé, OCDE *(2013) (1) Classement inversé (2) MPOC : maladie pulmonaire obstructive chronique*

1.1.2. Les (nombreux) dispositifs américains de paiement à la performance médicale

L'organisation du système de santé américain pose un problème important et légitime : comment est-il possible que le système de santé le plus coûteux du monde produise une qualité des soins et un état de santé si médiocres au regard des résultats des autres pays de l'OCDE ? Empreinte de culture économique *mainstream*, l'une des réponses fréquemment donnée est le déficit de performance des professionnels de santé – dont les médecins. Les médecins ne seraient pas incités à produire de la qualité mais à profiter de leur position dominante. Le P4P serait alors la solution idéale puisqu'elle permettrait non seulement d'améliorer la qualité des soins mais aussi d'éviter un gaspillage indécent de ressources.

Dans une lettre ouverte collective (Berwick et al., 2003), universitaires, administrateurs et , politiques en appellent au paiement à la performance du fait du coût humain et financier des erreurs médicales et de pratiques non alignées sur les standards de la médecine moderne. Il faudrait alors investir dans des systèmes informatiques couplés à des incitations financières permettant de réduire les erreurs et d'adopter des bonnes pratiques – comme c'est le cas d'autres secteurs de l'industrie (les auteurs citent l'aviation ou le nucléaire). Le prestigieux *Institut of Medicine* (IOM) produit des rapports sur l'organisation des soins diagnostiquant le même problème. Dans son rapport *Crossing the Quality Chasm* (2001), l'IOM s'insurge de la faible qualité des soins du système américain causée notamment par la forte variabilité des pratiques des médecins. Il faudrait alors que la rémunération prenne en compte la qualité des soins et l'efficience des pratiques grâce au P4P médicale (IOM, 2006).

C'est sans doute pour cette raison que les États-Unis sont les premiers à avoir expérimenté les mécanismes de P4P médicale (Moore, 1980), si bien que d'après Rosenthal et al. (2007), l'introduction du P4P est (avec la prise en charge des maladies chroniques) l'évolution majeure du système de santé depuis le développement des *Managed care organisations* au début des années 1970. La récente réforme « Obamacare » (Patient Protection and Affordable Care Act, 2010) prévoit à ce titre de développer ces dispositifs jusque dans les programmes publics de *Medicare*.

Le développement important du P4P – dans le domaine des soins primaires et à l'hôpital (Kahn III et al., 2006, Sautter et al., 2007), dans le public et dans le privé, etc. – ainsi que l'organisation décentralisée du système de santé américain rend extrêmement difficile un exercice de description exhaustif des dispositifs. Au-delà des grands programmes très

médiatisés (Epstein, 2006, Encadré 4.2.), il est possible de scinder l'analyse selon que l'assureur soit privé ou public.

- Le secteur privé : L'article de Rosenthal et al. (2006) permet d'avoir une vue d'ensemble de l'utilisation des mécanismes de P4P dans les réseaux intégrés de soins du secteur privé. Nous cherchons ici à avoir une idée de la proportion de patients qui peuvent avoir à faire à un médecin rémunéré à la performance. L'étude a été réalisée sur plus de 250 réseaux de soins portant sur 41 aires métropolitaines dispersées sur tout le pays (avec un taux de réponse supérieur à 95 %). Réalisée entre juillet 2005 et janvier 2006, elle porte sur des réseaux choisis de façon aléatoire couvrant au moins 100 000 personnes. Le résultat principal est la confirmation statistique de l'idée selon laquelle le P4P est un mode de régulation désormais largement implanté aux États-Unis. En effet, 52.1% des réseaux représentant plus de 80 % de la population étudiée utilisent des dispositifs de P4P. L'usage de ces mécanismes n'est pas réservé à la médecine de ville puisque sur 126 réseaux utilisant le P4P, 89.7 % l'utilisent pour les médecins et 38 % pour hôpitaux. Lorsqu'il s'adresse au médecin, le P4P rémunère soit les individus (13.3 %), soit les groupes (56.6 %), soit les deux (23%). Toujours dans ces cas-là, 100 % des programmes comportent des indicateurs de qualité clinique, 80.4 % des indicateurs d'utilisation de technologies, 68.6% des indicateurs de satisfaction des patients et 42 % des indicateurs d'efficacité clinique. Le tableau 4.2. précise les résultats obtenus.

- Le secteur public : Dans le secteur public, le rapport *Pay-for-Performance in State Medicaid Programs* (Kuhmerker et Hartman, 2007) offre une première approche intéressante. Ce rapport analyse les plans actuels et prévus de P4P dans le cadre des programmes *Medicaid* de chaque État. Les données recouvrent 47 États, près de 97 % du total du budget de *Medicaid* (296 milliards de dollars) et plus de 96.3 % des bénéficiaires États-uniens. L'un des auteurs (Kathryn Kuhmerker) étant elle-même ancienne directrice du programme *Medicaid* de New York, les données ont été recueillies notamment à l'aide d'entretiens avec les directeurs du programme dans chaque État. Selon ce rapport, au premier juillet 2006 plus de la moitié des États étudiés (n=28) avaient mis en place dans leur programme *Medicaid* un ou plusieurs dispositifs de P4P. La proportion passe à 85 % si l'on compte les plans prévus mais pas encore réalisés. Le P4P est donc bien implanté dans le secteur public et cela depuis assez longtemps. En effet, plus de la moitié des programmes (n=100) ont plus de 5 ans d'existence – 9 programmes en ayant plus de 10 au moment de l'étude (soit 1996).

Tableau 4.2. Les programmes de P4P médicale dans le secteur privé aux États-Unis

	Programmes (en %)
Caractéristiques des programmes étudiés	
Indicateurs cliniques cibles	
Pression sanguine	35 (31.0)
Asthme	80 (70.8)
Diabète	98 (86.7)
Antidépresseurs	42 (37.2)
Mammographie	85 (75.2)
Cholestérol	70 (61.9)
Bonus associés	
Moyenne du bonus	
<5 %	40 (35.4)
≥5%	34 (30.1)
Inconnu	39 (34.5)
Bonus maximum	
<5 %	32 (28.3)
≥5%	47 (41.6)
Inconnu	34 (30.1)
Structure du bonus	
Uniquement pour les plus performants	36 (31.9)
Atteinte d'objectifs prédéterminés	70 (61.9)
Amélioration de la performance	23 (20.4)
Inconnu	8 (7.1)

Source : à partir de Rosenthal et al. (2006), p. 1899

Les programmes de P4P ont longtemps été limités à *Medicaid*. Désormais, avec la « loi Obama », les programmes de *Medicare* s'ouvrent à ces dispositifs (Nix, 2013). La loi prévoit en effet deux dispositifs renforçant la logique de P4P. D'une part, le *Value based purshasing* à l'hôpital est destiné à réduire les financements des *Diagnostic-related groups* (DRG) et à redistribuer les gains aux hôpitaux performants (sur la base de 12 indicateurs cliniques). D'autre part, le *Physician Value-Based Payment Modifier* prévoit la mise en place du P4P pour la médecine de ville à partir de 2015. L'objectif de neutralité budgétaire, si important pour les États-Unis, est alors assuré par la mise au point d'un système qui compense la sur rémunération des médecins gagnants par la sous rémunération des médecins perdants. Comme demandé notamment par l'IOM et Berwick et al. (2003), *Medicare* est destiné à devenir le vaisseau amiral diffusant les pratiques de rémunération à la performance. Dans leur lettre ouverte, les auteurs en appelaient à l'acheteur de soin le plus puissant du système – *Medicare* – pour offrir des incitations qui permettraient de faire une avancée décisive dans

l'amélioration du système de santé. En ce sens, *Medicare* est voué à devenir l'opérateur de la standardisation des soins aux États-Unis.

Encadré 4.2. Trois programmes de P4P médicale emblématiques

- *Bridges to Excellence* : introduit en partenariat avec General Electric dans le Massachusetts, ce programme récompense des médecins qui se dotent de modes d'organisation permettant d'améliorer la qualité des soins (dossier médical, coordination, guides de bonnes pratiques, etc.).

- The *Integrated Healthcare Association's Physician Payment Program* : introduit en 2004 en Californie et couvrant 8 millions de personnes pour un coût de 100 millions de dollars, ce programme repose sur des contrats de performance entre des groupes de médecin et les assureurs parties prenantes. La rémunération dépend d'un score de performance calculé à partir de mesures cliniques, de la satisfaction des patients et de l'usage des technologies de l'information. Les indicateurs cliniques portent sur les soins préventifs (mammographies, dépistage du cancer du col de l'utérus, etc.) et le traitement des maladies chroniques (asthme, diabète, etc.)

- *Anthem Blue Cross and Blue Shield Plan* : Lancé en 1999 dans le New Hampshire, ce programme rémunère les médecins qui fournissent des soins préventifs (prévention des différents types de cancer, vaccins, etc.) et qui améliorent le suivi du diabète.

Il serait présomptueux de vouloir conclure sur le cas des États-Unis uniquement avec ces quelques pages. Néanmoins, il est possible de résumer et de questionner le raisonnement général ayant conduit à une dynamique de généralisation progressive du P4P. Ce type de dispositif tire sa légitimité de l'inefficacité avérée du système de santé américain. Bien qu'il soit le plus cher du monde, il produit une qualité de soins médiocre. Le P4P est apparu comme une réponse crédible à ces difficultés alors qu'elles relèvent essentiellement de problèmes structurels d'organisation de l'accès aux soins. L'accent mis, au travers du P4P, sur la qualité des soins qui serait insuffisante relègue au second plan la thématique des inégalités de santé. Le P4P cible des problèmes d'organisation clinique des soins et non des difficultés imputables aux barrières financières. Or, les mauvais indicateurs d'état de santé sont liés aux inégalités sociales (Wilkinson et Pickett, 2010).

Ce contexte politique qui entoure le rôle que joue le P4P aux États-Unis rend son évaluation malaisée. C'est pourquoi, nous nous concentrons maintenant sur le programme britannique de *Quality Outcome Framework*.

1.2. Le paiement à la performance au Royaume-Uni : le *Quality Outcome Framework*

Dans un premier temps, nous analysons les éléments distinctifs du système de santé britannique qui se caractérise par la centralisation au sein du *National Health Service* (NHS), des restrictions budgétaires conduisant au rationnement des soins et une « culture » de la performance (1.2.1). Ensuite, non présenterons le système de P4P britannique : le *Quality Outcome Framework* (1.2.2.).

1.2.1. L'organisation des soins au Royaume-Uni

À sa création en 1948, le NHS prévoit un accès aux soins universel financé par l'impôt et un mode de gestion uniforme tant en ambulatoire qu'à l'hôpital. Les médecins spécialistes se voient proposer des contrats salariés pour exercer exclusivement à l'hôpital – bien que soit prévu la possibilité d'exercer marginalement en libéral avec la création de lits privés à l'hôpital. La médecine de ville est réservée aux médecins généralistes et autres professionnels médicaux et paramédicaux rassemblés au sein de cabinets gérés à parité entre médecin et représentants des acteurs politiques locaux. Les médecins généralistes sont payés à la capitation à laquelle viennent s'ajouter quelques forfaits et allocations couvrant les dépenses immobilières.

En dépit d'objectifs initiaux exigeants en termes d'accès aux soins et de qualité des soins, l'histoire du NHS va rapidement devenir l'histoire de son contrôle budgétaire. En effet, dès les premières années les budgets vont augmenter plus vite que prévu et la réaction sera de mettre en place des co-paiements visant à partager le financement avec les patients. Dès 1951 la gratuité complète n'existait plus et la hausse du budget était contenue en dessous de la hausse du PIB – alors que dans les autres pays de l'OCDE les dépenses de santé croissaient plus vite que le PIB. L'encadré 4.3. retrace l'évolution de la régulation du système de santé britannique, en recherche constante d'efficacité.

Malgré les réformes destinées à améliorer l'« efficience » du système de santé britannique, il est apparu dès les premières années du XXIème siècle que le plus grand mal du NHS était l'insuffisance de son financement. L'objectif du premier ministre Tony Blair, dès son élection en 2000, était alors d'amener le budget du NHS en termes de pourcentage de PIB à un niveau comparable aux autres grands systèmes de santé. Il s'agit de réduire les listes d'attente aux soins spécialisés et d'améliorer la qualité des soins par une politique d'investissement massive. Mais cette réforme n'a pas été « gratuite » : elle s'est accompagnée de l'élargissement de la mise en concurrence et du pilotage par des indicateurs chiffrés (dont la systématisation du P4P n'est qu'un élément). En ce sens, le gouvernement Blair poursuit la logique de quasi-marché et de NPM des conservateurs.

Encadré 4.3. La recherche de l'efficience : une pluralité de réformes

Les premières réformes véritablement marquantes commencent au début des années 1970. En 1973 le gouvernement cherche à réorganiser l'organisation des hôpitaux en confiant la gestion aux administratifs au détriment du pouvoir médical jugé trop peu économe. De nouvelles mesures de maîtrise des dépenses sont prises en 1976 avant l'élection en 1979 de Margaret Thatcher – qui promettra dans un premier temps de ne pas baisser le budget du NHS. À ses débuts, les mesures prises concernent l'organisation du système dans la logique où « il ne faut pas dépenser moins, mais dépenser mieux ». Les conditions d'exercice en libéral sont élargies et des avantages fiscaux sont mis en œuvre pour les entreprises qui offrent des assurances de santé privées à leurs employés.

En 1983, de nouvelles mesures d'efficience sont mises en place et le budget du NHS est réduit de 1% ce qui conduira à une réduction drastique de l'activité en fin d'année (fermeture de lits et annulation d'opérations). Plutôt que de revenir sur la baisse du budget pour relancer l'activité, le gouvernement choisira d'encourager la vente du foncier (le parc hospitalier) et de mettre en place – pour la première fois – un pilotage par des indicateurs de performance. Mais, les problèmes de dépenses ne s'estompent pas et, à partir de 1984, de nouvelles responsabilités de gestion sont transférées des médecins vers les administratifs, ces derniers étant responsabilisés par la contractualisation et l'usage d'indicateurs de performance. Cependant, selon les réformateurs, les indicateurs de performance ne portant pas sur des données cliniques, le pouvoir médical restait suffisamment fort pour éviter de jouer le jeu de l'efficience administrative – uniquement endossé par les responsables administratifs.

Pour répondre à ces problèmes le début des années 1990 voit la modification en profondeur de la stratégie de réforme britannique. Plutôt que de vouloir centraliser le contrôle, sous l'influence du *New Public Management* (NPM), le gouvernement va mettre en œuvre une politique qui s'amplifie encore aujourd'hui : le quasi-marché. L'idée est de séparer le financeur de soins (NHS) de l'acheteur et/ou des producteurs de soins : les cabinets de médecins ambulatoires se voient confier les financements nécessaires pour rémunérer la médecine hospitalière. Les hôpitaux de leur côté sont dotés d'une autonomie de gestion plus forte. La mise en concurrence des hôpitaux entre eux doit permettre de tirer les prix payés par les cabinets de médecins généralistes vers le bas – ceux-ci récupérant la différence entre ce que paie le NHS et ce qu'ils dépensent effectivement pour les soins primaires et hospitaliers.

Cette période voit également le développement de l'individualisation des modes de rémunération avec l'autonomisation des établissements qui peuvent mettre en place –de façon encore marginale – la rémunération à la performance. En effet, bien que la loi de 1991 donnait aux établissements autogouvernés la liberté de moduler les salaires et les conditions de travail des médecins, cette possibilité n'était pas utilisée par les établissements qui respectaient les recommandations des autorités professionnelles – pourtant les salaires toutes professions comprises représentent ¾ du budget du NHS.

A cette époque, les médecins généralistes commençaient à se voir proposer l'accomplissement d'objectifs de santé publique quantifiés (par exemple, un pourcentage de vaccination infantile) contre un paiement forfaitaire. La réforme de 1991 s'articule autour d'un système d'accréditation et de démarche qualité encadré par le *British Standards Institute* et le *King's Fund*. Une petite vingtaine d'hôpitaux expérimentent un programme appelé *Total Quality Management* destiné à améliorer la qualité des soins jugée de loin la plus mauvaise des grands pays Européens. Des comités d'audit médical sont créés dans chaque district sanitaire (*District Medical Audit Advisory Committe*). Toujours dans la même optique, le NHS octroie un financement au *Royal College of Physicians* pour élaborer des normes et organiser des audits à grande échelle. Si la thématique du contrôle-qualité et de la standardisation est ancienne, dans les faits cette stratégie a rapidement trouvé ses limites puisque les médecins décidaient eux-mêmes de leur propre évaluation – celle-ci n'étant pas assortie de sanction. Néanmoins, sur le long terme cette période constitue une double victoire de la réforme dans la mesure où désormais les médecins acceptent les principes d'évaluation quantifiée et d'élaboration de normes de bonnes pratiques.

Dès 2002, le NHS délègue la définition des politiques de santé et le contrôle de la performance à une nouvelle organisation : les *Strategic Health Authorities* (SHA) qui elles-mêmes financent 152 *Primary Care Trusts* (PCT) (les cabinets ambulatoires) chargés d'organiser au niveau local les soins primaires, l'achat des soins secondaires et la mise en œuvre de la politique de santé publique. Le montant des allocations aux PCT est fondé sur la

base d'une capitation modulée selon les caractéristiques démographiques et socio-économiques de la population et les coûts locaux du travail et du capital. Ce sont désormais ces PCT qui contractent avec les cabinets de médecine générale.

La suite des années 2000 sera le témoin du prolongement de la logique de quasi-marché et de NPM : hausse de l'autonomie des établissements, mise en place de la tarification à l'activité et externalisation de nouvelles missions à des cliniques privées. Le livre blanc *Delivering the NHS* (2002) annonce le futur du NHS avec notamment le développement de programmes de financement de type partenariat public-privé pour la construction d'établissements (*Private finance initiative*) et l'externalisation d'une partie des opérations sur liste d'attente au secteur privé (*Diagnostic and Treatment Centres*). L'universalité du NHS de 1948 s'éloigne chaque jour un peu plus laissant place au secteur privé – réservé aux plus favorisés. Cependant, le secteur privé reste trop peu développé pour intégrer ces nouvelles activités. En conséquence, un nombre croissant d'actes chirurgicaux sont réalisées à l'étranger (250 000 actes).

En médecine de ville le rôle des cabinets est renforcé toujours dans l'optique de faire jouer la concurrence pour l'achat des soins hospitaliers. En 2008, 80 % du budget du NHS est géré par les cabinets. Ils ont désormais la responsabilité sur tous les soins de premier recours qu'il est possible maintenant d'acheter soit au privé soit au public. La concurrence se fait également sur la qualité avec la multiplication de normes (*Health care standards* scindés en normes minimales - *Core standards* – et normes avancées – *Developmental standards*).

C'est dans ce contexte de réforme permanente et de pilotage par la performance que les médecins généralistes britanniques ont vu l'introduction généralisée du P4P en 2004. À nos yeux, l'un des points les plus importants à souligner est la différence structurelle des systèmes américain et britannique. En effet, le P4P a été introduit au États-Unis pour résoudre le paradoxe d'un système de santé cher pour des états de santé médiocres. On peut discuter de l'opportunité d'imputer le faible état de santé des américains à la qualité des pratiques, mais le paradoxe justifie en première approximation une politique de maîtrise des coûts. À l'inverse, dans le cas de la Grande-Bretagne, cette justification n'est pas pertinente, le NHS étant sous-financé dès les années 1970 jusqu'à aujourd'hui encore.

1.2.2. Le dispositif de paiement à la performance britannique

La généralisation du P4P en médecine de ville en Grande-Bretagne a été réalisée sous le gouvernement travailliste de Tony Blair en 2004. La réforme est l'aboutissement des négociations entre le NHS et la *British Medical Association* (BMA) – seul syndicat de médecin habilité à négocier les contrats passés avec le NHS. Le dispositif, nommé *Quality and Outcome Framework* (QOF), a en outre bénéficié de travaux d'experts et d'universitaires extérieurs à la BMA. En principe, le QOF est fondé sur le volontariat mais, étant donné l'importance des sommes en jeu, l'ensemble des cabinets de médecins ont adhéré au programme. Comme aux États-Unis l'incitation financière est censée augmenter la motivation des professionnels pour améliorer la qualité des soins et l'organisation des cabinets.

Si l'extrême diversité des systèmes de P4P aux États-Unis rendait vaine la tentative de présentation détaillée des dispositifs, le cas britannique permet de décrire ce à quoi ressemble concrètement le P4P. Le dispositif est présenté dans un manuel publié par le NHS réactualisé chaque année. La version que nous utilisons est la suivante : *2013/14 general medical services (GMS) contract quality and outcomes framework. Guidance for GMS contract 2013/14*, NHS England, BMA, NHS Employers.

Les indicateurs composant le QOF sont classés en 4 domaines d'intérêt du NHS : domaine clinique, domaine de santé publique, domaine de qualité et de productivité et domaine de satisfaction des patients (Tableau 4.3.). À chaque domaine correspond une variété d'indicateurs, ces derniers étant associés à un certain nombre de points (la pondération exprimant la valeur accordée par le NHS à chaque objectif). Le principe de rémunération est simple : les médecins obtiennent des points pour chaque indicateur pour lesquels ils réussissent à atteindre un niveau objectif déterminé *ex ante* ; à chaque point gagné correspond une somme d'argent.

Tableau 4.3. Synthèse des indicateurs du QOF

Domaine	Nombre d'indicateurs	Nombre de point associé
Clinique	93	610
Santé publique	18	157
Qualité et de productivité	9	100
Satisfaction des patients	1	333
Total	121	900

Dans le domaine clinique les indicateurs sont organisés par catégorie de maladie. Le choix de ces indicateurs est réalisé en fonction de trois critères. D'une part, il faut que la

responsabilité de gestion de la maladie en question incombe principalement au médecin généraliste et, plus largement, à l'organisation de soins primaires en cabinet – de façon à ne pas attribuer au médecin un résultat pour lequel il n'est pas responsable. D'autre part, l'indicateur en question doit être légitimé par l'existence de preuves solides en termes de bénéfices de santé – notamment via l'existence de *guidelines* nationales unanimement acceptées par la profession. Enfin, les indicateurs sont plus probablement mis en place lorsque la maladie en question est une priorité de santé publique.

Concrètement, pour chaque maladie au moins deux types d'indicateurs sont mis en place. Un premier type d'indicateur a pour objectif d'inciter à la tenue de registres informatiques pour chaque patient. L'idée en filigrane est de construire un dossier médical informatisé du point de vue individuel et de pouvoir ensuite agréger les résultats pour obtenir des indicateurs d'état de santé de la population. Un second type d'indicateur est plus spécifiquement lié à la prise en charge de la maladie. À titre d'exemple, le tableau 4.4. présente les indicateurs liés au traitement de l'asthme[60].

Tableau 4.4. Le traitement de l'asthme dans le QOF

Asthme		
Indicateur	**Points**	**Seuil de réussite**
Dossiers		
Le contractant établit et maintient un registre de patient asthmatique, excluant les patients asthmatiques à qui l'on n'a pas prescrit de médicaments liés à l'asthme dans les 12 mois précédents.	4	Sans objet
Diagnostic initial		
Le pourcentage de patients asthmatiques âgés de 8 ans ou plus, dans les registres, avec des mesures de variabilité ou de réversibilité enregistrées entre les 3 moins avant ou n'importe quand après le diagnostic.	15	45-80 %
Gestion continue		
Le pourcentage de patients asthmatiques, dans les registres, qui ont eu une consultation liée à l'asthme dans les 12 moins incluant une évaluation de contrôle d'asthme.	20	45-70 %
Le pourcentage de patients asthmatiques ayant entre 14 et 19 ans, dans le registre, pour qui est enregistré le statut de fumeur dans les 12 mois précédents	6	45-80 %

[60] Voici les autres pathologies répertoriées pour le domaine clinique : *Atrial fibrilation, Secondary prevention of coronary heart disease, Heart failure, Hypertension, Peripheral arterial disease, Stroke and transient isachaemic attack, Diabetes mellitus, Hypothyroidism, Asthma, Chronic obstructive pulmonary disease, Dementia, Depression, Mental health, Cancer, Chronic kidney disease, Epilepsy, Learning disability, Osteoporosis : secondary prevention of fragility fractures, Rheumatoid arthritis, Palliative care*.

Il est intéressant de noter que tous les indicateurs ne sont pas à proprement parlé des objectifs chiffrés dans la mesure où il n'y a pas de seuil de réussite. Par exemple, la tenue d'un registre pour les patients asthmatiques relève d'un traitement binaire de la performance : soit l'action est réalisée, soit elle ne l'est pas. Par contre, le pourcentage de patients asthmatiques pour lesquels on connait le statut de fumeur dépend d'un seuil de réussite global (entre 45 et 80 %) en dessous duquel il n'y a pas de rémunération.

Le domaine de santé publique a été introduit en avril 2013 afin de reconnaître l'engagement des cabinets de généralistes dans le domaine de la prévention. Le domaine qualité et productivité produit des indicateurs binaires d'organisation du cabinet (Tableau 4.5.). Par exemple, 5 points sont accordés aux cabinets qui réalisent une synthèse de données portant sur les soins de seconde ligne pour chaque patient inscrit sur ses listes. Enfin, le domaine de satisfaction du patient ne comporte qu'un seul indicateur lié au temps de consultation des patients.

Tableau 4.5. Les indicateurs de qualité et de productivité dans le QOF

Domaine de qualité et productivité	
Indicateur	Points
Le contractant procède à une synthèse de données sur les soins hors ceux ambulatoires, pour les patients enregistrés sur sa liste, fournie par le NHS	5
Le contractant participe à des examens externes par les pairs avec d'autres contractants du même *Clinical commissioning groups* pour comparer ses données sur les soins de seconde ligne avec les autres cabinets. Le contractant accepte de procéder à des améliorations de la conception de ses services.	5
Le contractant s'engage à développer et à suivre trois sentiers d'amélioration des soins, agréés par le NHS, portant sur la gestion des patients dans le cadre des soins primaires pour éviter l'adressage inapproprié.	11
Le contractant procède à une synthèse sur les admissions en urgence, pour les patients enregistrés sur sa liste, fournie par le NHS.	5
Le contractant participe à des examens externes par les pairs avec d'autres contractants du même CCG pour comparer ses données sur les admissions en urgence avec les autres cabinets. Le contractant accepte de procéder à des améliorations de la conception de ses services.	15
Le contractant s'engage à développer et à suivre 3 sentiers d'amélioration des soins, agréés par le NHS, dans la gestion et le traitement des patients afin d'éviter les admissions en urgences.	28
Le contractant procède à une synthèse sur les accidents et la fréquentation des urgences pour les patients inscrits sur sa liste, fournie par le NHS. La synthèse inclura des informations sur l'opportunité de l'accès aux urgences en liaison avec la problématique des accidents et de la fréquentation des urgences.	7
Le contractant participe à des examens externes par les pairs avec d'autres contractants du même CCG pour comparer ses données sur les accidents et la fréquentation des urgences. Le contractant accepte de procéder à des améliorations de la conception de ses services.	9
Le contractant met en place un plan d'amélioration pour réduire les accidents et la fréquentation inutile des urgences.	15
Total	**100**

Le dernier point d'importance pour comprendre le QOF porte sur la possibilité d'exclure des patients du calcul des indicateurs. Cette procédure, nommée e*xception reporting*, est rendue possible par une série de critères dont voici les plus importants. (1) Lorsque les patients ont refusé de consulter en cabinet au moins trois fois dans l'année. (2) Lorsqu'il n'est pas pertinent d'examiner les indicateurs de maladie chronique à cause de circonstances particulières (par exemple, pour les malades en phase terminale ou pour les plus fragiles). (3) Lorsque les patients sont nouvellement inscrits ou nouvellement diagnostiqués (pour les maladies chroniques) et qu'il serait illégitime d'imputer leur état au cabinet en question. Cette exclusion est par exemple possible pour les indicateurs de pression sanguine et de taux de cholestérol. (4) Lorsque les patients ont besoins de doses de traitement différentes

des niveaux standards. (5) Lorsque la prescription standard d'une médication n'est pas médicalement appropriée (par exemple, pour les allergies ou lorsqu'il existe une contre-indication). (6) Lorsque les patients ne tolèrent pas la médication. (7) Lorsque les patients n'acceptent pas de subir des examens ou un traitement et que ce désaccord informé (*informed dissent*) a été consigné après discussion avec le patient. (8) Lorsque les services de soins secondaires sont indisponibles.

Le QOF est certainement le système le plus important du monde par l'ampleur des financements et le nombre d'indicateurs. À ce titre, il sert de référence à la diffusion mondiale des mécanismes de P4P à la quelle est consacrée la section suivante.

1.3. Une synthèse des différents dispositifs de paiement à la performance

Au plan international, Schoen et al. (2009) nous permettent de mesurer à quel point les mécanismes d'incitations financières auxquels sont soumis les médecins se sont développés tout en se diversifiant. En effet, en plus des mécanismes de P4P, les pouvoirs publics introduisent d'autres types d'incitations financières comme les forfaits qui, à la différence du P4P, sont des rémunérations non conditionnées à des résultats. En France par exemple, les médecins prenant en charge des patients atteints d'une affection longue durée (mucoviscidose, paraplégie, etc.) bénéficient d'un forfait de 40 € par an et par malade sans que cette rémunération ne soit conditionnée à l'atteinte d'objectifs prédéfinis. Le Tableau 4.6. donne la probabilité dans différents pays en 2009 qu'un médecin se voit proposer une incitation forfaitaire selon différents critères : la prise en charge d'une maladie chronique, la satisfaction des patients ; la réalisation d'un objectif de performance de type P4P.

Tableau 4.6. Incitation et P4P dans les grands systèmes de santé internationaux

Type d'incitation / Pays	Maladie chronique (%)	Satisfaction des patients (%)	P4P (%)	Toutes incitations confondues (%)
USA	17	19	28	33
Suède	2	4	5	10
Norvège	9	1	1	35
Royaume Uni	82	49	84	98
Pays-Bas	61	4	23	81
Italie	56	19	51	70
France	42	2	6	50
Allemagne	48	4	6	58

A partir de Schoen et al. (2009

Nous pouvons noter la faiblesse dans la plupart des pays de la probabilité pour un médecin de percevoir une incitation de type P4P. Même si cette probabilité dépasse les 50% pour les médecins italiens et est supérieure à 80% pour les médecins britanniques, le P4P semble assez peu répandu dans les autres pays. Mais, il existe dans tous les pays étudiés par Schoen et al. Le tableau 4.7. présente une liste des principaux programmes de P4P médicale (hors États-Unis).

Tableau 4.7. Les principaux programmes de P4P médicale

Nom du programme	Pays d'origine	Année d'introduction
Advancing Quality	Grande-Bretagne, régional	2008
Clalit P4P	Israël	1998
Clinical Practice Improvement Pay	Australie, Queensland	2008
Ergbnis Orientierte Vergütung	Allemagne, régional	2001
Maccabi P4P	Israël	2001
National Health Insurance P4P	Taïwan	2004
Performance Management Program	Nouvelle-Zélande	2006
Physician Integrated Network	Canada, Manitoba	2004
Practice Incentive Program	Australie	1998
Primary Care P4P	Pays-Bas, régional	2005
Primary Care Renewal Models	Canada, Ontario	2007
Program of Quality Improvement	Argentine, Buenos Aires	2005
Quality and Outcomes Framework	Grande-Bretagne	2004

A partir d'Eijkennar (2012, complété par nous)

À cette liste de programme de P4P médicale, il est possible d'ajouter un grand nombre d'autres pays ayant expérimenté ou expérimentant encore ce type de dispositif : Argentine (Rubinstein, 2009), Nicaragua (Jack, 2003), Rwanda (Logie et al. 2008), Canada (Canadian Medical Association, 2010), Allemagne et Italie (Fiorentini, et al. 2011), Suisse et Philippines (Peabody et al. 2011), Espagne (Gené badia et al. 2007, Benavent et al. 2009), Australie (Duckett et al. 2008 ; Scott, 2007), Nouvelle Zélande (Buetow, 2008), Israël (Gross et al, 2008, Gross, 2003), Taiwan (Lee et al. 2010), etc.

Le modèle de légitimation du P4P médicale est à chaque fois le même. La maîtrise des dépenses de santé et l'impératif d'amélioration de la qualité - indépendamment des positions relatives des pays eu égard à ces critères – sont des objectifs indiscutables – en tous cas indiscutés – qui pourraient être atteints par la mise en place de P4P dont la théorie économique standard vante les mérites. Avec la mise en place plus ou moins récente de ces préconisations, il convient alors de procéder à une évaluation de ces dispositifs.

2. Quelle performance du paiement à la performance ?

La première partie de cette thèse a remis en cause la pertinence le principe du P4P par une critique de la théorie économique *mainstream* qui en fournit une légitimation. Dans la première section de ce chapitre, nous avons vu à quel point ce type de dispositif s'est développé dans le monde. Il est maintenant temps de procéder à une évaluation du P4P. Est-ce que ce dispositif permet d'atteindre les objectifs qu'il se donne ? Rend-il les médecins performants ? Répondre à cette question pose un problème méthodologique important. En effet, la littérature étudiant l'impact du P4P est très large et s'intéresse à divers effets potentiels du P4P. On y trouve une variation importante de la conduite des travaux (méthodologie de recherche), des effets testés (maladie, instrument de rémunération, etc.) et des pays où sont menées les études. Cette situation a conduit depuis de nombreuses années à la production par divers auteurs et équipes de recherche de document de synthèse ou de revue de la littérature.

Il est alors pertinent de se demander ce qu'il faut faire de cette littérature. Faut-il l'étudier exhaustivement ? Cette possibilité n'est pas dénuée de pertinence. Ce serait le meilleur moyen d'évaluer comparativement la qualité des différentes expériences sur le sujet. Deux raisons nous poussent à adopter une position différente. D'une part, l'ambition de revisiter une à une l'ensemble des expériences menées mérite une thèse à elle seule. D'autre part, vouloir réaliser ce travail signifierait implicitement que les différentes revues de la littérature existantes n'apportent pas suffisamment à la connaissance du sujet. Or, nous pensons que l'on peut retirer de l'examen des différentes revues de la littérature une lecture originale du champ. Si on prend au sérieux le résultat principal de cette littérature – le P4P est incapable de faire la preuve de son efficacité et il peut produire des effets pervers –, il est intéressant de porter un regard critique sur la façon – relativement uniforme – dont les résultats sont présentés dans la littérature : selon cette dernière, le contexte et la mise en application de la théorie (*design*) sont imparfaits – et non la théorie elle-même.

Nous commençons par présenter la littérature permettant d'évaluer l'efficacité du P4P (2.1.). Ensuite, nous nous attardons sur la question des effets pervers engendrés par ce mode de rémunération (2.2). Enfin, nous proposons une synthèse critique du « contexte » et du *design* désignés comme facteurs explicatifs de l'(in)efficacité du P4P (2.3.).

2.1. Une évaluation du P4P : efficacité, conception et contexte

Dans un premier temps, nous analysons l'efficacité du P4P au regard des objectifs pour lesquels il a été mis en place (2.1.1.). Ensuite, nous nous intéressons à ce que la littérature nomme les choix de conception (*design*). La question posée ne porte plus sur l'efficacité du P4P en général mais sur la meilleure façon de concevoir le P4P. En effet, si théoriquement on parle toujours de P4P au singulier, dans les faits, les conditions matérielles mises en place du P4P sont plurielles. Par ailleurs, selon la littérature, il faut non seulement étudier le meilleur type de P4P mais aussi ses interactions parfois problématiques avec le « contexte » (2.1.2.).

2.1.1 L'efficacité et ses différents critères : des résultats mitigés

D'après de larges pans de la littérature théorique *maintream* en économie de la santé, le P4P médical améliorerait la qualité des soins ainsi que l'allocation des ressources. C'est donc en priorité en vertu de ces deux critères qu'il convient d'évaluer le P4P.

L'efficacité du P4P sur la qualité des soins

Le plus instructif pour commencer l'analyse de l'impact du P4P sur la qualité des soins est de détailler, au moins une fois, la façon dont sont conduites ces expériences.

L'expérience d'Hilman et al. (1998) est à ce titre un bon point de départ. Les auteurs s'intéressent à la question du dépistage du cancer dans les réseaux de soins aux États-Unis. De nombreuses études ont montré le rôle moteur des médecins dans le dépistage, pourtant ceux-ci ne semblent pas adhérer aux recommandations de bonnes pratiques qui leurs sont adressées. Pour modifier cet état de fait, de nombreux outils sont envisageables – dont l'utilisation d'incitation financière de type P4P. Le P4P n'ayant pas fait la preuve de son efficacité, les auteurs vont chercher à comparer une politique d'incitation financière à une politique d'incitations non financières – toutes deux destinées à faire adhérer les médecins aux recommandations de bonnes pratiques.

L'étude, reposant sur la méthode d'essais cliniques randomisés (ECR), est menée entre 1993 et 1995 dans 52 réseaux de soins comprenant plus de 100 médecins à Philadelphie. Elle porte sur le dépistage du cancer du sein et du cancer du col de l'utérus chez les femmes âgées de 50 ans et plus. Les médecins – tous payés à la capitation – se voient incités financièrement par une hausse de leur revenu à la capitation de 20 % pour les meilleurs (bonus total), et de 10 % pour les suivants (bonus partiel) – les derniers 70 % de médecins ne

recevant rien. Il s'agit donc d'une performance relative aux résultats des autres médecins. Dans le groupe de contrôle, les médecins ne reçoivent que des informations sur leurs taux de dépistage (*feedbacks*) tous les 6 mois (sur une période totale de 18 mois). Ces informations sont également distribuées au groupe de médecins incités, si bien que la seule différence entre les deux groupes est la présence ou non du P4P. Contrairement aux prévisions théoriques, les résultats ne montrent pas de différence significative entre les deux groupes de médecins. Payer à la performance a le même effet qu'informer de la performance.

L'expérience précédente est intéressante mais relativement ancienne. On pourrait se demander si des expériences plus récentes ne donneraient pas d'autres résultats. Giuffrida et al. (2009) présentent des résultats similaires pour deux expériences du même type. Une première étude, reposant également sur la méthode d'ECR, analyse l'effet d'une hausse des honoraires dans le cadre du paiement à l'acte pour des médecins réussissant à améliorer leur score de vaccination antigrippale pour des patients de 65 ans et plus. 54 médecins généralistes participent à l'expérience pour plus de 20 000 patients éligibles à la vaccination du groupe test et près de 17 500 dans le groupe de contrôle. Une seconde étude impliquant 95 médecins généralistes porte sur la vaccination des enfants. Dans les deux cas, le protocole d'expérimentation est proche du précédent. De façon concordante avec l'étude précédente, les expériences montrent que l'introduction d'objectifs chiffrés de paiement permet d'accroitre les taux de vaccination mais il n'y a pas de différences significatives entre les groupes tests (avec rémunération) et les groupes de contrôle (sans rémunération).

Ces résultats sont spécifiques et ne prétendent pas à une portée générale. Cependant, ils partagent les mêmes conclusions que les différentes revues de la littérature sur le sujet. Par exemple, sur le programme britannique QOF, Peckham et Wallace (2010) prétendent que son impact sur la qualité des soins est fortement limité. Si l'on observe une amélioration des indicateurs, celle-ci ne peut pas être attribuée au P4P. Il est possible de trouver dans la littérature deux raisons majeures venant expliquer cette position concernant le QOF.

D'une part, la hausse de la qualité mesurée ne fait que poursuivre la tendance à l'amélioration observée les années précédant l'introduction du QOF en 2004. Le point de vue de Martin Roland – économiste britannique ayant eu un rôle de conseiller lors de l'introduction du QOF – permet de préciser le propos. Selon lui, la qualité des soins était déjà en amélioration avant le QOF et il est probable que le NHS aurait pu parvenir aux mêmes résultats en dépensant moins (Galvin, 2006).

D'autre part, il est statistiquement impossible d'attribuer au QOF l'amélioration de la qualité dans la mesure où il n'existe aucun groupe de contrôle. Bien que ce dispositif repose sur le volontariat, les sommes en jeu sont si importantes que tous les cabinets médicaux l'ont accepté. Il faut alors nuancer le propos : le QOF est incapable de faire la preuve de son efficacité – nous ne disposons pas de groupe de contrôle – mais, pour les même raisons, on ne peut pas dire de façon certaine qu'il est inefficace. L'hypothèse est toutefois crédible au regard de ce qui vient d'être dit sur l'amélioration continue de la qualité des soins en Grande-Bretagne, des expériences avec groupes de contrôle faites sur le P4P et de l'évolution du budget du NHS. En effet, on se rappelle qu'au début des années 2000 Tony Blair a décidé d'augmenter significativement le budget du NHS – notamment par l'intermédiaire du QOF. Il est alors plausible que l'amélioration de la qualité des soins ne vienne pas des vertus incitatives du QOF mais… de la hausse du budget santé au Royaume-Uni.

Ce qui vient d'être dit sur l'efficacité du P4P sur l'objectif de qualité des soins est généralisable par l'étude des différentes revues de la littérature existante.

Dans le domaine de la prévention, la revue de Town et al. (2005) cherche à évaluer l'impact des incitations financières sur les performances des médecins en termes de prévention à partir de la méthode d'ECR (essais cliniques randomisés). Sur les 6 articles finalement retenus, 4 s'intéressent aux campagnes de vaccination, 2 au dépistage du cancer et 2 portent sur ces deux thèmes. Une seule démontre un impact positif des incitations financières sur le score des médecins (Kouides et al., 1998). Par ailleurs, une étude intéressante montre que s'il n'y a pas de différence significative entre l'utilisation du P4P et celle des *feedbacks* (retour d'informations sur les taux de dépistage des médecins), ne rien faire génère de plus mauvais résultats (Fairbrother et al. 1999). L'amélioration de la performance serait alors due à l'amélioration de l'information sur la pratique et non pas au fait que les médecins soient plus motivés. L'interprétation des résultats mérite d'être mentionnée. En effet, de façon assez étrange les auteurs concluent que ce qui est en cause n'est pas le fait de savoir si les incitations influent sur le comportement. Selon eux, c'est nécessairement le cas. La seule question pertinente ne serait donc pas de savoir s'il faut mettre en place des dispositifs de P4P mais combien il faut payer pour obtenir les résultats théoriques (Town et al. 2005, p. 237-238).

La revue de la littérature de Sabatino et al. (2008) parvient au même type de conclusion sur la question du dépistage du cancer du col de l'utérus, du cancer du sein et du cancer du côlon. Les expérimentations étudiées portent toute sur une comparaison P4P contre *feedback*.

Encore une fois, si les études portant sur le P4P parviennent le plus souvent – mais pas toujours – à démontrer une hausse des scores de prévention, cette hausse n'est jamais significativement différente des groupes avec *feedbacks*. Si l'on s'intéresse à l'effet du P4P sur le suivi des maladies chroniques (de Bruin et al., 2011), les résultats sont similaires.

De façon plus générale les différentes revues de littérature sur le sujet concluent toutes… qu'il n'est pas possible de conclure. Les difficultés méthodologiques à isoler l'effet du P4P sont présentées comme l'obstacle majeur à l'évaluation « toutes choses égales par ailleurs » de ce dispositif. La seule étude allant légèrement à l'encontre de ces conclusions est celle de Van Herck et al. (2010) pour qui le P4P permet en moyenne d'augmenter la performance des médecins de 5 %. C'est à notre connaissance la seule à être optimiste dans ses interprétations de l'effet du P4P sur la qualité des soins… ce qui encourage à la méfiance. D'ailleurs, ces auteurs concluent – comme tous les autres – sur la précarité méthodologique des résultats ainsi que sur la nécessité de poursuivre les recherches. Au final, il semble raisonnable de retenir que le P4P est incapable de faire la preuve de son efficacité sur l'amélioration de la qualité de soins.

L'efficacité du P4P sur la réduction des dépenses de santé

Assez paradoxalement, les études portant sur l'efficacité du P4P à réduire les dépenses de santé sont beaucoup plus rares dans la littérature empirique. Néanmoins, il existe une dizaine de revues de la littérature sur le sujet qui concluent toutes sur l'incapacité du P4P à démontrer son efficacité sur ce critère. C'est le cas par exemple des travaux d'Emmert et al. (2012) qui cherchent à identifier le bénéfice coût-avantage du P4P à partir de trois expériences sont analysées.

Dans le même registre Nahra et al. (2006) ont procédé à une analyse coût-utilité pour évaluer l'efficacité du P4P sur les soins cardiaques hospitaliers. Les primes, pouvant s'élever jusqu'à 2 % de la rémunération habituelle, étaient liées à l'amélioration de trois indicateurs de qualité sur une période de 4 ans. Au final, les coûts par Qaly[61] s'élevait à des niveaux compris entre 12 000 et 30 000 $ – soit à des niveaux bien en-dessous de ce qui est recommandé par la littérature et appliqué dans les divers systèmes de santé. Si ce test paraît couronné de succès, il présente des limites méthodologiques importantes. En effet, les résultats présentés se fondent sur une comparaison entre l'effet du P4P la première année et son effet la 4ème année alors

[61] L'évaluation médico-économique vise à mettre en rapport un critère de résultat ou gain en santé d'une stratégie de santé avec le coût d'une telle stratégie. Le critère des qaly (*Quality Adjusted Life Year* » ; ou « année de vie pondérée par la qualité de vie) est un critère de résultat souvent utilisé

qu'il aurait fallu comparer l'effet du P4P à la 4ème année par rapport à une situation antérieure sans P4P. Par ailleurs, cette étude – comme la plupart des études sur la question – ne comporte pas de groupe de contrôle. Or, il est possible que, comme pour le cas de l'amélioration de la qualité de soins, l'usage uniquement de *feedbak* ait le même type de résultats que l'usage de P4P – mais avec un meilleur bilan coût-efficacité puisqu'il n'aurait pas de prime à payer !

D'autres études sur le même sujet parviennent à des résultats semblables. Une étude de Curtin et al. (2006) trouve que pour chaque dollar dépensé, 2.5 $ sont économisés. La revue de littérature de de Bruin et al. (2011) mentionnent deux études qui rapportent un gain de 70.95 $ par dépistage pour l'une et 46.40 $ par dépistage pour l'autre. Nous pourrions multiplier le nombre d'études présentant un effet positif du P4P sur l'allocation des ressources. Ces résultats ont d'ailleurs conduit Van Herck et al. (2010) à conclure sur le fait que le P4P est efficace du point de vue de l'économie des ressources. Mais, comme le soulignent Eijkennar et al. (2013), cette conclusion est hâtive dans la mesure où la méthodologie utilisée dans ce type de littérature ne cherche jamais à comparer la situation avec P4P à une situation similaire sans P4P. Il est en conséquence difficile de conclure.

L'efficacité du P4P sur d'autres critères

En supplément des études portant sur la qualité des soins et les économies réalisées, la littérature évaluant le P4P médicale s'intéresse à d'autres critères dont il est question ici. Le premier d'entre eux est l'égalité d'accès aux soins. L'une des justifications du gouvernement britannique et de la *British medical association* pour soutenir le QOF a été de dire que celui-ci permettrait de réduire les inégalités d'accès aux soins. Mais, il est également probable qu'il les aggrave si l'on considère par exemple l'étude américaine de Krave et al. (2008). Selon cette étude menée dans près de 3 500 hôpitaux du pays, les hôpitaux avec plus de 20 % d'afro-américains ont une probabilité plus faible de voir les scores de performance s'améliorer pour le traitement du myocarde et de la pneumonie (comparativement à des hôpitaux à minorités plus faibles).

La revue de littérature d'Alshamsan et al. (2010) a pour objet d'étudier cette question (sur les 22 études retenues, 20 portent sur le QOF). Du point de vue des inégalités socioéconomiques, une seule étude montre que la pauvreté est positivement corrélée à de meilleurs scores durant la première année du QOF. Les autres études ont montré une plus faible qualité des soins dans les zones pauvres que dans les zones plus riches – cette

différence étant corrigée au fil du temps. Par exemple, Ashworth et al. (2007b) trouvent une différence entre les zones les plus pauvres et les plus riches de 64.5 points de QOF la première année (soit 6.1 % de différence) – cette différence passant à 30.4 la deuxième année (soit 2.9 % de différence). McGovern et al. (2008) montrent que les patients avec une maladie coronarienne vivant dans des zones pauvres ont une probabilité plus faible d'avoir leur statut d'état asthmatique enregistré, leur pression sanguine enregistrée, ou de recevoir une thérapie de bétabloquants.

Concernant les inégalités de sexe et d'âge, les auteurs trouvent que les inégalités existantes persistent – le QOF n'ayant pas l'effet correctif annoncé. Par exemple, pour une même maladie, une femme a une probabilité inférieure à celle d'un homme d'avoir un contrôle de pression sanguine après l'introduction du QOF (Simpson et al., 2006). Du point de vue des inégalités ethniques, Ashworth et al. (2007a) montrent que lors de la première année du QOF britannique les zones avec de fortes populations afro-caribéennes et asiatiques présentaient moins de prescriptions de statine. Au sein de la capitale Londonienne, les noirs et les asiatiques ont une probabilité plus faible d'atteindre les résultats intermédiaires concernant le suivi du diabète (pression sanguine, taux HbA1c[62] et cholestérol) que le reste de la population. Lorsque le QOF présente des résultats positifs, ceux-ci sont plus forts pour les blancs britanniques que pour les noirs caribéens, les indiens, les pakistanais et les bangladais.

Au final, il est tentant de dire que le QOF a accru les inégalités. Ce serait une erreur. Encore une fois, pour des raisons méthodologiques évidentes, rien ne permet de dire que l'évolution des inégalités après l'introduction du QOF lui est imputable. Ce qui est certain par contre, c'est que QOF ou non, les inégalités d'accès aux soins restent fortes et se développent. La possibilité d'exclure certains types de patients ainsi que les inégalités sociales d'éducation à la santé invitent néanmoins à ne pas exclure la possibilité d'un impact du QOF sur la hausse des inégalités.

Pour finir, nous pouvons citer deux derniers critères d'évaluation du P4P : l'effet sur la coordination et la continuité des soins et l'effet sur la prise en compte du patient dans sa prise en charge. Dans la revue de Van Herck et al. (2010), le P4P est présenté comme permettant d'améliorer ces deux critères. Une étude citée par les auteurs montre ainsi que rémunérer à la performance la coordination peut avoir un impact positif. En ce qui concerne la prise en compte du patient, deux études trouvent des résultats contradictoires sur la satisfaction des

[62] L'HbA1c est un examen sanguin qui mesure le risque de complications du diabète à long terme.

patients. Si ces deux critères sont intéressants, trop peu d'études ont été réalisées pour conclure.

Maintenant que nous avons analysé l'efficacité – peu probante – du P4P au singulier, il est temps de nous intéresser au P4P au pluriel.

2.1.2. Conception et contexte : expliquer les mauvais résultats du P4P médicale

Si le P4P « au singulier » est incapable de faire la preuve de son efficacité, la littérature empirique cherche depuis quelques années à évaluer le P4P « au pluriel ». Le P4P au singulier peut se définir comme un principe général relativement abstrait. Or, lorsqu'il s'agit de mettre en place concrètement le P4P, on s'aperçoit que sur ce principe général peuvent s'adosser diverses méthodes de conception. Par exemple, faut-il payer une somme fixe quelle que soit la performance ou un montant variable du niveau de la performance ? Faut-il payer l'individu particulier (le médecin) ou le groupe dans lequel il s'insère (le cabinet) ? L'objectif de cette littérature est finalement assez simple : parmi les diverses modalités possibles du P4P, il s'agit de trouver celle qui engendre une plus grande efficacité. Mais là encore les résultats sont bien maigres et la littérature s'attarde sur les effets de contexte.

Choix de conception du P4P médicales

Le premier critère important pour la construction du P4P médicale porte sur le choix des indicateurs de qualité. Il est possible de distinguer des indicateurs de processus – on mesure l'amélioration de pratiques dans le temps – et des indicateurs de résultats on mesure uniquement le résultat final, indépendamment de l'ampleur de l'amélioration des pratiques. L'intérêt majeur des indicateurs de processus est qu'ils permettent de valoriser l'engagement de ceux qui sont considérés comme les plus mauvais médecins quand ils ne respectent pas les règles de bonne pratiques Alors qu'un objectif trop difficile peut décourager un médecin dont les résultats initiaux sont particulièrement faibles, le fait de rémunérer le médecin lorsqu'il parvient à un niveau minimal de progression favorise l'effort. Par ailleurs, les médecins ayant déjà des niveaux de performance conformes à l'objectif cible ne sont pas incités à améliorer leur activité. Pour ces raisons, en moyenne, on observe que les indicateurs de processus engendrent une meilleure efficacité que les indicateurs de résultats – les indicateurs intermédiaires (demi-processus, demi-résultat) obtenant des résultats… intermédiaires.

L'étude de Rosenthal et al. (2005) permet d'illustrer le propos. Dans le cadre d'une « expérience naturelle » en Californie les auteurs cherchent à comparer deux conceptions du

P4P – l'une rémunérant l'atteinte d'un objectif prédéterminé (méthode majoritairement étudiée par la littérature), l'autre rémunérant l'amélioration des pratiques. De façon attendue, les indicateurs de processus parviennent à de meilleurs résultats que les indicateurs de résultats. Ainsi, dans le cas d'indicateurs de résultats, les meilleurs sont payés sans pour autant améliorer leur performance et les plus faibles ne tentent pas leur chance. La revue de littérature de Van Herck et al. (2010), parvient aux mêmes conclusions : il est plus efficace de rémunérer la progression.

Un deuxième critère de choix de conception porte sur les instruments de mesure de la qualité. Faut-il privilégier les dossiers papier, les dossiers électroniques, les données collectées par les médecins eux-mêmes, des données collectées par les pairs ? La littérature s'intéresse assez peu à la question et lorsqu'elle le fait aucune différence n'est soulignée entre les différentes méthodes de récolte des informations (Eijkenaar, 2013).

Une question de conception beaucoup plus documentée porte sur la forme que doit prendre l'incitation financière :

- Le signe de l'incitation : l'incitation doit-elle être positive – c'est une récompense du bon comportement – ou négative – c'est une sanction du mauvais comportement ? Il semble que les mécanismes de P4P reposant sur une sanction financière présentent des résultats moins bons que les autres. Si les dispositifs budgétairement neutres de ce type – générant des gagnants et des perdants – seraient moins efficaces, la relation n'est pas simple et dépend d'autres paramètres tels que la taille de l'incitation.

- L'ampleur de l'incitation : l'incitation doit-elle représenter une forte part de la rémunération totale ou peut-elle se contenter d'être une rémunération d'appoint ? Encore une fois, il n'y a pas de relation bien établie entre la taille de l'incitation et efficacité. de Brantes et d'Andrea (2009) trouvent une forte relation d'après laquelle le niveau de la récompense expliquerait entre 89 et 95 % de la variation de la performance. Mais cette démonstration est la seule à trouver de tels résultats si bien qu'il est difficile de conclure.

- La continuité de l'incitation : l'incitation doit-elle être une fonction continue et croissante de la performance ou doit-elle dépendre d'un seuil ? Là encore, les études divergent et selon les expériences les deux méthodes peuvent obtenir des résultats positifs.

- La cible de l'incitation : qui doit percevoir l'incitation – le médecin ou le groupe de médecin (en cas d'exercice en cabinet de groupe) ? La littérature définit l'exercice de groupe comme une situation où au moins deux médecins exercent en partenariat avec un centre de

profit commun, des revenus communs, des dépenses communes, une distribution des profits au groupe (Town et al. 2004). À ce titre, deux médecins partageant leurs dépenses mais avec des revenus individuels ne peuvent être considérés comme exerçant en groupe. Dans ce cadre, la littérature distingue deux cas de figure. Si la performance nécessite un effort collectif, il est préférable d'inciter les groupes. Pour la performance qui est sous le contrôle individuel du médecin, l'incitation doit porter sur l'individu. Cependant, dans la mesure où l'exercice de la médecine est de plus en plus collectif, il devient difficile d'attribuer des performances aux seules individualités. Les incitations devraient alors tendre à rémunérer les groupes (Eijkenaar, 2013). Une raison supplémentaire plaidant pour l'incitation de groupe porte sur le *case-mix*. L'exercice individuel et l'exercice de petit groupe ne permettant pas d'avoir un échantillon de patients suffisamment grand pour limiter les risques, le P4P peut s'avérer problématique. En effet, ce type d'exercice est vulnérable à des particularités locales qui ne permettent pas de mettre tous les médecins sur un pied d'égalité lorsqu'il s'agit de juger de leur performance. À l'inverse, lorsque l'échantillon de la population est suffisamment grand la performance des médecins n'est plus susceptible d'être impactée par la spécificité de la patientèle (en termes de revenu, d'état de santé, d'environnement, etc.). Evidemment, le point négatif dans cette évolution porte sur la hausse prévisible des comportements de passager clandestin (*free-riding*). Malgré ces discussions théoriques sur les différentes cibles possibles pour l'incitation, la revue de littérature de Damberg et al. (2009) ne permet pas de conclure sur la supériorité de l'une des trois possibilités étudiées (individuelle, groupe, hôpital).

Un dernier choix de conception traité dans la littérature porte sur la façon dont est introduit et présenté le dispositif de P4P. L'introduction peut se faire sous la forme de volontariat ou d'obligation. Si le volontariat présente le risque de sélectionner les médecins déjà performants et/ou déjà motivés (qui ne sont pas nécessairement les mêmes), les résultats sont contradictoires : Coleman et al. (2007) ne trouvent pas de biais de sélection alors que Ryan (2009) en trouve un. Outre la question du volontariat, certaines études insistent sur l'importance de la communication faite autour du P4P. S'il n'existe pas d'études statistiques sur la question, certains auteurs émettent l'hypothèse que l'absence de résultats peut provenir d'une mauvaise communication (Van Herck, 2010). Enfin, il est intéressant de noter que plusieurs études trouvent que les résultats sont meilleurs lorsque les dispositifs de P4P sont financés par de nouveaux budgets plutôt que par des transferts budgétaires (Lindenauer et al. 2007 et Ryan, 2009, par exemple). Ce point peut s'expliquer selon nous par la même

hypothèse que pour le QOF : l'amélioration des résultats peut être causée… par la hausse du budget de la santé.

Il ressort finalement peu de choses des études sur les choix de conception du P4P. De nombreux auteurs expliquent alors la variabilité des résultats par la variabilité des contextes.

L'effet du « contexte » sur le P4P médicale

Les revues de littérature portant sur l'efficacité du P4P mentionnent presque toutes l'existence d'un effet de « contexte » permettant d'expliquer en creux les résultats limités obtenus. La revue de littérature de van Herck et al. (2010) est la plus complète à ce sujet en distinguant quatre éléments faisant partie du contexte et pouvant avoir un impact sur l'efficacité du P4P médical. Le premier des éléments de contexte mentionné n'est rien de moins que les caractéristiques du système de santé. En ce sens, le P4P n'est qu'un élément faisant partie d'un système de santé plus vaste. Or, les systèmes de santé sont très différents les uns des autres. Cette diversité peut être une source d'explication pertinente pour expliquer les difficultés du P4P à faire la preuve de son efficacité. Ainsi, mettre en place un programme de P4P au niveau national produirait de meilleurs résultats que lorsque le système de santé est fragmenté et ne permet pas de mener une politique publique uniforme. Les meilleurs résultats du QOF britannique seraient alors à mettre au crédit de l'unité du NHS alors que les mécanismes américains pâtiraient de leur diversité. En influençant les choix de conception du P4P de façon plus autoritaire, un niveau de décision plus élevé et plus centralisé permettrait de trancher dans le sens de l'efficacité.

Un deuxième élément de contexte – assez lié au précédent – porte sur les modes de rémunération principaux des médecins. La théorie économique *mainstream* établit que le paiement à la capitation engendre une sous-utilisation des ressources alors que le paiement à l'acte engendre une sur-utilisation des ressources[63]. En conséquence, il devrait être important de concevoir le paiement à la performance de façons différentes selon qu'il faut encourager à utiliser plus de ressources ou à réduire l'usage des ressources. Or, l'indifférenciation de la conception du P4P au regard du mode de rémunération principal du médecin peut être à l'origine des difficultés à faire la preuve de son efficacité.

[63] Cf. Chapitre 1. La théorie prévoit par exemple qu'avec le paiement à l'acte les médecins sont incités à prescrire plus facilement dans la mesure où ils doivent satisfaire le patient pour qu'il revienne. Dans le cas de la capitation, c'est l'inverse : les patients étant attachés à leur médecin et ce dernier ayant intérêt à minimiser son effort, il n'est pas intéressant de prescrire – cela est d'autant plus vrai si le médecin gère une enveloppe globale.

Ensuite, il est possible de mettre en évidence la variabilité contextuelle de l'organisation (*leadership*, culture, histoire d'engagement dans la qualité, propriété des moyens de production, taille de l'organisation) et des caractéristiques individuelles des médecins (démographique, expérience, formation, etc.). Si nous avons discuté de certains de ces aspects, ils n'ont pas tous fait l'objet d'études spécifiques et lorsque cela a été le cas, les études sont trop peu nombreuses et souvent défaillantes méthodologiquement.

Enfin, le dernier élément de contexte qui revient dans la littérature sur l'efficacité du P4P porte sur les caractéristiques des patients. Aucune information ou presque n'est disponible sur la patientèle alors qu'il est de notoriété publique que le comportement du patient peut avoir une influence sur la performance du médecin (style de vie, coopération, etc.). À l'avenir la conception des mécanismes de P4P devrait prendre en compte l'hétérogénéité des patients.

2.2. Les effets pervers du paiement à la performance médicale

L'évaluation du P4P médicale ne permet pas de tirer de conclusion sur son (in)efficacité en dépit du grand nombre d'expériences mises en place. La littérature empirique a cependant fait la preuve que ce type de dispositif pouvait avoir des effets pervers. Les effets pervers sont des effets inattendus au regard de la théorie, ayant pour particularité d'avoir un impact négatif sur le système de santé. Il est important de noter que ces effets inattendus – positifs et négatifs – sont par construction sous-évalués. En effet, sachant que les études empiriques ne sont pas pensées pour les mesurer, on ne peut que les sous-évaluer. Il est alors évident que la plupart du temps la qualité de la méthodologie est affectée par la nature même de ce dont on parle.

Si l'on peut s'attendre à ce que des effets inattendus positifs se réalisent, la littérature n'en fait pas ou presque pas mention – c'est pour cela que nous nous focalisons sur les effets pervers que nous rangeons en 4 catégories : la sélection des patients (2.2.1), la sélection des tâches (2.2.2), le crowding out effect médical (2.2.3) et les effets de « gaming » où les acteurs peuvent se jouer de l'indicateur (2.2.4). Au titre des effets inattendus positifs, on peut néanmoins citer deux études qui décèlent un effet de débordement (*spill over*) positif du P4P sur la continuité des soins (Cameron et al., 1999, Srirangalingam et al., 2006). Le fait d'inciter à mieux prendre en charge certains patients aurait pour effet de généraliser la bonne pratique au-delà du seul groupe des pathologies pour lesquelles les médecins sont sur-rémunérés.

2.2.1. La sélection des patients

L'un des effets pervers du P4P les plus documentés porte sur la possible sélection des patients qu'engendrerait ce type de dispositif.. Si le médecin sait que, pour une raison ou une autre, certains types de patients sont susceptibles de détériorer son score de performance, alors il est profitable pour lui de les exclure du calcul de l'indicateur. La sélection des patients n'est rien de moins qu'un processus de sélection des risques. Il est alors tentant pour un médecin, selon la conception de l'indicateur, d'évincer du calcul les plus malades, les moins conciliants, les moins éduqués à la santé, etc. – toutes ces catégories étant alors potentiellement corrélées à des inégalités sociales. Longtemps l'absence de preuve de sélection des patients a permis de dire qu'il s'agissait d'un risque théorique. Désormais l'accumulation de preuve permet à Eijkenaar de dire que la « risk selection is not just a theoretical concept » (Eijkenaar, 2013, p. 119).

Par exemple, Shen (2003) s'intéresse au phénomène de sélection des risques en médecine ambulatoire lorsqu'un contrat de P4P est proposé aux médecins[64]. Son étude porte sur des patients couverts par les programmes *Medicaid* et *Main office substance abuse* (OSA). Les patients sont catégorisés en deux groupes selon la sévérité de la maladie : sévérité normale – les patients donnant alors lieu à une rémunération à la performance – et sévérité élevée – les patients étant exclus du calcul de l'indicateur. L'étude compare la probabilité pour un patient d'être catégorisé dans le groupe de sévérité élevée avant et après la mise en place du P4P pour les patients couverts par l'OSA (les patients couverts par *Medicaid* étant le groupe de contrôle). L'auteur trouve alors que la probabilité d'être classé dans le groupe de maladie plus sévère croît suite à l'introduction d'une incitation de plus de 7 % ($p<0.01$). Cela suggère qu'à la suite de l'introduction du P4P, les médecins sont incités à choisir les patients sur la base desquels leur performance est calculée.

Le phénomène de sélection des patients s'observe également dans le QOF britannique (Doran et al., 2006, Doran et al. 2008). On se rappelle en effet que le QOF prévoit une procédure d'exclusion des patients nommée *exception reporting*. Au départ, l'idée est de protéger les patients contre un traitement inapproprié d'un médecin qui souhaiterait maximiser sa rémunération à la performance. La logique peut également être d'exclure des patients pour lesquels les mesures quantifiées de qualité des soins ne sont pas adaptées à la spécificité du cas. Cependant, cette procédure peut également permettre au médecin d'exclure

[64] On pourra se référer à Ryan (2009) pour une étude dans le cas de l'hôpital.

des patients pour lesquels il anticipe qu'il sera difficile d'atteindre les objectifs. L'amélioration de performance peut alors être qualifiée d'artificielle.

Afin de démontrer cet effet de sélection des patients, Doran et al. (2006) ont distingué et calculé deux types de performance lors de la première année du QOF : la performance reportée – qui sert de base au calcul de la performance – et la performance totale – qui inclut les exclus. Si l'on suppose que, pour une maladie donnée, 100 personnes sont atteintes de la maladie en question, 80 sont choisies pour le calcul de la performance et 75 obtiennent le score attendu, la performance reportée est de 94 % (75/80) et la performance totale est de 75 % (75/100). Les auteurs ont calculé que la médiane de la performance reportée la première année approchait les 83 %. Le pourcentage médian de sélection des patients était de 6 % la première année avec une variabilité importante puisque pour le 1 % des cabinets ayant le plus exclu de patients, en ont exclu en moyenne plus de 15 %. Si le pourcentage médian d'exclusion des patients ne semble pas nécessairement élevé, il est pourtant le facteur explicatif le plus important de la performance médicale : 1% d'exclusion supplémentaire apportait ainsi 0.31 % de performance supplémentaire – associée à une rémunération supplémentaire.

2.2.2. La sélection des tâches

Dans le cadre de la sélection des risques, en plus du phénomène de sélection des patients, la littérature en économie du travail est familière depuis longtemps avec la question de la pluralité des objectifs que doit satisfaire un agent. Le modèle de « multitâche » (*multitask*) présenté par Holmström et Milgrom (1991) est à ce titre exemplaire : lorsqu'un mécanisme de P4P ne rémunère qu'une seule composante de l'activité, il est rationnel pour les employés de ne pas exercer d'effort sur les dimensions du travail non incitées. Par exemple, l'étude de Campbell et al. (2009) montre que le P4P aurait un effet négatif sur certaines dimensions non incitées de la qualité des soins (asthme et maladies coronariennes, continuité des soins). Une autre expérimentation suggère par ailleurs dans le cas de l'hôpital que l'introduction du P4P aurait eu pour effet de réduire le temps consacré par les médecins à l'enseignement et à la recherche (Simon et al., 1999).

L'article de Roski et al. (2003) sur le traitement du tabagisme conduit à des résultats similaires. La littérature en santé publique a mis en évidence l'intérêt d'identifier les fumeurs et de les conseiller sur les méthodes permettant d'arrêter la cigarette notamment par la distribution de documents informatifs. Dans cette étude, l'hypothèse de P4P comme élément

moteur stimulant les médecins à adopter les bonnes pratiques est testée. Si, de manière habituelle, il n'est pas possible de démontrer l'efficacité du P4P face aux différents groupes de contrôle mis en place, l'originalité de cette étude est de démontrer un effet de sélection des tâches. Deux objectifs étaient liés au P4P. D'une part, les médecins devaient identifier et enregistrer dans un dossier médical informatisé au moins 75 % des patients âgés de plus de 18 ans. D'autre part, 65 % des fumeurs devaient avoir reçu des conseils (notamment sous forme de documentation) pour arrêter de fumer (également enregistré dans le dossier médical). En plus de ces deux critères, un troisième critère quantitatif non incité a été calculé : le taux de patients déclarant avoir reçu une assistance proactive (conversation téléphonique, relance du médecin, etc.) pour arrêter de fumer. De façon assez surprenante, alors que ce taux était de 14.3 % dans le groupe de contrôle, il descendait à 8.8 % dans le groupe soumis au paiement à la performance – les médecins faisant plus d'effort sur l'identification des fumeurs (6.2 % contre 14.1 % et la distribution de documentation (18.3 % contre 24.2 %). L'amélioration des pratiques incitées s'est faite au détriment des pratiques non incitées – sans toutefois produire de modification significative du point de vue de l'*outcome* pour les patients.

2.2.3. Le crowding out effect médical

La littérature grandissante en économie sur les motivations s'appuie notamment sur la distinction entre motivation intrinsèque (MI) et motivation extrinsèque (ME) issue de la psychologie sociale (Deci et Ryan, 2002)[65]. Dans ce cadre théorique, le P4P est l'exemple typique de la ME car la tâche est rendue attrayante par l'intéressement financier. Seule la récompense est valorisée et non l'objectif lui-même. Il faut prescrire des médicaments génériques non pas parce que c'est valorisant en soi mais parce que c'est profitable. Ou bien, il faut entreprendre des actions de dépistage et de prévention parce que c'est intéressant financièrement et non parce que cela rentre dans la définition d'une bonne pratique professionnelle ou dans un souci de santé publique.

L'existence de plusieurs registres de motivation pose néanmoins un nouveau problème à l'économiste. En effet, l'hypothèse de substituabilité des motivations semble contraire au cadre économique standard où les motivations sont complémentaires, comme nous l'avons vu au chapitre 3. On dit qu'il y a *crowding out effect* (COE) lorsque l'introduction d'une ME, alors que l'action visée par l'incitation était auparavant intrinsèquement motivée, détruit les

[65] Cf. Chapitre 3.

MI. La ME a détruit la MI et le P4P pourrait avoir pour effet de démotiver les médecins en évinçant les MI. L'incitation peut « désinciter ». Il s'agit alors pour comprendre l'effet du P4P de déterminer sous quelles conditions les motivations sont complémentaires (*crowding in*) ou substituables (*crowding out*). Si un nombre croissant d'éléments de preuve vient alimenter l'hypothèse de COE, celles-ci sont, dans un souci de cohérence, présentées dans la section 4.3. Il suffit ici de noter que le COE est un effet pervers additionnel venant relativiser l'évaluation du P4P médical.

2.2.4. Gaming et effets additionnels

Lorsque la littérature s'intéresse aux effets pervers du P4P médicale, elle insiste sur la notion de *gaming*. Dans ce cadre, les auteurs reconnaissent que la mise en place d'une rémunération sur indicateur implique nécessairement que les acteurs vont « jouer » l'indicateur – c'est l'effet recherché au moins en partie – et se jouer de l'indicateur. L'amélioration de la performance mesurée passe ainsi par l'accroissement de l'effort sur la tâche incitée éventuellement au détriment des autres (jouer l'indicateur) et par des stratégies visant à améliorer la performance mesurée en exploitant les failles de la mesure (se jouer de l'indicateur). Le premier cas recouvre l'effet de sélection des tâches tandis que le second recouvre toute une série de stratégies dont fait partie la sélection des patients, l'effet salami et le surcodage. Dans ce dernier cas, les médecins peuvent avoir un comportement de fraude – parce qu'ils codent mal leur performance – ou d'optimisation – parce qu'ils usent de l'incomplétude de la règle pour compter (ou ne pas compter) des éléments discutables de la performance. Il est également possible qu'il se trompe dans leur travail de codage.

Ainsi, en plus de la sélection des tâches, un autre effet pervers de la rémunération sur indicateur porte sur le codage de la performance. Cette idée peut être illustrée à partir de l'étude de Norton (1992) s'intéressant à l'effet du P4P sur le profil des patients dans les maisons de retraite. L'auteur observe qu'après mise en place du P4P, un nombre important de personnes étant qualifiées de très fortement handicapées recouvraient « miraculeusement » leurs capacités après leur entrée dans l'institution. Malheureusement aucun miracle n'a eu lieu : en raison du mode de rémunération, les praticiens étaient incités à déclarer des états de morbidité plus importants afin d'être ensuite plus confortablement rémunérés.

Cet effet de codage est bien documenté en France comme ailleurs, au-delà des mécanismes de P4P avec la tarification à l'activité dans le cadre de l'hôpital (Belorgey, 2010, Domin, 2013a et 2013b, Mas et al., 2011). Ce nouveau mode de rémunération des hôpitaux

est centré sur l'usage d'une nomenclature d'actes chacun donnant lieu à un niveau de rémunération différent. Dans ce cadre, on a observé pour certaines pathologies des phénomènes de surcodage : plutôt que de coder les actes effectivement réalisés, les médecins ont tendance à coder des actes similaires mais mieux rémunérés. Un exemple célèbre est celui de l'accouchement : un accouchement avec césarienne étant mieux rémunéré qu'un accouchement par voie basse, il a été observé une hausse anormale du premier type d'actes (Milcent et Rochut, 2009).

Cette incitation à manipuler l'indicateur plutôt qu'à améliorer la qualité peut prendre d'autres formes. L'une d'entre elles est baptisée par Beauvallet (2010) l'effet « salami ». L'effet salami énonce que la mise en place d'un indicateur de performance est toujours suivi de l'amélioration de l'indicateur sans pour autant entraîner une amélioration de la performance. Des individus soumis à un nouvel indicateur vont chercher à jouer l'indicateur plutôt que la performance. Cet effet tire son nom du constat fait dans une boucherie spécialisée dans la confection de salami en tranche. Afin d'améliorer sa production le manager a introduit une rémunération à la performance fondée sur le nombre de tranches découpées. Alors que l'objectif initial de ce manager était d'augmenter sa production en volume, le résultat observé a été la multiplication du nombre de tranches. Les salariés ont découpé des tranches plus fines mais pour un volume constant. Le manager a ainsi dû payer plus cher des salariés performants sans toutefois augmenter son chiffre d'affaire.

Il est alors possible de mettre en évidence un décalage entre la notion de performance telle qu'elle est mesurée par les indicateurs (performance mesurée) et la notion de performance telle qu'elle est imaginée et souhaitée par le régulateur (performance idéalisée). On peut interpréter l'effet salami comme une mauvaise traduction de l'objectif (performance idéale) par l'indicateur (performance mesurée) ayant pour résultat une hausse de la performance mesurée sans effets positifs sur la performance idéale.

Deux derniers effets suggérés mais non étudiés explicitement par la littérature peuvent être mentionnés. D'une part, certains auteurs pensent que l'usage de recommandations de bonnes pratiques peut être problématique dans la mesure où selon les *guidelines*, des actions différentes sont recommandées pour des situations similaires (Kravitz et Greenfield, 1995 ; Rodwin, 1997 ; Van Horn et al. 1997). D'autre part, d'autres auteurs soulignent la possibilité que le P4P provoque un conflit d'intérêts entre patient et médecin (Allen, 1993, Thirlby et al., 1996, Chaix-Couturier et al., 2000). En effet, il n'est pas difficile d'imaginer une situation où, quelle que soit la raison, le patient ne veuille pas être traité à partir des référentiels communs

alors que le médecin y a intérêt. C'est d'ailleurs l'un des intérêts de la procédure d'exclusion du QOF que d'éviter ce type de problème… en engendrant d'autres.

2.3. Une synthèse des évaluations du P4P médicale

L'évaluation des politiques de P4P peut se faire à partir de trois critères : efficacité, efficience et effets pervers (2.3.1) pour déboucher sur une analyse de ses limites méthodologiques (2.3.2).

2.3.1. Le P4P médicale : efficacité limitée, efficience douteuse et effets pervers

L'efficacité d'une politique économique se définit comme l'atteinte des objectifs fixés. Le point d'accord majeur entre les différentes revues de la littérature sur le sujet porte sur l'incapacité actuelle à prouver un effet strictement positif (ou négatif) du P4P tant sur la dimension qualitative qu'économique. Plus ou moins optimistes sur le développement des mécanismes de P4P, elles ne peuvent cependant prétendre donner une réponse claire quant à l'efficacité de ce dispositif (Tableau 4.8.). La qualité des données et celle des méthodes statistiques sont un rempart difficilement surmontable pour obtenir des résultats dénués d'ambiguïté.

Tableau 4.8. L'évaluation du P4P médicale

Evaluation du P4P médicale	
Critères d'évaluation	***Résultats***
Efficacité clinique	Etudes à « fort niveau de preuve » : forte littérature sur les soins préventifs (notamment vaccination et dépistage) mais peu ou pas d'impact du P4P. Etudes à « faible niveau de preuve » : beaucoup d'études montrent une amélioration dans quelques indicateurs qualitatifs. Mais, il est impossible d'attribuer les résultats au P4P dans la mesure il y a une tendance d'amélioration continue et qu'il n'y a pas de groupe de contrôle.
Efficacité économique	Evaluation économique complète : ce type d'évaluation établit une liaison explicite entre les coûts du programme et la qualité par le calcul de ratios coût-efficacité. Toutes les études trouvent un bilan coût/bénéfice positif. Mais, il n'y a jamais de groupe de contrôle. Evaluation économique partielle : ce type d'étude n'établit pas de liaison explicite entre coût et qualité, seules des déductions sont faites. Les bilans sont positifs mais la méthodologie n'est pas satisfaisante. Au total : le P4P présente un potentiel d'efficacité économique mais aucune preuve rigoureuse ne vient confirmer cette hypothèse.
Lutte contre les inégalités	Inégalités socioéconomiques : différence de qualité des soins entre zones pauvres et riches mais comblée à moyen terme. La tenue du dossier médical informatisé est plus souvent réalisée dans les zones riches (ex : tabagisme). Inégalités de sexe, d'âge et ethnique : persistance des inégalités malgré l'introduction du P4P. Les femmes, les personnes âgées et celles issues des minorités ethniques sont le plus souvent moins bien soignées. Au total : Il semble y avoir une neutralité du P4P (pas d'effet).
Incitations non financières	Feedback : plusieurs études semblent montrer qu'il n'y a pas de supériorité du duo P4P + feedback *versus* feedback seul. Public reporting : la publicité des résultats peut être plus efficace avec le P4P. Mais on ne peut pas distinguer ce qui vient du P4P ou de la publicité. Les auteurs suggèrent que ce soit la publicité qui soit la plus efficace.
Choix de conception (design)	Type de mesure de la performance : il vaut mieux des mesures simples et spécifiques que complexes ; des mesure de processus plutôt que de résultats ; des niveaux de base faibles laissant une marge de progression forte. Entité ciblée : individus ou petites équipes plutôt que grands groupes. Arbitrage entre *free-riding* et taille du *case-mix*. Indicateurs : il vaut mieux des cibles absolues que relatives et plusieurs cibles qu'une seule. Montant de l'incitation : très peu d'information disponible (pas de lien clair entre montant de l'incitation et la motivation). Il vaut mieux de nouveaux budgets qu'un transfert. Engagement des offreurs : la collaboration et la communication semblent importantes pour l'efficacité des programmes.
Effet du contexte (à étudier)	Caractéristiques du système de santé Mode de rémunération principal (paiement à l'acte, salariat ou capitation) Caractéristiques de l'organisation et caractéristiques individuelles du médecin Caractéristiques des patients

A partir d'Eijkenaar et al. (2013), complété par nous

Ces résultats – confirmés par une récente « revue de la littérature des revues de la littérature » (Eijkenaar et al., 2013) – permettent de mettre en exergue les résultats financiers du P4P sur le plan de l'efficience. En effet, au-delà de l'efficacité, la nouvelle politique publique se justifie par son efficience, c'est-à-dire qu'elle doit atteindre ses objectifs à moindre coût. Il faut alors se demander si une autre politique que le P4P aurait pu obtenir les mêmes résultats à moindre coût. Très peu d'études s'intéressent à l'efficience des politiques de P4P en raison de la difficulté à la mesurer. Pour ce faire, il faudrait en effet pouvoir mettre en place deux types de politiques publiques différentes dans deux situations très semblables. À notre connaissance, dans le champ de la santé, la question de l'efficience n'a jamais été posée aux politiques de P4P. Cependant, bien qu'il n'y ait pas eu d'autres politiques de régulation mises en place permettant la comparaison, on peut essayer de commenter l'efficience du QOF britannique. Au regard de résultats qualitatifs et économiques assez minimes, le coût engendré ne peut apparaître que prohibitif. Entre l'exercice 2002-2003 et l'exercice 2005-2006 le surcoût du QOF s'élevait à plus de 500 millions de livre sterling. On peut raisonnablement supposer qu'à la vue de résultats aussi modestes, tant sur le plan économique que qualitatif, d'autres types de régulations auraient pu être mises au point pour un coût inférieur. L'efficience du QOF est, pour le moins, douteuse.

Enfin, la politique publique de P4P doit être jugée à l'aune de ses effets inattendus. Il s'agit d'effets qui n'étaient pas prévus par la théorie (comme par le législateur) et qui doivent entrer en ligne de compte pour l'évaluation des mécanismes de P4P. Même si théoriquement les effets inattendus peuvent être positifs ou négatifs, dans le champ qui nous intéresse ici, il s'agit essentiellement d'effets pervers. Or, comme nous l'avons vu précédemment, le P4P engendre toute une série d'effets pervers qui remettent en cause la pertinence de son utilisation (Tableau 4.9.).

Tableau 4.9. Les effets pervers du P4P médicale

Nom de l'effet pervers	Description	Bibliographie indicative
Sélection des tâches	Le médecin privilégie la tâche pour laquelle il est rémunéré au détriment d'autres tâches qui font partie de l'activité mais pour lesquelles aucune rémunération spécifique n'est prévue.	Shen (2003), Dorant et al. (2006), Doran et al. (2008)
Sélection des patients	Le médecin cherche à exclure (ou à choisir) certains types de patients dans la mesure où il sait ou pense savoir qu'il sera plus difficile (ou plus facile) de réaliser un bon score avec ces patients spécifiques.	Simon et al. (1999), Roski et al. (2009), Campbell et al. (2009)
Crowding out effect	L'incitation financière constitue une motivation extrinsèque qui évince la motivation intrinsèque du médecin - l'effet total pouvant aller jusqu'à une réduction de la motivation. L'incitation dés-incite.	Forest (2008), Lievaut (2011, 2012), Videau et al. (2010), Videau et al. (2011)
Surcodage	Le médecin manipule le codage de sa performance de façon à améliorer son score. Ce surcodage n'est pas nécessairement frauduleux, il peut relever de l'optimisation lorsque l'indicateur laisse une forte marge d'interprétation.	Milcent et Rochut (2009), Belorgey (2010)
Incohérence des guidelines	Le P4P repose sur des recommandations de bonnes pratiques qui peuvent parfois être opposées à d'autres *guidelines*.	Kravitz et Greenfield (1995), Rodwin (1997), Van Horn et al. (1997)
Conflits d'intérêts	Le patient peut avoir intérêt (légitimement ou non) à ce que le médecin ne suive pas les objectifs quantifiés issus du P4P alors que ce dernier a intérêt à suivre la recommandation pour percevoir sa sur-rémunération.	Allen, (1993), Thirlby et al. (1996) Chaix-Couturier et al. (2000)

2.3.2. Des limites méthodologiques et un programme de recherche aveugle

L'une des principales conclusions sur la littérature du P4P médicale porte sur les difficultés méthodologiques des économistes. En effet, si la conclusion qui s'impose est que le P4P ne fait pas la preuve de son efficacité, ce serait avant tout à cause des limites méthodologiques et d'un accès compliqué à des données complexes (Eijkenaar, et al. 2013, p. 128).

Le monde réel ne permet pas de faire des évaluations « pures ». Différents problèmes se posent : (i) il est souvent impossible d'avoir un groupe de contrôle permettant de distinguer l'effet du P4P d'autres effets. Le cas du QOF est exemplaire. (ii) Même lorsque c'est le cas, il est difficile de distinguer l'impact de la rémunération – censée motiver le médecin – du reste du P4P – feedbacks, indicateurs, objectifs, etc. (iii) On ne comprend pas pourquoi les études P4P contre feedback donnent des résultats similaires. (iv) Il est difficile de mesurer les effets inattendus – dont les effets pervers – dans la mesure où ne sait pas à quoi s'attendre. (v) Lorsque l'effet de l'incitation existe, on ne sait pas par quel mécanisme il opère : motivation ou investissement ? (vi) La majorité des études portent sur les systèmes anglo-saxons. Plus généralement, il n'y a jamais de données contextuelles dans les études sur le P4P.

Le manque de résultats probants ne remet jamais en cause l'enthousiasme des auteurs quant aux effets potentiellement bénéfiques de ce type de dispositif. Quelle que soit l'interprétation des résultats – plutôt optimiste comme van Herck et al. (2010) ou plutôt pessimiste comme Eijkenaar et al. (2013), le programme de recherches reste le même : il faut poursuivre les expériences sur le P4P médicale car il est hors de question de le remettre en cause :

> *« Nonetheless, in general, the potential of P4P to improve performance remains undisputed. There is a consensus that the way in which P4P is designed has important consequences for the incentives that physicians experience and how they might respond to them. As argued by several authors, the fact that P4P has not been very successful has partly been a consequence of flaws in program design » (Eijkenaar, 2013, p. 118).*

Si le P4P ne fonctionne pas (ou mal), c'est parce que l'on n'a pas trouvé… dans quelles conditions il devrait fonctionner. En d'autres termes, les systèmes de santé doivent s'adapter à l'outil P4P pour qu'il soit performant. Ainsi, il faudrait poursuivre les recherches en s'intéressant plus particulièrement aux questions de la conception (*design*) du P4P médical et du contexte :

> *« If P4P is to achieve desired results, careful consideration of program design is vital. The inconclusive evidence on the effectiveness of P4P suggests that design has not been optimal. In this respect, insight into the design of current P4P programs would be useful in improving P4P performance » (Eijkenaar, 2012, p. 252).*

Ce n'est pas le modèle théorique qui est à remettre en cause mais la façon dont il a été appliqué dans la réalité. Ce n'est pas le modèle théorique qui est à remettre en cause mais la réalité ?

3. Quelle motivation professionnelle pour le paiement à la performance ?

Comme la performance du P4P est limitée, on peut se poser la question de la motivation des médecins. Sont-ils intéressés à l'intéressement ? Depuis quelques années, la littérature empirique s'appuie sur l'hypothèse de substituabilité des registres de motivation (cf. chapitre 3) pour interroger l'adéquation des motivations médicales aux dispositifs incitatifs. Le P4P ne vient pas sur un terrain vierge de valeurs. Les médecins ont différents mobiles pour l'action et le succès ou le rejet du P4P est le fruit de la rencontre de ces motivations et du dispositif qui les conforte ou les modifie. Les médecins interprètent le P4P en fonction de leurs motivations professionnelles.

On peut répertorier trois types d'études empiriques sur les motivations professionnelles. On mobilise dans le tableau 4.10., les études empiriques sur les effets du P4P pour en tirer des enseignements sur les motivations professionnelles. Le tableau 4.11. rassemble les études empiriques cherchant à déterminer les effets de dispositifs incitatifs autres que le P4P sur la motivation des médecins. Enfin, le tableau 4.12. reprend les grandes lignes méthodologiques et théoriques des travaux empiriques cherchant directement à tester la MI ou le COE.

Tableau 4.10. Ce que révèle le paiement à la performance sur la motivation médicale

Auteur	Champ d'étude	Population étudiée	Problématique	Dispositif étudié	Effets majeurs du P4P
Géne-Badia et al. (2007)	Satisfaction des professionnels de santé (infirmiers et médecins) et des patients	257 Primary Care en Catalogne (Espagne) entre 2002 et 2003	Quelle efficacité des incitations sur la qualité des utilisateurs finaux ?	P4P et paiement après décision d'une commission de professionnels	Augmentation du stress Hausse du bien-être si rémunération par commission
Bruni et al. (2008)	Prise en charge du diabète de type 2	Médecins de la région Emilie Romagne	Quelle efficacité qualitative selon différents modes de paiement ?	P4P, 'pay for compliance', 'pay for participation'	Sans P4P, hausse de l'autonomie et, en conséquence, de la qualité
McDonald et Roland (2009)	Effet du P4P sur les pratiques médicales	Enquête auprès 20 Primary Care Britanniques et 20 Primary Care Californiens	Quelles sont les conséquences involontaires du P4P ?	P4P	Menaces sur la relation patient – médecin ; perte d'autonomie perçue ; changement dans la nature de la visite
McDonald et al. (2007)	Effet du P4P sur les pratiques médicales	2 Primary Care Britaniques	Effet du QOF sur la qualité des soins, la MI et l'autonomie de la clinique	P4P (QOF)	Pas d'effets pervers si : incitations faibles ; performance sur la qualité (pas économies) et organisation autonome
Spooner et al. (2001)	Appétence pour le P4P	Enquête auprès 36 Primary Care Britanniques (East Kent)	Pourquoi les médecins français ont-ils adhéré volontairement à un programme de P4P ?	P4P auprès des médecins et infirmiers	Pas d'effet pervers si : amélioration des soins ; faible incitation financière ; autonomie ; reconnaissance
Saint Lary et al. (2011)	Appétence pour le P4P	Enquête auprès de 15 médecins généralistes franciliens	Quels sont les déterminants de l'adhésion au P4P ?	Rapport aux caisses ; jugement porté sur le P4P ; Effets sur la pratique clinique et sur la relation avec le patient	Dichotomie : dépend de la perception de l'environnement institutionnel (notamment rapport avec les caisses)

Tableau 4.11. Ce que révèlent les effets d'autres types d'incitations sur la motivation médicale

Auteur	Champ d'étude	Population étudiée	Problématique	Dispositif étudié	Effets majeurs
Ammi (2011)	Politique de prévention	301 médecins généralistes libéraux de Bourgogne	Quel type d'incitation serait le plus efficace pour inciter à la prévention ?	Incitations : monétaires ; non monétaires : organisationnelles	Guide de bonnes pratiques non participatives mal perçu ; 25% des médecins ont une appétence pour le P4P ; 60% des médecins se plaignent du manque d'effectifs
Belorgey (2011)	Qualité des soins hospitaliers	Service des Urgences de l'hôpital Maubourg	Effet pervers du management par objectif	Objectif de réduction du temps de passage des patients en urgences	Baisse de la qualité avec management par objectif : opposition established / outsiders
Barley et al. (2011)	Préférences en termes de rémunération	1900 médecins généralistes français (2008)	Que pensent les médecins de leur rémunération ?	Pas de dispositif spécifique	Les médecins exerçant en groupe préfèrent le changement ; pas de caractéristiques sociodémographiques déterminantes
Franc et al. (2009)	Prévention de l'obésité infantile	611 pédiatres français	Quelle perception des contraintes économiques chez les pédiatres ?	Durée de consultation ; tarif conventionnel ; formation ; perception de son efficacité	Problème de formation ; les médecins ne pense pas qu'ils sont efficaces à du fait de l'origine environnementale et sociale de l'obésité
Bachimont et al. (2006)	Prévention du diabète de type 2	Médecins généralistes de la région centre	Pourquoi les médecins ne suivent pas les recommandations de bonnes pratiques ?	Recommandations de bonnes pratiques	Problème si : prescription de norme opposé aux valeurs libérales ; objectif économique ; maladie sociale

Tableau 4.12. Les tests sur le niveau de motivation intrinsèque des médecins et sur le *crowding out effect*

Auteur	Champ d'étude	Echantillon	Problématique	Dispositif étudié	Résultats majeurs
Forest (2008)	Motivation des praticiens hospitaliers	Médecins hospitaliers français	Quelle relation entre job design, motivation autonome et satisfaction au travail ?	Structure des tâches	Forte motivation autonome (MI) des médecins
Lievaut (2011)	Pratiques tarifaires considérées comme 'justes'	7896 omnipraticiens observés entre 1980 et 2000	Quelles sont les pratiques tarifaires ?	Nombre de dépassements d'honoraires et d'actes gratuits	Ouverture secteur 2 et politique publique sur l'offre de soins développe la marchandisation
Lievaut (2012)	Pratiques tarifaires considérées comme 'justes'	7896 omnipraticiens observés entre 1980 et 2000	Comment expliquer l'érosion des actes gratuits au profit d'actes chers ?	Nombre d'actes gratuits	Politique sur l'offre de soins (discours gestionnaire) développe la marchandisation
Videau et al. (2010)	Promotion de la santé (prévention et éducation à la santé)	528 médecins généralistes libéraux de région PACA (2002)	Part des MI dans la motivation totale et COE	Pas de dispositif spécifique	Juger les recommandations de bonnes pratiques injustes est un signe de faible MI ; cycle de vie des MI
Videau et al. (2011)	Promotion de la santé (prévention et ETP)	1 875 médecins généralistes libéraux du panel 5 régions	COE	Incitation, niveau de l'incitation et engagement professionnel	Pas de COE sauf si situation de burnout (épuisement professionnel)

La lecture des différentes expériences empiriques sur les motivations médicales permet de comprendre l'effet caché des dispositifs de P4P. Cette revue de la littérature permet d'écarter le rôle des caractéristiques sociodémographiques (âge, sexe, région d'exercice…) dans l'adhésion ou le rejet du P4P (Barley et al., 2011). Il n'y a pas de profil type du médecin acceptant ou ayant une préférence exogène pour le P4P. Les médecins sont diversement intéressés par l'intéressement et leur intérêt pour l'intéressement dépend globalement de deux grandes caractéristiques : l'autonomie de la profession médicale et le contexte.

La question de l'autonomie de la pratique médicale semble transversale à toutes les études empiriques étudiées. L'effet souvent contre-productif des mécanismes de P4P sur l'autonomie du médecin est en effet souligné par un grand nombre d'études. Par exemple, il apparaît que les mécanismes de P4P réduisent considérablement l'autonomie des médecins italiens ce qui, en conséquence, les décourage de prendre en charge la prévention du diabète de type 2 (Bruni et al., 2008). Par ailleurs, les guides de bonnes pratiques peuvent être jugés de façon négative et réduire la motivation justement parce qu'ils remettent en cause l'autonomie des médecins (Ammi, 2010 ; Videau et al., 2010). Ce sont alors les valeurs libérales de la médecine qui sont contrariées par la prescription de normes (Bachimont et al., 2006). Cette logique d'autonomie est confirmée par Spooner et al. (2001) pour lesquels les programmes volontaires de P4P implantés à la fin des années 1990 en Angleterre ont été plébiscités par les médecins du fait de l'encouragement des cliniques à l'autonomie.

Il semble que le « contexte » conditionne la bonne ou la mauvaise réception des mécanismes de P4P. Si les médecins français sont plus réticents au P4P que les médecins anglais, c'est notamment du fait d'objectifs initiaux différents et de relations différentes avec la tutelle (Bras, 2011)[66]. Alors que la mise en œuvre du mécanisme anglais s'était accompagnée de l'augmentation du budget du *National Health Service*, le mécanisme français s'inscrit dans le contexte d'une réduction des dépenses socialisées de santé. La mention d'un champ « optimisation des prescriptions » cadre le dispositif comme un outil de réduction des dépenses. Par ailleurs, si le mécanisme du P4P anglais est plus facilement accepté que le mécanisme français, c'est du fait de rapports plus pacifiés avec la tutelle.

Pourtant, le fait qu'une partie des médecins français aient choisi de s'engager dans le P4P nous montre que, outre le contexte institutionnel, ce sont les représentations que se font les médecins du contexte institutionnel qui déterminent les effets du P4P. Les dispositifs de

[66] Nous revenons sur le cas français dans le chapitre suivant.

P4P n'ont pas de qualités intrinsèques. Ils ne valent que par les représentations qui y sont associées (Saint-Lary et al., 2011). Ce n'est que de cette façon que l'on peut comprendre leur rationalité cachée. Ce n'est pas exclusivement l'intéressement qui intéresse les médecins signataires mais aussi les valeurs véhiculées par le dispositif (amélioration de la qualité des soins, changement dans la relation avec la tutelle, reconnaissance du travail accompli, etc.). À l'inverse, s'il est encore vivement rejeté par une autre partie de la profession, c'est justement parce que d'autres représentations qu'il véhicule sont antinomiques à leur conception de l'activité médicale.

La bonne réception des mécanismes de P4P dépend de l'adéquation entre les valeurs qu'il véhicule et les valeurs médicales. Or, ni les unes ni les autres ne sont naturelles. Elles relèvent d'un construit social. Le P4P n'est pas un dispositif neutre. Ce mécanisme d'intéressement diffuse une rationalité marchande. Mais, cette perspective marchande semble s'opposer à la conception traditionnelle de la relation patient-médecin. Le P4P jette alors un flou sur l'activité médicale. Cette situation, intenable tant du côté de l'offreur que du demandeur de soins, risque fort, à terme, de contrarier durablement tant l'efficacité que l'équité du système de santé.

Conclusion

La conclusion qui s'impose à l'issue de notre revue de littérature sur l'efficacité du P4P médicale et son impact sur la motivation des professionnels est assez claire : en dépit d'un coût souvent important, ce dispositif est incapable de faire la preuve de son utilité. Tant sur l'objectif d'amélioration de la qualité des soins que sur celui de réduction des dépenses, la littérature empirique ne peut pas démontrer un impact – positif ou négatif – du P4P médicale. Si l'efficacité de ce mode de rémunération reste débattue, un nombre croissant de travaux attestent de la création d'effets pervers remettant plus radicalement en cause sa pertinence. Par exemple, les médecins ont intérêt à écarter les patients trop difficiles ou les tâches non rémunérées. Plus largement, en focalisant l'attention sur la performance chiffrée, ce mode de rémunération n'incite pas tant à améliorer les comportements qu'à jouer sur les indicateurs. Le principe même du P4P médicale semble en conséquence devoir être questionné.

Pourtant, le plus étonnant concernant cette littérature empirique porte sur les interprétations qui sont faites des résultats. Plutôt que de remettre en cause le cadre théorique

ayant mené à prescrire le P4P, les auteurs préfèrent chercher les obstacles qui dans, la réalité, empêchent le P4P d'être efficace.

À ce titre, deux explications majeures émergent. D'une part, le P4P souffrirait d'un problème de conception (*design*). Les médecins sont des *homo economicus* qui réagissent aux mécanismes d'incitation de façon tout à fait rationnelle : une incitation mal conçue implique nécessairement des inefficacités et des effets pervers. Il faudrait alors améliorer la conception du P4P médicale car il n'existe pas de préférences qui puissent résister à un bon schéma d'incitation. D'autre part, si le P4P est si difficile à concevoir, c'est à cause du « contexte ». En fait, ce qui pose problème à la thématique incitative, c'est l'impureté et la variabilité du monde réel.

La mise en exergue du « problème » de contexte est particulièrement intéressante dans la mesure où nous avons montré en début de chapitre que le P4P médicale est aujourd'hui devenu une « solution miracle » de politique publique à travers le monde. En effet, quels que soient les contextes locaux où il est intégré, ce mode de rémunération est doté du pouvoir – désormais fortement contestable – d'améliorer la qualité des soins et de réduire les dépenses de santé. La comparaison des cas américain et britannique est en ce sens emblématique : alors que ces deux pays font face à des problématiques radicalement différentes – l'un étant en sous-financement chronique par rapport aux pays comparables, l'autre étant le plus dispendieux du monde –, ils ont tous deux introduit le P4P médicale pour résoudre leurs problèmes.

Le chapitre suivant propose de prendre au sérieux ces questions du *design* et du contexte par une étude pionnière du dispositif de P4P médicale français.

Chapitre 5

Le paiement à la performance médicale en France : les performants ne sont pas forcément les gagnants

Introduction

D'après la théorie économique *mainstream,* le paiement à la performance (P4P) médicale permettrait d'améliorer la qualité des soins et d'économiser les précieuses ressources publiques. En brisant l'opacité de la relation patient-médecin et en fournissant des standards de production, la rente de situation de la profession médicale serait instantanément réduite à rien ou presque.

Malgré des situations nationales très différentes, c'est à partir de ce type de raisonnement générique qu'ont proliféré dans le monde les mécanismes de P4P médicale. Si les États-Unis et le Royaume-Uni ont tous deux mis en place ce type de dispositifs dans le but de réduire les dépenses de santé, leur situation initiale est loin d'être similaire : les États-Unis sont le pays au monde dépensant le plus pour la santé alors que les études sur le sujet jugent régulièrement que le système britannique est sous-financé. Bien que la médiocrité des résultats des différents systèmes de santé en termes d'état de santé, de qualité des soins et/ou de dépense puisse provenir de causes multiples, la solution privilégiée des politiques publique est le P4P médicale. Les politiques publiques visant l'activité médicale semblent s'uniformiser malgré les dépendances de sentier.

La diffusion du P4P médicale ne va pourtant pas sans problèmes. La revue de littérature réalisée dans le chapitre 4 sur l'évaluation du P4P a en effet montré que ce mode de rémunération éprouve d'immenses difficultés à faire la preuve de son utilité. L'efficacité de ce dispositif est limitée, son efficience est douteuse et il provoque des effets pervers nuisibles tant à la qualité des soins qu'à l'usage des ressources.

Malgré ces faits biens documentés, le P4P a été importé en France à la fin des années 2000. En 2009, les médecins libéraux se sont vus proposés le Contrat d'amélioration des pratiques individuelles (CAPI). Ce contrat prévoyait que les médecins éligibles volontaires

puissent recevoir une prime lorsqu'ils atteignent des objectifs chiffrés décidés par l'assurance maladie. Au départ, deux grands types d'indicateurs étaient proposés aux médecins signataires : prévention et suivi des maladies chroniques, d'une part, et optimisation des prescriptions, d'autre part. L'idée, toujours en vigueur, est d'inciter les médecins français à améliorer leurs pratiques en les récompensant financièrement pour le faire.

Fondé sur le volontariat, ce dispositif a été signé par un tiers des médecins éligibles (Polton et Aubert, 2010). Encouragés par cette forte participation et sans attendre l'analyse statistique des résultats du CAPI, l'assurance maladie et les syndicats de médecins se sont entendus en 2011 autour d'une convention médicale nationale prévoyant la généralisation du P4P au-delà des médecins volontaires. Ainsi, depuis le 1er janvier 2012, l'ensemble des médecins traitants français sont éligibles à la Rémunération sur objectif de santé publique (ROSP). En plus des indicateurs de prévention, de suivi des pathologies chroniques et d'optimisation des prescriptions, le ROSP prévoit une rémunération à la performance pour des indicateurs d'organisation et de qualité du service. Malgré un ciblage initial implicite sur les médecins généralistes, les diverses annexes à la convention négociées et signées depuis cette date étendent le P4P à la médecine spécialisée (spécialistes des maladies vasculaires, gastroentérologie et d'hépatologie). Aujourd'hui, la généralisation du ROSP est bien avancée et il n'y a aucune raison pour qu'elle s'arrête, tant au niveau des sommes allouées qu'au nombre d'indicateurs utilisés.

Cette généralisation semble néanmoins problématique dans la mesure où l'évaluation des expériences étrangères est loin d'être encourageante. Concernant le CAPI, bien que l'assurance maladie se félicite de ses bons résultats, rien ne permet de dire que le P4P médicale version française ait atteint ses objectifs. On ne sait d'ailleurs pas précisément quels sont les objectifs de la puissance publique. À notre connaissance, aucune étude n'a permis de montrer que le CAPI a permis d'augmenter significativement la qualité des soins et des prescriptions – mis à part la publicité faite par l'assurance maladie elle-même.

L'objectif de ce chapitre est de fournir des éléments de preuve empirique permettant d'évaluer le ROSP. La plus grande difficulté pour ce faire réside dans l'absence de groupe de contrôle permettant de comparer la situation avec ROSP et la situation sans ROSP[67]. Il est

[67] Comme dans le cas du *Quality Outcome Framework*, bien que la possibilité fût laissée de refuser le P4P médicale, les médecins s'y sont soumis. Dans le cas britannique, la raison est liée à l'ampleur massive des financements alors que, dans le cas français, la raison est liée aux démarches administratives nécessaires pour refuser (envoi d'une lettre recommandée dans un délai relativement court).

donc impossible d'évaluer l'efficacité et l'efficience de ce dispositif avec la rigueur et l'autorité scientifique nécessaires. En conséquence, nous adoptons une approche empirique tournée vers la détection d'effets pervers. La problématique générale de ce chapitre porte sur la signification de la « performance » dans le dispositif français de paiement à la « performance » – le ROSP. Notre hypothèse de travail est qu'il existe un écart très important entre la performance mesurée par le dispositif et la performance théorique. Deux questions principales animent nos travaux :

(i) Est-ce que l'amélioration de la performance mesurée est identifiable à une amélioration de la performance dans un sens plus profane ? Il ne s'agit ici pas de mesurer une possible amélioration de la performance consécutive à l'introduction de la ROSP : nous ne cherchons pas à évaluer l'efficacité de la ROSP, c'est-à-dire sa capacité à atteindre ses objectifs chiffrées. L'attention se porte plutôt sur la signification d'une hypothétique amélioration de la performance mesurée – telle qu'elle est définie par la ROSP. Améliorer ses indicateurs permet-il d'améliorer la qualité des soins apportés aux patients ?

(ii) Est-ce que la performance est imputable au médecin pour lequel sont calculés les indicateurs de performance ? Indépendamment de la question de l'efficacité de la ROSP ou de la pertinence de ses indicateurs, il s'agit de découvrir si la seule et unique cause de la variation des indicateurs est l'action du médecin. Autrement dit, est-ce que la performance médicale ne dépend que du bon vouloir du médecin ?

À ces deux questions la suite du chapitre fournit des éléments de preuve empirique permettant de répondre négativement. Pour cela, ce chapitre s'organise en trois sections.

La première section commence par faire une présentation critique de la ROSP (1.). Jusqu'à maintenant nous avons beaucoup parlé de P4P médicale, il est temps de décrire et d'analyser un cas concret, celui de la ROSP. L'étude minutieuse du dispositif peut nous permettre de comprendre avec précision ce que performance veut dire – d'après la formule de rémunération du ROSP et sa critique. Peut-on identifier « performance », et ROSP ? Peut-on identifier ROSP et qualité des soins ? Peut-on identifier « performance » et effort ou mérite du médecin ? A certains égards, cet objectif est plus ambitieux que de simplement vérifier si l'incitation incite. En effet, dans la perspective épistémologique ouverte par Alain

Desrosières, il s'agit de discuter de la réalité de la chose mesurée et incitée[68]. Sans même nous intéresser aux données empiriques, nous pourrons déjà nous apercevoir de l'écart problématique entre la performance mesurée et la performance théorique. Nous insisterons notamment sur la différence entre le calcul du taux de réalisation global de l'indicateur (en pourcentage de réussite ou en points) et le calcul de la prime correspondante (en euros). La définition de la performance est alors multiple selon que l'on parle de la performance en termes de taux de réalisation global ou de performance en termes de rémunération à la performance.

Les deux sections suivantes sont à proprement parler les sections empiriques de ce chapitre. À notre connaissance elles constituent le premier travail empirique sur la ROSP. En collaboration avec le Service Médical de l'Assurance Maladie d'Ile de France (SMAMIF), nous avons pu analyser la performance (aux différents sens du ROSP) des médecins généralistes de cette région.

La section 2. présente les données ainsi qu'une analyse des statistiques descriptives concernant les médecins généralistes et leur performance. Le principal résultat porte sur la mise en évidence empirique de la différence entre les médecins « gagnants » et les médecins « performants ». Nous définissons les premiers comme ceux ayant obtenu une rémunération à la performance supérieure au troisième quartile des rémunérations et les seconds comme ceux ayant obtenu un taux de réalisation global supérieur au troisième quartile des taux de réalisation. Contrairement à ce que l'on pourrait penser, les performants ne sont pas forcément les gagnants. Autrement dit, ceux qui ont les meilleurs taux de réalisation ne sont pas forcément ceux qui ont les plus fortes rémunérations à la performance.

La dernière section restreint le champ d'investigation afin de procéder à une étude économétrique (3.). Nous testons l'hypothèse selon laquelle la composition de la patientèle (le *case mix*) – *a priori* indépendante de la responsabilité et de l'implication du médecin – est corrélée à la performance du médecin. En ce sens, l'obtention de bons résultats en termes de taux de réalisation global ou de rémunération à la performance ne serait pas uniquement dépendante de la motivation du médecin mais aussi de facteurs indépendants de sa volonté – ici, la composition de sa patientèle (définie notamment par un indice de pauvreté et un indice de son état de santé).

[68] Il ne s'agit pas de savoir si la courbe du chômage s'inverse ou non mais de savoir ce que veut dire le chiffre du chômage – qu'il soit à la hausse ou à la baisse.

1 Le paiement à la performance médicale en France

Nous présentons d'abord le contexte dans lequel s'insère le dispositif de P4P médicale en France : la Rémunération sur objectif de santé publique (1.1.). Ensuite, nous étudions plus précisément la formule de calcul de la rémunération telle qu'elle est définie dans la convention nationale de 2011 (1.2.). Enfin, nous analysons les indicateurs ROSP au regard de l'épistémologie des objets statistiques proposée par Alain Desrosières (1.3).

1.1. La justification de la Rémunération sur objectif de santé publique

Comme toutes les réformes du système de santé depuis de nombreuses années, la première justification de la mise en place du P4P médicale en médecine ambulatoire repose sur la maîtrise des dépenses socialisées de santé. Le besoin de financement de la Sécurité sociale se renouvelle chaque année légitimant la nécessaire réforme d'une institution dépassée. Les différentes lois de financement de la Sécurité sociale peuvent ainsi se lire comme autant de plan de maîtrise de l'évolution des dépenses. Si elle est protégée par son statut libéral, la médecine ambulatoire n'échappe pas à cette règle. En introduisant le P4P médicale, la Caisse nationale d'assurance maladie (CNAM) a devancé l'État dans l'objectif de contrôle des dépenses. Les indicateurs d'optimisation des prescriptions peuvent ainsi se lire comme un investissement : en payant les médecins pour prescrire des médicaments génériques, on s'attend en retour à une réduction des dépenses de remboursement sur les médicaments princeps – c'est-à-dire les médicaments de marque, plus cher à efficacité thérapeutique égale.

La deuxième source de légitimité du P4P médicale dans le cas français repose sur l'ambition d'amélioration de la qualité des soins. Dans le cas du diabète par exemple, les études montrent que la prise en charge des patients diabétiques n'est pas conforme aux recommandations pourtant très précises de la Haute Autorité de Santé (HAS). Comme le souligne Dominique Polton (2010), directrice des études statistiques à la CNAM, dans un article présentant la démarche de l'assurance maladie, alors que la HAS préconise de réaliser un dosage d'HbA1C par trimestre, moins de 50 % des patients diabétiques en bénéficient. En France comme ailleurs, l'objectif du P4P médicale est également d'améliorer une qualité de soins jugée relativement médiocre.

Ce discours n'est pas neuf. C'est exactement le même qui a conduit à la généralisation du P4P médicale à travers le monde. Si l'on se rappelle de la discussion menée au chapitre précédent sur les cas américain et britannique, force est de constater qu'à des situations différentes – en termes de dépense publique, d'état de santé de la population, de qualité des soins –, il ne semble exister aucune alternation au P4P. La France n'échappe pas à cette uniformisation.

C'est dans le cadre général qu'est apparu le P4P en France. Après l'expérience du CAPI, la négociation conventionnelle entre la CNAM et les syndicats médicaux représentatifs a débouché sur le développement du P4P médicale à tous les médecins traitants libéraux. La convention médicale signée le 26 juillet 2011 est le témoin d'un infléchissement des syndicats médicaux qui acceptent désormais le principe du P4P médicale – alors qu'ils avaient vivement contesté le CAPI. La négociation conventionnelle ne porte donc plus sur l'existence du P4P médicale mais sur ses modalités. La convention substitue ainsi la Rémunération sur Objectif de Santé Publique (ROSP) au CAPI en donnant une nouvelle dimension au P4P médicale : le ROSP devient le troisième pilier de la rémunération de la médecine libérale – aux côtés du paiement à l'acte et des forfaits.

En 2012, le ROSP représentait une dépense de 282 millions d'euros pour environ 75 000 médecins, dont 51 000 généralistes (CCSS, 2013). La rémunération moyenne était alors de 3 750 € pour l'ensemble des médecins et de près de 5 000 € pour les médecins généralistes (soit une hausse de 3% de leur salaire). En 2014, l'assurance maladie a publié des chiffres jugés par elle positifs sur les différents objectifs du ROSP. Tous indicateurs confondus, les médecins auraient amélioré leur performance globale de 51.7% à 60.4 % (CNAM, 2014) – 49.5% à 56.1% pour le suivi des pathologies chroniques ; 55% à 63.3% pour l'optimisation des prescriptions ; 34.5% à 40 % pour la prévention ; organisation (pour omnipraticiens) : 58.3% à 70% pour les indicateurs d'organisation du cabinet. Les 85 187 médecins concernés en 2013 ont reçu en moyenne 4 003 €, 5 480 € pour les omnipraticiens, 5 774 € pour les médecins généralistes et 1 082 € pour les spécialistes concernés uniquement par les indicateurs d'organisation du cabinet. Au total la dépense effectuée par l'assurance maladie s'élevait en 2013 à 341 millions d'euros.

En dépit de l'optimisme de la CNAM, il n'est pas statistiquement possible de savoir si les améliorations constatées de la performance médicale sont dues au mécanisme incitatif ou à autre chose. Il est tout aussi possible que l'amélioration des scores soit le fait d'une évolution plus globale des pratiques ou d'un effet d'information – les médecins connaissant désormais

la marche à suivre. Dominique Polton (2010) rappelle que, sur la question du diabète, les médecins n'ont pas attendu le CAPI ou le ROSP pour améliorer leurs pratiques. Pour reprendre l'exemple précédent, entre 2001 et 2007 (soit avant le ROSP) le pourcentage de patients diabétiques ayant eu au moins trois dosages de HbA1C a augmenté de 10 points.

Il est n'est pas facile de conclure sur l'efficacité et l'efficience du P4P médicale français. Cependant, il reste l'opportunité d'étudier la signification de cette « performance » – qu'elle croisse ou non. La première étape pour cela est d'analyser la formule de rémunération du ROSP.

1.2. La formule de rémunération du ROSP

Cette section cherche à éclaircir la formule de rémunération du ROSP à partir de la convention signée en 2011 (Titre 3, sous-titre 5) et de ses annexes (Annexe XVII). L'objectif est de comprendre ce que signifie « performance » dans paiement à la « performance ». Que calcule cet indicateur ? Quelle définition de la performance y est sous-entendue ?

Pour le médecin i, la rémunération à la performance est donnée par :

$$Y_i = \sum_{n=1}^{z} (s_n * t_{ni} * v_{ni}) * \frac{l_i}{800}$$

Avec s_n, le nombre de points fixés par l'assurance maladie pour l'indicateur n et t_{ni} le taux de réalisation de l'indicateur n pour le médecin i. v_{ni} est la valeur du point en euros de l'indicateur n pour le médecin i. Dans la majorité des cas, la valeur du point est fixée à 7 €. Mais, l'article 2 de l'annexe XVII prévoit que la valeur du point, pour les indicateurs de la qualité de la pratique médicale, soit majorée pour les médecins nouvellement installés. La majoration est de 15 % la première année, 10 % la deuxième et 5 % la troisième. Aucune majoration n'est prévue pour les indicateurs d'organisation du cabinet.

l_i représente la patientèle médecin traitant du médecin i. En effet, le calcul de la ROSP ne se fonde pas sur l'ensemble de la patientèle d'un médecin mais uniquement sur la partie de celle-ci qui l'a déclaré Médecin traitant (MT). La déclaration de MT, qui n'est pas obligatoire mais fortement incitée, concerne tous les patients de plus de 16 ans quel que soit leur régime d'affiliation (Régime général, Mutualité sociale agricole, Régime social des indépendants, etc.). L'article 2 de l'annexe XVII à la convention médicale prévoit que le nombre de points par indicateur soit défini pour une « patientèle moyenne de référence de 800 patients ». En

conséquence, les points obtenus pour chaque indicateur doivent être pondérés par le rapport entre patientèle réelle MT et patientèle moyenne MT de référence. Si le médecin dispose d'une patientèle réelle MT égale à la patientèle de référence, le coefficient est 1 et le nombre de points associés à l'indicateur est précisément celui inscrit dans la convention. Si la patientèle réelle est supérieure, le nombre de points par indicateur est majoré, si elle est inférieure, il est minoré. Par exemple, un médecin disposant d'une patientèle médecin traitant de 1 600 patients verrait son nombre de points par indicateur doublé – indépendamment de son taux de réalisation.

Le taux de réalisation de l'indicateur n pour le médecin i (t_{ni}) est donné par la formule suivante :

$$t_{ni} = \begin{cases} 50\% * \dfrac{n_c - n_i}{o_i - n_i}, si\ n_c < o_i \\ 50\% + 50\% * \dfrac{n_c - o_i}{o_c - o_i}, si\ n_c \geq o_i \end{cases}$$

Avec n_c, le niveau constaté de l'indicateur n, n_i, le niveau initial de l'indicateur n, o_c, l'objectif cible de l'indicateur n et o_i, l'objectif intermédiaire de l'indicateur n. Pour aider à la compréhension, prenons l'exemple de l'indicateur de prévention du cancer du sein (thème prévention, sous thème cancer du sein). Cet indicateur comptabilise le nombre de patientes MT de 50 à 74 ans participant au dépistage (organisé ou individuel) du cancer du sein rapporté au nombre de femmes MT de 50 à 74 ans. Le tableau 5.1. montre comment se présentent les différents niveaux et objectifs.

Tableau 5.1. L'indicateur de dépistage du cancer du sein

Niveau initial (n_i)	Niveau constaté (n_c)	Objectif intermédiaire (o_i)	Objectif cible (o_c)	Seuil	Points associés (s_n)
Observé en T_0 pour chaque médecin i	Observé en T_1 pour chaque médecin i	70 %	≥ 80 %	20 patientes	35

Le niveau initial est celui observé à la date d'entrée en vigueur du dispositif (31 décembre 2011) ou à la date d'installation du médecin (en cas de première installation en libéral). Le niveau constaté est le niveau observé à chaque date anniversaire d'entrée en vigueur du dispositif. L'objectif intermédiaire est le seuil à partir duquel les médecins sont éligibles à au moins la moitié des points ROSP. L'objectif cible est le niveau pour lequel le

médecin peut prétendre au nombre de points maximal. À chaque indicateur est défini un seuil minimal pour que ce dernier soit pris en compte. Dans cet exemple, il faut que le médecin ait dans sa patientèle MT au moins 20 patientes ayant entre 50 et 74 ans pour que son indicateur soit calculé. S'il n'est pas calculé, cela ne veut pas dire que le médecin n'est pas performant mais qu'il n'a pas une patientèle suffisamment large pour que son activité soit évaluée. Enfin, à chaque indicateur correspond un nombre de points associés (s_n) qui est une mesure de la valeur qu'accorde la CNAM à tel ou tel indicateur. Dans notre exemple, si x % des patientes entre 50 et 74 ans du médecin traitant i participent au dépistage (organisé ou individuel) du cancer du sein, ce médecin obtient $\frac{x}{100} * 35$ points[69]. Notons que contrairement au calcul de la taille de la patientèle, le calcul des indicateurs porte sur la patientèle dite « consommante et fidèle ». Il s'agit de patients qui ont eu recours à des soins (consultation, traitement, examens, etc.) qui ont été remboursés sur une période de 12 mois précédent la période de calcul (consommante) et qui n'ont pas changé de médecin traitant au cours de l'année (fidèle). On peut simplifier l'écriture des taux de réalisation en notant t_i le taux de réalisation global pour le médecin i.

$$t_i = \sum_{n=1}^{z} (s_n * t_{ni})$$

Cette formule suppose soit, que tous les médecins sont installés depuis plus de trois ans soit, que les indicateurs d'organisation des cabinets sont exclus. Dans les deux cas, il est possible d'extraire v_{ni} de la somme, d'où la nouvelle formule de rémunération :

$$Y_i = t_i * v_i * \frac{l_i}{800}$$

Notons que s_n et v_i sont exogènes à la décision du médecin. En effet, le nombre de points attribués à chaque indicateur (s_n) et la valeur en euros du point pour le médecin (v_i) sont décidés dans le cadre conventionnel. Le médecin i ne peut agir que sur t_{ni} (que l'on a simplifié par t_i) et sur l_i pour améliorer sa prime (Y_i).

En anticipation de l'étude empirique des sections suivantes, nous introduisons deux définitions permettant d'éclairer la polysémie de la notion de performance dans le cas du ROSP :

<u>Définition 1 :</u> nous dirons qu'un médecin, est « performant » si et seulement si son taux de réalisation global (t_i) est supérieur ou égal au troisième quartile de la distribution des taux

[69] Cette valeur restant encore à pondérer par la taille de sa patientèle par rapport à la patientèle de référence.

de réalisation. La performance est ici définie en termes relatifs. Un médecin est performant, non par rapport à ses résultats absolus (indicateur du médecin i), mais par rapport à la distribution des résultats des autres médecins (indicateur du médecin i / indicateur des autres médecins). Nous nous écartons ainsi de la définition implicite de l'AM (Encadré 5.1.). Selon celle-ci, est performant un médecin qui est rémunéré à la performance. Notre objectif en modifiant la définition de la performance est de comparer les niveaux de performance pour affiner le sens que prend cette qualification. Il s'agit notamment d'expliquer la distribution des performances. Tous les médecins ne sont pas performants de la même façon.

Encadré 5.1. La performance selon l'assurance maladie

L'assurance maladie ne distingue pas le fait de recevoir une prime et celui d'obtenir un taux de réalisation global minimal. En effet, du point de vue de l'AM, la performance est définie par la fonction de prime. L'idée implicite, qui n'est pas sans fondements, est que la performance doit être définie en termes de santé publique. Dans ce cadre, il est non seulement important que le médecin obtienne de bons indicateurs mais aussi qu'il « touche » une population la plus grande possible. Une explication complémentaire de la présence d'une pondération par la taille de la patientèle est l'effort réalisé par le médecin. En situation de pénurie, il est demandé au médecin de répondre à la demande (donc de conserver une taille minimale de patientèle) et d'augmenter la qualité des soins (donc d'améliorer ses scores).

Plus largement il est possible de distinguer au moins deux définitions de la performance :

- Le taux de réalisation global : $s_n * t_{ni} = t_i$
- Le taux de réalisation global pondéré par la patientèle : $t_i * l_i / 800$

Notons ici que le chiffre de 800 est considéré implicitement comme la patientèle normale, c'est-à-dire celle qui ouvre le droit à un calcul de la prime à partir du nombre de points inscrits dans la convention. Le médecin qui dispose d'une telle patientèle voit en effet son « coefficient patientèle » se normaliser à 1. En dessous, le nombre de points diminue, au-dessus, il augmente.

Quelle que soit la définition, il est à chaque fois possible de calculer cette performance en euros en lui attribuant la valeur du point (v_i). Ici, nous séparons le calcul de la performance et de la prime. Par ailleurs, nous préférons la définition de la performance à partir du taux de réalisation global. L'argument principal est que l'augmentation de la patientèle ne constitue pas une amélioration de l'indicateur – lui-même destiné à améliorer la qualité des soins (c'est la justification première de l'AM). Il nous semble donc impropre de parler d'augmentation de la performance. Dirait-on d'un médecin qui réalise plus d'actes qu'il améliore la qualité des soins ?

Définition 2 : nous dirons d'un médecin qu'il est gagnant si et seulement si sa prime (Y_i) est supérieure ou égale au troisième quartile de la distribution des primes. À nouveau, il ne s'agit pas d'une définition absolue mais d'une définition relative. L'intérêt de cette définition est de montrer que, paradoxalement, ce n'est pas parce qu'un médecin est performant qu'il est gagnant. Autrement dit, ce n'est pas parce qu'un médecin obtient un bon taux de réalisation global (t_i) qu'il en sera récompensé par une meilleure prime (Y_i) – comme le voudrait la logique du paiement à la performance. Si cette hypothèse est vraie, le paiement à la performance n'est pas qu'une question de mérite individuel du médecin.

Maintenant que nous avons établi la formule de la rémunération à la performance dans le ROSP, nous allons analyser en profondeur les différents indicateurs la constituant.

1.3.Les indicateurs de performance du ROSP : entre réalisme et constructivisme

Les économistes ont pris pour habitude dans leur travail statistique de se satisfaire des nomenclatures qui leur sont offertes par des constructions antérieures sans se poser de questions sur leur « réalité » (Desrosières, 2008, p. 119). De ce fait, au moins implicitement, les économistes souscrivent à une épistémologie réaliste des objets statistiques dans la mesure où il n'est pas question de remettre en cause la réalité de l'objet observé. Cela ne veut pas dire que les économistes ne veulent pas le faire, cela veut simplement dire que bien souvent la question ne se pose même pas. À l'inverse, d'après l'épistémologie constructiviste des objets statistiques, il est toujours possible et nécessaire de questionner les choix – souvent arbitraires – qui ont conduit à utiliser telle ou telle méthode de traitement des données ou à construire telle ou telle nomenclature. Dans cette optique, le problème posé ne provient pas tant de la précision de la mesure mais de la définition même de ce qui est mesuré.

Par exemple, le travail de l'économiste du travail n'est pas seulement de dénombrer le nombre de chômeurs mais aussi, de discuter la définition du chômage. Ainsi, dans une note d'avril 2012, le collectif les Autres chiffres du chômage (ACDC, 2012) estimait ce dernier en 2010 en France métropolitaine à 44.9 % de la population active alors que l'Institut national de la statistique et des études économiques (INSEE, 2011) comptabilisait 9.2 % de chômeurs à la même période. Evidemment, l'écart entre les deux chiffres repose sur la définition de l'objet à observer.

Cette problématique n'est pas purement théorique. En effet, la statistique n'est pas neutre, elle constitue un argument de poids qu'il convient de maîtriser. S'ouvre ici le questionnement sur l'usage des statistiques dans le débat public. Restons sur l'exemple du chômage. Comme l'explique Robert Salais (2010), lorsqu'il est dit « le chômage a baissé de 5 % », l'énonciateur fait comme s'il était possible de passer de la proposition A – le nombre de demandeurs d'emploi inscrit à Pôle emploi en catégorie 1 a diminuer de 5 % – à la proposition B – le chômage a diminué de 5 % – sans s'interroger sur la légitimité de ce passage. En effet, le mot chômage renvoi à un ensemble d'idées préconçues qui lui donnent un sens dans le langage courant. C'est pourquoi la proposition A diffère grandement de la proposition B. La statistique n'est plus uniquement un fait empirique mais aussi un jugement

de valeur dans la mesure où il est possible de penser que l'« état du monde » s'améliore – alors qu'il n'y a pas forcément d'accord sur ce que veut dire chômage. Si la proposition A produit des effets cognitifs menant à penser que l'« état du monde » s'améliore – confondre A et B – alors le travail du statisticien est en cause. Les recherches sur la quantification doivent en conséquence être attentives aux modalités d'usage de l'argument statistique, c'est-à-dire comprendre sur quoi repose la validité des propositions A et B ainsi que discuter de la légitimité de leur identification.

Le même problème se pose dans le cas de la performance médicale. Dire « le taux de réalisation global s'est amélioré de 5 % » – proposition 1 – n'est pas nécessairement équivalent à dire « la rémunération à la performance des médecins s'est amélioré de 5 % » – proposition 2 – ou bien « la performance des médecins s'est amélioré de 5 % » – proposition 3 – ou encore « la qualité des soins s'est améliorée de 5 % » – proposition 4. Si la rémunération à la performance et le taux de réalisation global sont des indicateurs chiffrés dont il est possible de comparer l'évolution, rien ne garantit qu'il soit possible d'identifier ces états à une amélioration de la performance des médecins ou à une amélioration de la qualité des soins. Or, l'usage de l'argument statistique peut produire une identification problématique de ces 4 propositions.

Dans ce chapitre, nous ne souhaitons pas seulement étudier les relations des nombres entre eux. Nous souhaitons également étudier la façon dont ont été conçues les nomenclatures qui nous occupent. C'est pourquoi nous consacrerons dans la suite de l'analyse une place importante à la discussion des nomenclatures utilisées. Nous pensons que de cette manière nous récolterons des éléments permettant de comprendre la signification de la « performance » et, en creux, la plus ou moins grande sensibilité des médecins à la politique publique de ROSP. Nous commençons par présenter les indicateurs de la ROSP (1.3.1.) pour ensuite en faire une évaluation critique (1.3.2.).

1.3.1. Les indicateurs de la ROSP

La convention médicale de 2011 distingue deux dimensions de la pratique médicale dans lesquelles sont mentionnés les 29 indicateurs de la ROSP.

- Les indicateurs d'organisation et de qualité de service :

Ces indicateurs (Tableau 5.2.) sont calculés pour tous les médecins mis à part les indicateurs d'élaboration et de mise à disposition d'une fiche de synthèse annuelle intégrée au

dossier médical informatisé (uniquement pour les médecins traitants)[70]. Il s'agit de mesurer l'utilisation de logiciels professionnels pour la gestion des données du dossier médical, l'usage de logiciels d'aide à la prescription, la transmissions d'information à destination des patients (horaires d'ouverture, plage de consultation avec et sans rendez-vous, accessibilité pour les personnes atteintes d'un handicap, etc.).

Tableau 5.2. Indicateurs portant sur l'organisation du cabinet et de qualité du service

Champs	*Indicateurs*	*Points*
Organisation du cabinet	Mise à disposition d'un justificatif comportant un descriptif de l'équipement permettant la tenue du dossier médical informatisé et la saisie de données cliniques pour le suivi individuel et de la patientèle	75
	Mise à disposition d'un justificatif témoignant de l'utilisation d'un logiciel d'aide à la prescription certifié *	50
	Mise à disposition d'un justificatif d'équipement informatique permettant de télétransmettre et d'utiliser des téléservices	75
	Affichage dans le cabinet et sur le site ameli** des horaires de consultations et des modalités d'organisation du cabinet, notamment pour l'accès adapté des patients	50
	Elaboration à partir du dossier informatisé, et mise à disposition, d'une synthèse annuelle par le médecin traitant pour ces patients	150
Total de points		**400**

* ou pour les médecins non prescripteurs (anatomo-cytopathologiste) d'un module de création de compte-rendu standardisé alimentant le dossier médical informatisé ou le DMP.
** dès lors que le service sur ameli sera opérationnel.

Remarquons que pour ce qui concerne presque tous les indicateurs d'organisation du cabinet et de qualité des services, il n'y a pas d'objectifs chiffrés proprement dit. Au mieux, l'objectif est binaire : 0 si l'objectif est atteint, 1 sinon. Le seul indicateur qui propose un objectif chiffré non binaire est celui de télétransmission. Pour être rémunéré, il faut que le médecin atteigne l'objectif de 2/3 des feuilles de soins télétransmises.

- Les indicateurs de qualité de la pratique médicale

Ces indicateurs sont supposés mesurer la qualité de la prise en charge médicale en termes de suivi des pathologies chroniques (Tableau 5.3.), de prévention (Tableau 5.4.) et d'efficience (Tableau 5.5.). Plus précisément encore, il s'agit de mesurer « l'implication des médecins » dans ces trois thèmes. Selon la convention, les indicateurs sont fixés à partir de la loi de santé publique et en tenant compte des avis et référentiels émis pas la Haute Autorité de Santé et des recommandations internationales « afin d'être adaptés aux données acquises de la

[70] Il n'y a donc pas de seuil minimal de patientèle (MT ou autre).

science » (Convention, 2011, p. 36). Notons que lorsque l'intitulé de l'indicateur est précédé de la notation (d), cela signifie que l'indicateur est déclaratif. Autrement dit, ce n'est pas par l'enregistrement électronique d'une consommation qu'est calculé l'indicateur mais, par la déclaration que le médecin concerné fait à la Caisse primaire d'assurance maladie.

Tableau 5.3. Les indicateurs de qualité de la pratique médicale. Thème suivi des pathologies chroniques

Sous thème	*Indicateurs*	*Objectifs intermédiaire et cible*	*Seuil minimal*	*Points*
Diabète	Nombre de patients traités par antidiabétiques ayant choisi le médecin comme « médecin traitant » (MT) et bénéficiant de 3 à 4 dosages d'HbA1c dans l'année parmi l'ensemble des patients traités par antidiabétiques ayant choisi le médecin comme MT.	54% et ≥ 65%	10 patients	30
	(d) Nombre de patients MT diabétiques de type 2 dont les résultats de dosages d'HbA1c sont < 8,5 % parmi l'ensemble des patients diabétiques de type 2 ayant choisi le médecin comme MT.	80% et ≥ 90%	10 patients	15
	(d) Nombre de patients MT diabétiques de type 2 dont les résultats de dosages d'HbA1c sont < 7,5 % parmi l'ensemble des patients diabétiques de type 2 ayant choisi le médecin comme MT.	60% et ≥ 80%	10 patients	25
	(d) Nombre de patients MT diabétiques de type 2 dont les résultats du dosage de LDL cholestérol est < 1,5 g/l parmi l'ensemble des patients diabétiques de type 2 ayant choisi le médecin comme MT.	80% et ≥ 90%	10 patients	10
	(d) Nombre de patients MT diabétiques de type 2 dont les résultats du dosage de LDL cholestérol est < 1,3 g/l parmi l'ensemble des patients diabétiques de type 2 ayant choisi le médecin comme MT.	65% et ≥ 80%	10 patients	25
	Nombre de patients MT traités par antidiabétiques et bénéficiant d'une consultation ou d'un examen du fond d'œil ou d'une rétinographie dans les deux ans rapporté à l'ensemble des patients MT traités par antidiabétiques.	68% et ≥ 80%	10 patients	35
	Nombre de patients MT traités par antidiabétiques dont l'âge est > 50 ans pour les hommes et > 60 ans pour les femmes et traités par antihypertenseurs et statines parmi l'ensemble des patients MT de mêmes tranches d'âge traités par antidiabétiques et antihypertenseurs.	65% et ≥ 75%	10 patients	35
	Nombre de patients MT diabétiques dont l'âge est > 50 ans pour les hommes et > 60 ans pour les femmes et traités par antihypertenseurs et statines et bénéficiant d'un traitement par aspirine faible dosage ou anticoagulant, rapporté à l'ensemble des patients MT diabétiques de mêmes tranches d'âge traités par antihypertenseurs et statines.	52% et ≥ 65%	10 patients	35
Hypertension artérielle	(d) Nombre de patients MT traités par antihypertenseurs dont la pression artérielle est ≤ à 140/90 mm Hg, rapporté à l'ensemble des patients MT traités par antihypertenseurs.	50% et ≥ 60%	20 patients	40d
Total de points				**250**

Tableau 5.4. Les indicateurs de qualité de la pratique médicale. Thème prévention

Sous thème	*Indicateurs*	*Objectifs intermédiaire et cibles*	*Seuil minimal*	*Points*
Vaccination antigrippale	Nombre de patients MT âgés de 65 ans ou plus vaccinés, rapporté à l'ensemble des patients MT de 65 ans ou plus.	62% et ≥ 75%	20 patients	20
	Nombre de patients MT âgées de 16 à 64 ans en ALD ciblés par la campagne de vaccination et vaccinés, rapporté à l'ensemble des patients MT âgées de 16 à 64 ans en ALD ciblés par la campagne de vaccination.	62% et ≥ 75%	10 patients	20
Cancer du sein	Nombre de patientes MT de 50 à 74 ans participant au dépistage (organisé ou individuel) du cancer du sein rapporté au nombre de femmes MT de 50 à 74 ans.	70% et ≥ 80%	20 patientes	35
Vasodilatateur	Nombre de patients MT âgés de plus de 65 traités, rapporté au nombre des patients MT âgés de plus de 65 ans.	7% et ≤ 5%	20 patients	35
Benzodiazépines à demi-vie longue	Nombre de patients MT âgés de plus de 65 ans traités, rapporté au nombre de patients MT âgés de plus de 65 ans.	11% et ≤ 5%	20 patients	35
Durée de traitement par benzodiazépines	Nombre de patients MT ayant débuté un traitement par BZD à partir de 01/01/2012 et d'une durée supérieure à 12 semaines, rapporté au nombre de patients MT ayant débuté un traitement par BZD sur la même période.	13% et ≤ 12%	5 patients	35
Cancer du col de l'utérus	Nombre de patientes MT de 25 à 65 ans ayant bénéficié d'un frottis au cours des 3 dernières années, rapporté à l'ensemble des patientes MT de 25 à 65 ans.	65% et ≥ 80%	20 patientes	35
Antibiothérapie	Nombre de traitements par antibiotiques pour les patients MT de 16 à 65 ans et hors ALD rapporté au nombre de patients MT de 16 à 65 ans et hors ALD.	40 et ≤ 37	20 patientes	35
Total de points				**250**

Tableau 5.5. Les indicateurs de qualité de la pratique médicale. Thème efficience

Sous thème	*Indicateurs*	*Objectifs intermédiaire et cibles*	*Seuil minimal*	*Points*
Antibiotiques	Prescription (en nombre de boîtes) d'antibiotiques dans le répertoire des génériques / l'ensemble des antibiotiques prescrits (en nombre de boîtes).	85 % et ≥ 90 %	40 boites	60
Inhibiteurs de la pompe à protons	Prescription (en nombre de boîtes) des IPP dans le répertoire des génériques / l'ensemble des IPP prescrits (en nombre de boîtes).	68 % et ≥ 85 %	35 boites	60
Statines	Prescription (en nombre de boîtes) des statines dans le répertoire des génériques / l'ensemble des statines prescrites (en nombre de boîtes).	46 % et ≥ 70 %	30 boites	60
Antihypertenseurs	Prescription (en nb de boîtes) des antihypertenseurs dans le répertoire des génériques / l'ensemble des antihypertenseurs prescrits (en nombre de boîtes).	64 % et ≥ 65 %	30 boites	55
Antidépresseurs	Prescription (en nb de boîtes) des antidépresseurs dans le répertoire des génériques / l'ensemble des antidépresseurs prescrits (en nombre de boîtes).	75 % et ≥ 80 %	30 boites	55
IEC	Prescription (en nb de boîtes) d'IEC / l'ensemble des IEC + sartans prescrits (en nombre de boîtes).	45 % et ≥ 65 %	50 boites	55
Aspirine	Nombre de patients MT traités par aspirine à faible dosage, rapporté à l'ensemble des patients MT traités par antiagrégants plaquettaires.	84 % et ≥ 85 %	10 patients	55
Total de points				**400**

Le tableau 5.6. synthétise le nombre de points maximal auquel le médecin à patientèle de référence moyenne peut prétendre.

Tableau 5.6. Synthèse des points ROSP

Thème	*Points*
Organisation du cabinet	400
Suivi des pathologies chroniques	250
Prévention et santé publique	250
Efficience	400
Total de points	1 300

La question restant à poser est alors de savoir de ce que disent ces indicateurs de l'activité médicale. Que veut dire « performance » dans le ROSP français ?

1.3.2. Des indicateurs de performance ?

Afin de mener la discussion sur le sens que prennent les indicateurs, nous proposons de repartir sur les différentes propositions qu'il est possible d'établir à partir d'une hausse de 5 % de l'indicateur statistique :

Proposition 1 : « le taux de réalisation global s'est amélioré de 5 % ». On parle ici de l'indicateur statistique t_i tel qu'il est donné plus haut.

Proposition 2 : « la rémunération à la performance des médecins s'est améliorée de 5 % ». On parle ici de la rémunération à la performance du médecin Y_i telle qu'elle est définie plus haut.

Proposition 3 : « la performance des médecins s'est améliorée de 5 % ». Lorsque l'on dit que la performance des médecins s'est améliorée, on porte un jugement de valeur sur l'activité individuelle du médecin qui se serait amélioré. Il est performant parce qu'il s'est conformé à ce que l'Assurance maladie lui demande de faire. Pourtant, ce n'est pas tant l'indicateur t_i qui est en cause mais le jugement porté sur l'activité du médecin dans son ensemble. Tout se passe comme si l'évaluation portait sur un ensemble P comprenant t_i mais aussi plus grand que lui. Il s'agit en quelque sorte de mesurer le mérite individuel du médecin.

Proposition 4 : « la qualité des soins s'est améliorée de 5 % ». Ici, le jugement ne porte plus tant sur l'activité du médecin – et donc sa contribution individuelle – que sur le bénéfice que peuvent attendre les patients dans l'ensemble du système de santé. La question qui se pose est celle de l'intérêt de ce qui est demandé au médecin pour la santé du patient. En ce sens, cette dernière proposition n'est pas si éloignée de la suivante : « l'état de santé s'est amélioré de 5 % ».

A partir de cette brève cartographie des significations possibles de l'indicateur, il est possible d'en dresser une lecture critique. Pour cela il suffit d'analyser les différentes dimensions de la critique comme autant d'erreur d'identification des propositions. On peut synthétiser la critique dans le tableau suivant (Tableau 5.7) :

Tableau 5.7. Type d'indentification et critiques des indicateurs de performance

Propositions	**P1 : taux de réalisation global (t_i)**	**P2 : rémunération à la performance (Y_i)**	**P3 : performance du médecin**
P2 : rémunération à la performance (Y_i)	Critique gagnants – performants (effet taille de la patientèle)		
P3 : performance du médecin	Critique multitâche Critique de l'inestimable Critique du codage Critique de l'hétérogénéité de la patientèle	Critique multitâche Critique de l'inestimable Critique du codage Critique de l'hétérogénéité de la patientèle	
P4 : qualité des soins apportée au patient	Critique clinique Critique organisationnelle	Critique clinique Critique organisationnelle	Critique de la production individuelle de la qualité

Présentons les huit types de critiques possibles[71] :

(i) Critique gagnants – performants (identification P1/P2) : ceux qui ont le plus de points (t_i) ne sont pas toujours récompensés par de meilleures rémunérations (Y_i) du fait de l'effet taille de patientèle. En pondérant le calcul de la rémunération à la performance par un coefficient lié à la taille de la patientèle, le dispositif de ROSP introduit une discontinuité

[71] Certaines critiques ont déjà été présentées dans le chapitre précédent, dans la section portant sur les effets pervers du P4P médicale.

entre les gagnants et les performants. Comme le confirmera notre étude empirique, les performants ne sont pas toujours les gagnants.

(ii) Critique multitâche (identification P1/P3 et P2/P3) : en focalisant l'attention sur certains indicateurs plutôt que d'autres, certaines tâches ne sont plus – ou plus aussi bien – réalisées. La santé publique s'intéresse à certains types de pathologies mais pas à d'autres : tabagisme, l'obésité, alcoolisme, etc. Cette sélection des tâches peut également se décliner en sélection des patients. Par exemple, étant donné que les indicateurs portent sur les patients ayant un médecin traitant, tous les patients âgés en dessous de 16 ans ne sont pas comptabilisés. Soigner la jeunesse n'est pas rentable.

(iii) Critique de l'inestimable (identification P1/P3 et P2/P3) : ce qui n'est pas mesurable, n'est pas mesuré donc ni valorisé et ni récompensé. C'est ici une variation particulière de la sélection des tâches dans une logique de *care*. Ce qui est exclu de l'activité est ce qui ne peut pas être mesuré tout en étant indispensable : l'attention, l'écoute, le respect, la compassion, etc. que l'on peut résumer par l'appellation « inestimable » (Chanial, 2010).

(iv) Critique du codage (identification P1/P3 et P2/P3) : ce n'est pas parce qu'un indicateur est renseigné qu'il l'est en raison de l'activité du médecin. Dans le cas du ROSP, il faut savoir que les indicateurs de performance d'un médecin donné ne sont pas tous dépendants de ses propres décisions. Pour les indicateurs de suivi des pathologies chroniques, de prévention et d'efficience, les actes effectués par d'autres médecins… comptent pour le médecin traitant. Ainsi, si une patiente pratique un dépistage du cancer du sein sans que son médecin traitant ne lui en fasse la demande, la performance est tout de même attribuée à son médecin traitant. Il n'a rien fait mais il voit son indicateur s'améliorer. Pour les indicateurs d'optimisation des prescriptions, le médecin doit prescrire dans le répertoire de génériques et pas directement des génériques. En fait, si le médecin prescrit un non-générique dans le répertoire, il est performant – alors que c'est le pharmacien (incité à la performance pour cela) qui va substituer le non-générique en générique.

En outre, lorsque l'indicateur est déclaratif, rien ne garantit que la déclaration soit conforme à l'observation. Pour l'indicateur de dosage de Hb1ac, c'est le médecin qui déclare le taux de son patient. Ce fait est paradoxal avec la justification théorique *mainstream* du P4P : si le P4P est nécessaire à cause de l'opportunisme naturel du médecin, pourquoi lui donner la liberté d'exercer son opportunisme sur des indicateurs déclaratifs ?

(v) Critique de l'hétérogénéité de la patientèle (identification P1/P3 et P2/P3) : il est possible que la composition de la patientèle influe sur la probabilité d'avoir un bon score à un indicateur (par un effet d'éducation à la santé, par exemple). Il est également possible que la patientèle d'un médecin soit plus malade que d'autres (selon le secteur géographique, par exemple). Ainsi, pour certains indicateurs, comme par exemple le dépistage du cancer du sein, si le médecin peut prescrire une mammographie, il n'est pas acquis que la patiente en question la réalise. Pour les indicateurs de diabète, à activité médicale équivalente, il n'est pas difficile d'imaginer que certains patients seront plus enclins à suivre les conseils des médecins que d'autres – ce qui aura notamment une répercussion sur les taux de Hb1ac attribués aux médecins et donc sur leur performance.

(vi) Critique clinique (identification P1/P4 et P2/P4) : les objectifs quantifiés par l'assurance maladie ne font pas nécessairement l'objet d'un consensus du monde médical. L'application indifférenciée de la norme peut s'avérer contreproductive pour la qualité des soins. À ce titre, il est possible de discuter ici de la critique des indicateurs de performance du ROSP réalisée par la revue indépendante *Prescrire* (2013a, b, c, d, e, f). Dans différents numéros publiés en 2013, la revue s'est prêtée au jeu de l'évaluation des critères d'évaluation. Sur les 26 indicateurs du ROSP, seulement 9 d'entre eux sont considérés par *Prescrire* comme pertinents pour la qualité des soins (2013f).

D'autres indicateurs ne sont que partiellement pertinents parce que leur application uniforme pose un problème pour certains types de patients. Par exemple, l'objectif de 80% des patients diabétiques de type 2 dont les dosages de Hb1ac sont inférieurs à 7.5% est cohérent en toute généralité mais, pour certains types de patients, il ne serait pas pertinent. Pour les diabétiques de type 2 dont l'espérance de vie est inférieure à 15 ans, sans complication microangiopathique sévère, le dosage de Hb1ac n'est pas un problème majeur. Pour les patients dont l'espérance de vie est de 5 ans, un dosage satisfaisant est compris entre 8 et 9% (*Prescrire*, 2013b).

Pour d'autres indicateurs encore, les seuils choisis ne sont pas pertinents bien que l'objectif soit louable. Par exemple, pour les patients diabétiques de type 2, les seuils de dosage de LDL-cholestérol inférieur à 1.3 et 1.5 g/l seraient trop élevés et mériteraient d'être placés à 1.15 g/l (*Prescrire*, 2013b). Pour l'objectif de prescription de 5% aux patients âgés de plus de 65 ans traités par vasodilatateurs, étant donné le manque d'efficacité de ces médicaments et leur nocivité, le seuil adéquaté serait de 0% (aucune prescription de ce type de traitement).

Enfin, pour plusieurs indicateurs, l'objectif est totalement « inacceptable ». C'est notamment le cas pour l'objectif de dépistage du cancer du sein. Selon la revue médicale Prescrire, les mammographies n'ont pas fait la preuve de leur efficacité. Parfois le dépistage détecte des tumeurs non géantes qui sont traitées sans raisons, et parfois même, les irradiations répétées causée par la mammographie génèrent des cancers du sein. Cet objectif est en contradiction avec l'information équilibrée des femmes. Les limites certaines et les effets pervers du dépistage seraient établis – contrairement à l'efficacité du dépistage. Dans ces conditions, c'est l'information des patientes qui doit être privilégiée (*Prescrire*, 2013d).

(vii) Critique organisationnelle (identification P1/P4 et P2/P4) : les objectifs d'organisation du cabinet et de qualité du service sont de type binaire. La plus grande partie d'entre eux ne nécessite pas d'effort continu (par exemple, une fois que les horaires sont affichés sur le site internet ameli.fr, il n'y a plus d'effort à renouveler). Ces indicateurs de performance ne font que cacher une aide à l'investissement dans les technologies de l'information et de la communication. En ce qui concerne les objectifs de télétransmission, il est même possible de détecter un effet pervers potentiel : étant donné la difficulté de télétransmettre à partir du domicile des patients, l'incitation à la télétransmission est aussi une incitation à éviter les visites à domicile (*Prescrire*, 2013g).

(viii) Critique de la production individuelle de la qualité (identification P3/P4) : tout se passe comme si le médecin était le seul responsable de la qualité des soins. Or, la qualité des soins est une construction sociale (environnement propice, système de soin développé, état des connaissances, etc.). Il ne semble pas raisonnable de rendre responsable de la qualité de soins uniquement le médecin – aussi malveillant soit-il. Dans le cas du diabète par exemple, des politiques agressives visant à réduire les matières grasses, sucrées et salées dans les aliments auraient sans doute un effet macroéconomique plus sensible sur l'obésité et le diabète que l'action isolée en bout de chaîne des médecins. On peut reprocher au garagiste d'avoir mal fait son travail mais on ne peut lui reprocher un défaut de fabrication du véhicule.

La définition de la performance médicale n'est pas si simple et le dispositif de ROSP est sujet à de nombreuses critiques – tout comme ses variantes internationales[72]. Après lecture des différents types de critiques que nous venons d'établir, la proposition selon laquelle « l'augmentation de la performance des médecins revient à améliorer l'état du monde[73] »,

[72] Cf. Chapitre 4.

[73] Au sens de Salais (2010), voir *supra*.

devient suspecte. D'ailleurs, nous ne sommes même plus sûrs que cette augmentation de la performance soit le produit exclusif de la volonté du médecin. Pour appuyer ces premières remarques, nous devons désormais passer à l'étude empirique.

2 Une analyse statistique du ROSP en Ile de France : les performants ne sont pas forcément les gagnants

Après la discussion générale sur le ROSP, nous allons désormais entreprendre une étude exploratoire sur les médecins traitants généralistes d'Ile de France. La première section présente les données généreusement mises à disposition par le Service médical de l'assurance maladie Ile-de-France (2.1.). La deuxième section analyse le profil des médecins étudiés et de leur patientèle (2.2.). La troisième section présente les résultats ROSP en termes de taux de réalisation et de rémunération à la performance (2.3.). La quatrième section présente les résultats des médecins généralistes selon les critères de gagnants et de performants (2.4.). Enfin, nous montrons dans la dernière section qu'en Ile-de-France, les médecins performants ne sont pas forcément gagnants (2.5.)

2.1. Les données : les bases du Service médical de l'assurance maladie d'Ile-de-France

Les données à disposition pour notre étude portent sur les médecins exerçant en Ile-de-France en 2012, c'est-à-dire la première année de mise en service du ROSP. Du début à la fin de l'étude les données ont été en la possession et traitées par le Service Médical d'assurance maladie d'Ile-de-France (SMAMIF)[74].

Le système d'information de l'Assurance Maladie a pour origine le système de production des services administratifs et médicaux. Il permet d'accéder à des données sur le fonctionnement du système de soins. On distingue cependant plusieurs « outils d'extraction » des informations des assurés sociaux ou plusieurs « bases de données » source n'ayant pas toujours le même périmètre d'information, soit parce que les niveaux d'accès sont différents (local, régional, national) soit parce que l'alimentation des données dans les outils est

[74] Les données médicales sont très sensibles et leur utilisation doit passer par une demande d'autorisation à la Commission nationale de l'informatique et des libertés. Afin de gagner un temps précieux et parce que la SMAMIF dispose de statisticien(ne)s compétent(e)s, nous avons collectivement décidé de travailler de cette manière.

différente (un ou plusieurs régimes), soit parce que le traitement de l'information est différent (régulations des remboursements, traitement de la qualité des flux de données). Trois sources ont été utilisées et articulées pour notre étude : ERASME, SNIIRAM et la ROSP-CNAMTS.

La base ERASME – Extraction, Recherches et Analyse pour un Suivi Médico-Économique – est constituée par les informations saisies au moment de la liquidation de l'ensemble des actes et prestations des Caisses primaires d'assurance maladie et des mutuelles gérant le Régime général (RG). Elle contient des données détaillées sur la consommation de soins donnant lieu à un remboursement de l'Assurance maladie (biais éventuel du fait de la non exhaustivité des soins), des bénéficiaires du Régime général, et des données administratives relatives aux professionnels de santé et aux bénéficiaires.

La base ERASME contient les prestations donnant lieu à un codage comme les médicaments, la biologie, les actes médicaux (Classification commune des actes médicaux), la Liste des produits et prestations et les séjours en établissements privés (Groupes homogènes de séjours). Il s'agit aussi de remboursements de pharmacie délivrés en médecine de ville (biais éventuel quand les soins ont été réalisés à l'hôpital). Le système est alimenté quotidiennement et les données sont conservées pendant une période de 24 mois glissants[75].

La base SNIIRAM – Système National d'Information Inter-Régimes de l'Assurance Maladie – contient la plupart des informations d'Erasme, enrichies de celles des autres régimes que le régime général (notamment le Régime social des indépendant et la Mutuelle social agricole) mais aussi des factures individuelles émises par les établissements de santé – toujours pour l'ensemble des régimes d'assurance maladie et les données informationnelles médicalisées de ces mêmes établissements.

Tandis que les fichiers Caisse nationale d'assurance maladie des travailleurs salariés (Cnamts) concernent un périmètre plus large (France entière), les données extraites de nos interrogations régionales sont plus restreintes (Région Ile de France). Ainsi pour notre étude, les informations relatives aux indicateurs ROSP sont issues des fichiers Cnamts (les numérateurs, dénominateurs et indicateurs ratios ROSP). Les informations relatives à la patientèle du médecin traitant sont issues d'Erasme régional (la patientèle ayant déclaré le médecin Médecin traitant (MT) du Régime général (RG), la part de patients MT RG en Affection longue durée (ALD), la part de patients MT RG Couverture maladie universelle

[75] A chaque nouveau mois enregistré, le plus ancien est supprimé. Par exemple, au 1er février 2014 les données du 1er janvier 2014 sont enregistrées sur la base de données et celle du 1er janvier 2012 sont supprimées.

complémentaire (CMUC). Les informations relatives aux honoraires du médecin traitant sont issues du Sniiram (honoraires issues de la patientèle MT RG, taux de dépassement, honoraires par patient MT RG). Les calculs réalisés à partir d'un numérateur d'une source et d'un dénominateur d'une autre amènent à des indicateurs approximatifs qui permettent toutefois d'approcher une information, à défaut de l'avoir vraiment. C'est le cas des part estimées de patients sous Antiagrégants plaquettaires (APP), part de diabétiques, part de femmes de 50-74 ans, part de femmes de 25-65 ans, part de plus de 65 ans, part de 16-65 ans sans ALD, etc.

La base de données finalement utilisée a été construite en croisant ces trois bases. Les données concernant les patients ne tiennent compte que de ceux affiliés au Régime général qui couvre 86% de la population INSEE. Initialement le fichier croisé comportait 17 910 observations pour lesquelles nous disposions des informations ROSP, des caractéristiques des médecins et des caractéristiques des patients.

Dans le but de faciliter l'interprétation des résultats, nous avons choisi de ne traiter que les médecins traitants généralistes (N=9 080). Cette option se justifie par le fait que la grande majorité des médecins traitants – à qui s'adresse le ROSP[76] – sont des médecins généralistes. Enfin, nous avons sélectionné dans cette base un sous-échantillon de médecins éligibles (N=7440). Autrement dit, nous avons exclus les médecins qui n'ont pas atteint dans un indicateur au moins le seuil minimal de patientèle ou de prescription requis pour que soit calculé l'indicateur.

Concernant les indicateurs disponibles, la base de données fournie par l'assurance maladie comporte les résultats des médecins sur les indicateurs de qualité de la pratique médicale sur les thèmes de suivi des pathologies chroniques, de prévention et d'efficience (mis à part les indicateurs déclaratifs). En conséquence, le nombre de points sur lequel nous allons étudier les résultats des médecins est de 785 et non de 1 300.

[76] La convention signée en 2011 ne prévoit pas de ROSP spécifique pour les médecins spécialistes. Ces deniers ne sont concernés que par les indicateurs d'organisation du cabinet et de qualité du service. Depuis, de multiples avenants à la convention sont venus élargir le ROSP à certaines spécialités.

2.2. Le profil moyen du médecin traitant généraliste et de sa patientèle

Le tableau 5.8. fournit de premières statistiques descriptives sur le profil des médecins traitants généralistes d'Ile de France en 2012. On voit qu'en moyenne chaque médecin[77] entretient une patientèle l'ayant déclaré comme médecin traitant (MT) d'environ 759 patients. Si l'on considère que ces données ne portent que sur les patients du régime général, la patientèle de référence telle que définie dans la convention médicale – 800 – correspond à peu de chose près à ce chiffre. La médiane est quant à elle plus basse et se situe à 729 patients. Mais, en considérant qu'il manque approximativement 10 % des patients affiliés à d'autres régimes, l'hypothèse de quasi équivalence de la patientèle de référence et de la patientèle médiane de notre étude n'est pas dénuée de fondement.

Tableau 5.8. Profil des médecins traitants généralistes en 2012

Libellé	N	Moyenne	Médiane	Minimum	Maximum
Nb patients RG l'ayant déclaré MT	7440	758.77	729	33	3207
Nb patients MT avec prestations 2012	7440	563.40	546	30	2277
Honoraires annuels MT RG (euros)	7440	76304	71593,76	1836	333643,74
Honoraires annuels RG (euros)	7440	136088	125896	2480	699531
Montant Dépassement (euros)	7440	12467	560	0	487828
Taux dépassement 2012	7440	0.0923	0.0049	0	0.8645
Nb patients CMUC MT (RG)	7440	34.90	18	0	634
Nb patients ALD MT (RG)	7440	149.93	140	0	873
Part d'ALD MT (RG)	7440	0.1951	0.1910	0	0.9444
Part de CMUC MT (RG)	7440	0.0424	0.0271	0	0.4349
Age du médecin	7439	55.52	57	30	91
Ancienneté installation	7440	24.72	26.00	2.09	64.05
Age moyen patients MT (RG)	7440	48.38	48.13	31.34	82.41

Par contre, nous observons une forte étendue du nombre de patients ayant déclaré leur médecin comme MT. Alors qu'en moyenne et en médiane le coefficient de patientèle est égale à 1, pour le médecin disposant de la patientèle maximum, celui-ci est de $\frac{3207}{800} \approx 4.01$. Ce qui signifie que pour ce médecin, le calcul des indicateurs ne se fait pas sur 900[78] points

[77] Dans la suite du texte, sauf indication contraire, nous identifions abusivement la catégorie médecin à la catégorie médecin traitant généraliste d'Ile de France en 2012. Il s'agit simplement de faciliter l'écriture et la lecture des résultats. En conclusion, nous reviendrons à une stricte séparation des deux catégories afin de préciser les résultats.

[78] Les 900 points valent hors indicateurs d'organisation du cabinet, hors médecins récemment installés et indicateurs déclaratifs compris (ces derniers n'étant pas disponibles dans notre base de données).

maximum mais $900 * 4.01 = 3609$ points (pour un taux de réalisation de 100 %). Si l'on tient le même raisonnement pour le médecin avec la patientèle MT minimale (patientèle MT = 33), le coefficient est d'environ 0.04 et le nombre de points maximal de 36 !

Il en découle une première intuition : le calcul de la prime est très fortement lié au nombre de patients suivis par le même MT. Il suffit d'étudier la fonction de rémunération pour le constater. En effet, un médecin qui obtient un taux de réalisation de son indicateur relativement bas peut compenser cette faiblesse assez facilement par une forte patientèle. Par exemple, le médecin à la patientèle maximum (patientèle MT = 3207) est éligible à la rémunération maximale d'un médecin à la patientèle de référence lorsqu'il obtient un taux de réalisation de 24.94 %[79] ! Il faut donc s'attendre à ce que la taille de la patientèle (l_i) ait un impact sur la rémunération (Y_i) tel que les médecins fortement rémunérés ne sont pas nécessairement les médecins les plus performants – au sens restrictif de l'obtention du meilleur taux de réalisation (t_i). C'est cette intuition que nous allons tester.

En ce qui concerne le calcul des indicateurs (et non celui de la patientèle MT), il faut noter que le nombre de patients MT consommants et fidèles[80] est sensiblement plus faible que le nombre de patients déclarés MT. Cela implique que le coefficient de patientèle MT prend en compte un nombre de patient plus élevé que le nombre de patient utilisé pour le calcul des indicateurs. En conséquence, même un patient qui ne consulte pas son médecin au cours de l'année permet d'augmenter la rémunération à la performance du médecin. Encore une fois, il semble important de soigneusement distinguer la performance comme concept abstrait et arbitraire de la rémunération à la performance. Le médecin peut en effet recevoir une prime à la performance… pour un patient qui ne le consulte pas. En moyenne, ce sont ainsi $759 - 563 = 196$ patients qui rapportent au médecin une prime à la performance sans consultation. Il s'agit d'une augmentation moyenne de 0.245 du coefficient lié à la patientèle de référence. En moyenne, le coefficient de patientèle est augmenté de 0,245 représentant des patients sans

[79] Lorsque le médecin a la patientèle de référence obtient un taux de réalisation de 100 %, il ne peut prétendre à plus de 900 points.

$$\frac{800}{800} * 900 * 100\% = 900$$

D'où x, le taux de réalisation pour lequel la rémunération maximum du médecin à patientèle de référence est égale à la rémunération du médecin à patientèle maximale :

$$\frac{800}{800} * 900 * 100\% = \frac{3207}{800} * 900 * x\% \; ; \; x \approx 24.94$$

[80] La patientèle consommante est celle qui a réalisé au moins une consultation chez son médecin traitant au cours de l'année de calcul de l'indicateur. La patientèle fidèle est la patientèle qui n'a pas changé de médecin traitant au cours de l'année de calcul de l'indicateur.

consultation. Pour un médecin ayant une patientèle de référence (coefficient = 1), cela signifie qu'il ne consulte effectivement que 75.5 % des patients pour lesquels il est rémunéré.

Concernant les honoraires, en moyenne, les médecins ont perçu 76 304 € dépassement inclus soit, en moyenne, 135 € par patient consommant. Sans surprise, l'étendue des rémunérations est forte puisque le médecin le moins bien rémunéré percevait 1836 € annuel tandis que le médecin le plus rémunéré percevait près de 333 644 € annuels pour ses patients MT. Cela semble indiquer deux choses au moins : les médecins les moins rémunérés ont probablement une activité complémentaire (comme le salariat par exemple) et l'étendue des rémunérations doit s'expliquer en grande partie par la possibilité de pratiquer des dépassements d'honoraires.

L'âge moyen du médecin est de 55.52 ans avec un minimum à 30 ans et un maximum à 91 ans ! Les médecins sont installés en moyenne depuis 25 ans avec de nouveaux installés (2 ans) et des médecins historiques (64 ans d'ancienneté !). L'âge moyen de la patientèle est de 48 ans. Celle-ci comporte en moyenne pour chaque médecin 4.24 % de patient titulaires de la CMUC et 19.51 % titulaires du dispositif ALD. Ces deux variables nous seront par la suite particulièrement importantes dans la mesure où le type de patientèle peut avoir un impact sur le taux de réalisation globale et la rémunération à la performance. Or, le dispositif CMUC dépend d'un seuil minimal de ressources fixé à environ 720 euros mensuels – c'est un indicateur de pauvreté – et le dispositif ALD ne dépend que de l'état de santé du patient – c'est un indicateur d'état de santé. On observe ainsi que pour certains médecins la part de patient CMUC dépasse les 40 % de la patientèle totale soit 10 fois plus que la moyenne. Or, si l'on suppose que les patients les plus pauvres sont les moins bien éduqués à la santé, il est probable qu'ils influent sur les résultats du médecin. Par exemple, pour l'indicateur du cancer du sein, le taux de réalisation du médecin dépend de la volonté et des possibilités que la patiente a de se faire dépister.

2.3. Les résultats ROSP : taux de réalisation et rémunération à la performance

Le tableau 5.9. recense les résultats ROSP des médecins. Y sont reportés les points obtenus par thème (suivi des pathologies chroniques, prévention et efficience de la prescription) et la rémunération totale qui en résulte. À chaque fois nous reportons le nombre de médecin pour lesquels au moins un indicateur du thème a été calculé. Nous observons que seuls 6630 médecins sur 7440 ont suffisamment de patients atteints de pathologies chroniques

pour que soit calculé leur indicateur. Ainsi, 810 médecins n'ont pas suffisamment de patients atteints de diabète et/ou d'hypertension[81] pour que l'assurance maladie juge de leur performance dans le thème suivi des pathologies chroniques. Cela ne veut pas dire qu'ils ne sont pas performants mais leur niveau d'activité est trop faible pour qu'ils soient considérés intéressants du point de vue de la santé publique. En conséquence, quel que soit leur type d'activité – conforme ou non aux indicateurs ROSP – les médecins ne reçoivent pas de prime à la performance pour ces indicateurs.

Tableau 5.9. Résultats ROSP des médecins traitants généralistes par thème en 2012

Points et rémunération 2012	N	Moyenne	Médiane	Q3	Min	Max	Max théorique	Note moyenne
Points pathologie chronique	6630	39.79	35	58.2	0	135	135	29.47 %
Points Prévention	7422	97.64	97.5	125	0	219.7	250	39.06 %
Points Efficience de la prescription	7391	229.88	231.5	276.4	0	400	400	57.47 %
Points au total	7440	361.23	362.55	428.3	0	742.4	785	45.99 %
Rémunération 2012* (euros)	7440	2464.01	2348.46	3282.02	0	14233.25		

**Majoration pour les nouveaux installés en 2010 2011 et 2012 (15% 1ère année, 10% 2nde année et 5% 3ème année)*

En moyenne les médecins ont réussi à obtenir 361 points sur 785 mis en jeu, soit environ 45.99%. Les résultats sont assez faibles pour les indicateurs de prévention avec 39.06 % de réussite et très faibles pour les indicateurs de suivi des pathologies chroniques avec 29.47 % du total. Ce n'est que grâce aux indicateurs d'efficience de la prescription que le taux de réalisation global moyen passe la barre des 45 %. En effet, étant donnée la forte part relative de ces indicateurs dans l'ensemble (400 points sur 785, soit 50.96 %) et la plus grande facilité à les réaliser (57.47 % de taux de réalisation moyen), il est assez logique que la moyenne générale augmente.

On peut ainsi dire que la performance des médecins est tirée à la hausse par leurs bonnes pratiques de prescription. À ce titre, on peut se demander s'il ne s'agit pas d'une performance en trompe l'œil. Si l'on considère que derrière la notion abstraite de performance

[81] Dans ce thème, le seuil minimal le plus faible est de 30 patients (hors indicateurs déclaratifs ici non pris en compte).

il y a l'idée de soins de meilleure qualité, au sens de suivi des pathologies chroniques et de prévention, et pas nécessairement celle de prescriptions génériques, bien que le taux de réalisation soit proche de la moyenne, on peut légitimement se demander s'il s'agit de performance. Cette interrogation est d'autant plus légitime qu'en ce qui concerne la prescription de génériques, les médecins ne sont pas obligés de les prescrire mais de prescrire dans le répertoire de génériques – le pharmacien substituant la prescription par un générique le cas échant. Par ailleurs, les médecins sont également incités dans le volet « organisation du cabinet » à acquérir les logiciels permettant de prescrire dans le répertoire de génériques. La même action est rémunérée plusieurs fois et la performance est réalisée pour les indicateurs les plus faciles : l'optimisation des prescriptions. Tout se passe comme si un lycéen de terminale scientifique obtenait le bac grâce à l'option sport !

Concernant la rémunération à la performance, en moyenne, les médecins ont gagné 2 464,01 € pour 361.23 points. S'en tenir à ces chiffres peut particulièrement induire en erreur : on pourrait en effet déduire une relation linéaire entre taux de réalisation et performance. Or, il suffit d'analyser le rapport entre la rémunération maximale et le taux de réalisation maximal pour se rendre compte que ce n'est pas le cas. Le médecin ayant obtenu le maximum de points a eu 742.4 points soit environ 2.06 fois plus que le médecin moyen. Pourtant, la rémunération maximale n'est pas de 2.06 fois la rémunération moyenne (5 075.86) : elle est de 14 233.25 €. C'est-à-dire trois fois plus que ce que prévoit la relation linéaire[82].

Le tableau 5.10. renforce le niveau de détail concernant les taux de réalisation des médecins. Le résultat de chaque indicateur est répertorié. Tout comme le montre le tableau 5.9., les médecins sont plus performants dans les indicateurs d'efficience de la prescription. Par exemple, en moyenne, 86,18 % des prescriptions d'IPP sont faites dans le répertoire de génériques. Le contraste est évident avec les prescriptions de génériques pour lesquelles au moins un médecin par indicateur parvient à atteindre le taux de réalisation maximal (100 %).

[82] Posons l'hypothèse que le médecin à la rémunération maximale (14 233.25 €) soit celui qui a le nombre de patients maximum (3 207). En divisant sa rémunération par 7 €, on retombe sur le nombre de points pondérés obtenus : 2033.32. Soit y sont taux de réalisation :

$$2033.32 = \frac{3207}{800} * y\,\% \,;\, y = 56.34$$

Ainsi, en dépit du fait que le médecin à la meilleure rémunération soit relativement plus performant que la moyenne (taux de réalisation supérieur de 22.50 % à la valeur moyenne), l'augmentation de sa rémunération est bien supérieure à celle de son taux de réalisation : elle subit une hausse de 477.65 %. Il semble évident dans ce cas que l'effet le plus important dans la variation de la rémunération est la taille de la patientèle.

L'hypothèse développée plus haut se précise : il semble plus « facile » de réaliser un taux de réalisation global élevé dans le thème efficience des prescriptions que dans les autres.

Tableau 5.10. Résultats ROSP des médecins traitants généralistes par indicateur en 2012

Résultat ratios ROSP du MT généraliste en 2012	N	Moyenne	Médiane	Q3	Min	Max
Diabétiques à 3-4 dosages Hba1c/an	6630	0,4268	0,4204	0,5405	0.00	0,9583
Diabétiques avec suivi ophtalmo. dans les 2ans	6630	0,6043	0,6041	0,6756	0,1428	1.00
Diabétiques sous antihypertenseurs sous statine	6011	0,6153	0,6190	0,7031	0,0769	1.00
Diabétiques sous anti-hypertenseurs	4952	0,5025	0,5	0,6	0.00	1.00
Pat. 50-74 ans avec une mammo. dans les 2ans	7130	0,6445	0,6543	0,7037	0,2236	0,9268
Pat. 25-65 ans avec frottis ans les 3 ans	7370	0,5812	0,5913	0,6538	0,1052	0,8823
Pat. >65 ans traités par vasodilatateurs	6886	0,0481	0,0394	0,0645	0.00	0,6363
Pat. >65 ans traités par benzodiazépines à ½ vie longue	6886	0,1190	0,1162	0,1488	0.00	0,3809
Pat. début traitement BZD durée >12 semaines	7347	0,1336	0,1291	0,1666	0.00	0,625
Prescriptions AB pour 100 patients 16-65ans hors ALD	7405	37,86	35,5437	48,8558	0.00	144,88
Pat. 65ans et + vaccinés contre la grippe saisonnière	6944	0,5256	0,5416	0,6	0.00	0,8611
Pat. 16-64 ans ALD vaccinés pour la grippe saisonnière	4883	0,2959	0,2916	0,3571	0.00	0,7551
Presc. AB dans rép. Génériques	7187	0,8239	0,8410	0,8906	0,2081	1.00
Presc. IPP dans le rép. des génériques	7152	0,8618	0,8902	0,9363	0,1791	1.00
Presc. Statines dans le rép. des génériques	7177	0,5479	0,5540	0,6565	0.00	1.00
Presc. Antihypertenseurs dans le rép. des génériques	7367	0,7184	0,7216	0,7689	0,0526	1.00
Presc. Antidépresseurs dans le rép. des génériques	7132	0,6968	0,7097	0,8072	0.00	1.00
Presc. IEC sur l'ensemble des IEC/Sartans	7183	0,3527	0,3440	0,4304	0.00	1.00
Pat. Sous APP traités par aspirine faible dosage	6502	0,8525	0,8620	0,9074	0,3	1.00

2.4. Des gagnants et des performants

Le tableau 5.11. analyse le profil des médecins performants – c'est-à-dire ceux dont le taux de réalisation global est supérieur au troisième quartile de la distribution des taux de réalisation (Q3 = 428). Ce tableau renseigne le thème où les points ont été obtenus, le montant de la prime, les caractéristiques des médecins et les caractéristiques de la patientèle. 1869 médecins représentant les 25 % de médecins ayant obtenus le plus de points ont eu un taux de réalisation global supérieur à 428 points pour une moyenne de 487.13 points. Ces médecins

ont ainsi perçu une rémunération moyenne de 3 446.11 € contre seulement 2 134.53 € pour les médecins peu performants (ceux en dessous du dernier quartile). Si cette différence de rémunération semble tout à fait normale, la comparaison entre les minima et maxima des deux catégories n'est pas sans surprise. On observe d'abord qu'il est possible d'être performant sans être rémunéré plus que la moyenne des « moins performants ». Ainsi, parmi les médecins performants, celui qui a reçu la plus petite rémunération (228.85 €) a obtenu près de dix fois moins (9.33) que la moyenne des moins performants (2 134.53 €). Inversement, le médecin le mieux payé de la catégorie des peu performants (10 427,9 €) a obtenu une rémunération trois fois plus élevée (3.03) que la moyenne des médecins performants (3 446.11 €). Le contraste est plus grand lorsque l'on compare le mieux rémunéré des médecins non performants (10 427.9 €) au plus mal rémunéré des médecins performants (228.85) : le non performant peut être payé 45.57 fois plus que le performant !

Les médecins performants le sont généralement grâce aux indicateurs d'efficience des prescriptions. En moyenne, ils obtiennent 297.78 points sur 400 possibles, soit 74.45 % de réussite. Or, leur niveau de réalisation global moyen n'est que de 62.05 %. Cette fois-ci ce sont bien les indicateurs de suivi des pathologies chroniques et de prévention qui tirent la performance à la baisse.

En moyenne, le nombre de patients MT et le nombre avec des prestations ne sont pas sensiblement différents entre les deux catégories de médecins. Les performants ont en moyenne 807 patients déclarés MT contre 742 pour les autres. Ils ont par ailleurs, 600 patients consommants et fidèles contre 551 pour les non performants. Même si ces différences sont significatives au risque d'erreur 5% ($p<0.01$) elles sont relativement faibles. La même chose vaut pour le montant des honoraires issus de la patientèle du régime général totale, qu'elle ait ou non déclaré un MT, (en moyenne 132 428 € pour les performants contre 137 317 € pour les autres, $p=0.003$). Par contre, il semble qu'en moyenne les médecins performants font moins de dépassements d'honoraire que les autres (9 293 € contre 13 533 €, $p<0.01$). Le taux de dépassement est d'ailleurs supérieur de 4 points de pourcentage pour les non performants (6.49 % contre 10.15 %, $p<0.01$). La composition de la patientèle (ALD, CMUC), ainsi que l'âge du patient, l'âge du médecin et son ancienneté ne semblent pas très différents d'une catégorie à l'autre. En moyenne, les médecins performants semblent toutefois avoir une patientèle avec un peu moins de patients CMUC et un peu plus de patients ALD que les autres.

Tableau 5.11. Caractéristiques des médecins performants

Profil MT des deux groupes	Performant (>=Q3, 428 points): Oui - n= 1869					Performant (>=Q3, 428 points): Non - n= 5571				
	N	Moyenne	Médiane	Min	Max	N	Moyenne	Médiane	Min	Max
Pathologie chronique	1837	61,28	61,5	0	135.00	4793	31,55	29.00	0.00	131,6
Prévention	1869	129,11	128,7	22,9	219,7	5553	87,04	86,4	0.00	207,2
Efficience	1869	297,78	296,5	135,3	400.00	5522	206,89	211,9	0.00	370,2
Points	1869	487,13	475,4	428.00	742,4	5571	318,99	332.00	0.00	427,9
Rémunération	1869	3446,11	3256,82	228,85	14233,25	5571	2134,53	2046,21	0.00	10427,9
Nb pat. MT	1869	807,49	766	59	2887	5571	742,43	712	33	3207
Nb pat. MT avec prest.	1869	599,75	570	40	2130	5571	551,21	532	30	2277
Hon. MT	1866	77859,3	72224,99	3592.00	333643,74	5558	75781,88	71422,1	1836.00	323274,2
Honoraires 2012	1869	132428	123705	7841	561538	5571	137317	126889	2480	699531
Dépassement	1869	9293	311	0	167682	5571	13533	674	0	487829
Tx dépassement	1869	0,0649	0,0026	0.0000	0,8317	5571	0,1015	0,0059	0.0000	0,8645
Nb pat. CMUC MT	1869	30,85	17.00	0.00	634.00	5571	36,26	19.00	0.00	614.00
Nb pat. ALD MT	1869	161,13	149.00	9.00	873.00	5571	146,17	136.00	0.00	658.00
Part ALD MT	1869	0,2008	0,1967	0,0736	0,7272	5571	0,1932	0,1894	0.0000	0,9444
Part CMUC MT	1869	0,0356	0,0230	0.0000	0,2892	5571	0,0447	0,0289	0.0000	0,4349
Age MT	1868	54,63	55,5	32.0000	90.0000	5571	55,8169	57.0000	30.0000	91.00
Ancienneté installation	1869	24,37	25,36	2,26	58,04	5571	24,84	26,14	2,08	64,04
Age moyen pat. MT	1869	48,62	48,43	35,69	62,62	5571	48,29	48,03	31,34	82,404

Après avoir comparé les médecins selon le critère du taux de réalisation global (performants *versus* non performants) nous pouvons faire le même travail à partir du critère de rémunération. Le tableau 5.12. dresse le profil des médecins gagnants (rémunération supérieure au troisième quartile de la distribution) et des médecins perdants (rémunération inférieure au troisième quartile de la distribution). La ligne de démarcation est la rémunération de 3 282 €. 75 % des médecins ont gagné moins et 25 % des médecins ont gagné plus au titre du ROSP. Une bonne partie des résultats sont symétriques à ceux obtenus précédemment. En moyenne, un médecin gagnant a obtenu 430.53 points. Ce médecin moyen peut donc être considéré comme gagnant et performant (puisqu'il fallait 428 point pour être dans la catégorie performant). Mais, et ce n'est plus une surprise, lorsque l'on s'intéresse au minimum pour être gagnant, on s'aperçoit qu'il suffit d'obtenir 189.8 points pour être gagnant. Pour ce médecin, il suffisait donc d'avoir une patientèle supérieur ou égale à k :

$$189.8 * \frac{k}{800} = \frac{3282}{7}$$

$$k = 1976.22$$

Ainsi, il existe un médecin qui a obtenu 24.18 % des points – alors que la moyenne générale est de 361.23, soit 45.99 % du total – mais qui, ayant une patientèle supérieur ou égale à 1973 patients, s'est hissé dans les 25 % de médecins les mieux rémunérés. Celui-là est un gagnant du ROSP. L'un des gagnants a même réussi à obtenir plus de 3 282 € en obtenant 0 point sur 250 dans le thème « pathologie chronique » et/ou 12.8/250 sur le thème « prévention ». Pour le ROSP comme pour le bac, il n'y a pas de note éliminatoire. À l'inverse un médecin perdant à réussit à obtenir… 714.9 point sur 785, soit un taux de réussite de 91.07 %. La raison de cette catégorisation se trouve encore dans la taille de la patientèle. Il suffit pour cela que ce médecin ait une patientèle inférieure ou égale à n :

$$3282 = 715.9 * 7 * \frac{n}{800}$$

$$n \approx 524.67$$

Tableau 5.12. Profil du médecin gagnant

Profil MT des deux groupes	*Gagnant (>=Q3, 3282 E): Oui - n= 1860*					*Gagnant (>=Q3, 3282 E): Non - n= 5580*				
	N	**Moyenne**	**Médiane**	**Min**	**Max**	**N**	**Moyenne**	**Médiane**	**Min**	**Max**
Pathologie chronique	1856	51.87	48.95	0	135	4774	35.09	31.6	0	135
Prévention	1860	110.34	109.5	12.8	212.5	5562	93.39	93.3	0	219.7
Prescription	1860	268.42	267.3	93.7	400	5531	216.91	219.4	0	400
Total de points	1860	430.53	426.9	189.8	742.4	5580	338.13	341.4	0	714.9
Rémunération	1860	4356.07	4031.9	3282.39	14233.25	5580	1833.32	1932.26	0	3281.65
Nb pat. MT	1860	1192.68	1116	549	3207	5580	614.14373	622	33	2294
Nb pat. MT avec prest.	1860	884.34	834	351	2277	5580	456.42491	461	30	1616
Honoraire 2012	1860	188700.98	173258.6	57847	683394.96	5580	118551	112247.6	2480	699531
Montant Dépassement	1860	11375.94	367.55	0	180113.03	5580	12831.86	650	0	487828.5
Tx dépassement	1860	0.0563	0.002015	0	0.6881	5580	0.1044	0.0067	0	0.8645
Nb pat. CMUC MT	1860	59.58	31	1	634	5580	26.68	15	0	378
Nb patients ALD MT	1860	233.08	218	61	873	5580	122.21	116	0	557
Part d'ALD MT	1860	0.1962	0.1929	0.0491	0.5031	5580	0.1948	0.1905	0	0.9444
Part de CMUC MT	1860	0.0465	0.0280	0.0009	0.3265	5580	0.0410	0.0269	0	0.4349
Age MT	1860	54.89	55	33	77	5579	55.72	57	30	91
Ancienneté installation	1860	24.83	25.35	2.50	45.99	5580	24.68	26.17	2.08	64.04

Evidemment, s'il y a un écart important entre la moyenne de point des gagnants et celle des perdants (430.53 contre 341.4, soit 26.11 % de plus), ce qui est le plus marquant est la différence de taille de patientèle. Un médecin gagnant a en moyenne une patientèle presque 2 fois plus élevée que celle d'un perdant. C'est vrai pour le nombre de patient MT (1192.68 patients contre 614.14, soit un coefficient multiplicateur de 1.94) mais aussi pour le nombre de patients avec prestation (884.34 contre 456.42, soit un coefficient multiplicateur de 1.94).

Si les honoraires sont plus élevés pour les médecins gagnants, c'est tout simplement parce qu'ils consultent plus. Par contre le taux de dépassement est sensiblement plus élevé chez les médecins perdant (10.44 % contre 5.63 %). Le type de patientèle (CMUC ou ALD), l'âge du médecin ou l'ancienneté d'installation ne varient pas entre les deux catégories de médecin.

Le tableau 5.13. s'intéresse plus spécifiquement au profil de la patientèle du médecin gagnant étant donné les indicateurs ROSP. Concernant la plupart des indicateurs, les parts de patients ciblées par le ROSP sont insensibles à la catégorisation gagnant/performant. Par exemple, en moyenne, la part de femmes entre 50 et 74 ans est de 17.18 % pour les gagnants et de 18.57 % pour les perdants. La part moyenne de plus de 65 ans est également stable : 16.90 % pour les gagnants contre 17.91 chez les perdants. Par contre, on observe une plus forte différence de patients sur l'indicateur de patients ALD âgés de 16 à 64 ans. Ils sont 4.61 % chez les gagnants contre 3.93 % chez les perdants. Même s'il y a une différence, elle ne semble pas très importante. Les patientèles paraissent être sensiblement les mêmes quel que soit le statut du médecin.

Tableau 5.13. Profil de la patientèle du médecin gagnant

Profil MT gagnant	*Gagnant (>=Q3, 3282 E): Oui - n= 1860*					*Gagnant (>=Q3, 3282 E): Non - n= 5580*				
	N	**Moyenne**	**Médiane**	**Min**	**Max**	**N**	**Moyenne**	**Médiane**	**Min**	**Max**
Age moyen pat. MT	1860	47.96	47.69	35.69	64.58	5580	48.517	48.32	31.34	82.40
Nb diabétiques	1860	74.63	65.5	8	498	5580	38.28	34	0	219
Part de diabétiques	1860	0.0617	0.0592	0.0064	0.2571	5580	0.0599	0.0572	0	0.7473
Nb diab./antihypertenseurs	1860	44.22	40	1	221	5580	23.39	21	0	130
Part diab./antihypertenseurs	1860	0.0371	0.0359	0.0008	0.1260	5580	0.0368	0.0347	0	0.4421
Nb pat. 50-74 ans	1860	203.17	191	16	529	5580	110.28	106	0	446
Part pat. 50-74 ans	1860	0.1718	0.1714	0.0160	0.5064	5580	0.1857	0.1789	0	0.7266
Nb pat. 25-65 ans	1860	446.32	423.5	78	1229	5580	233.77	230	1	823
Part pat. 25-65 ans	1860	0.3754	0.3705	0.0781	0.6244	5580	0.3875	0.3745	0.0044	1
Nb pat. >65ans	1860	197.53	180	20	993	5580	107.82	96	0	531
Part de +65 ans	1860	0.1690	0.1593	0.0168	0.5510	5580	0.1791	0.1649	0	0.9591
Nb 16-65 ans sans ALD	1860	770.22	727.5	194	2257	5580	394.71	392	1	1432
Part 16-65 ans sans ALD	1860	0.6466	0.6448	0.2639	0.9317	5580	0.6438	0.6422	0.0069	1
Nb 16-64 ans en ALD ciblés	1860	55.85	47	8	459	5580	27.58	24	0	383
Part 16-64 ans en ALD ciblés	1860	0.0461	0.0429	0.0098	0.4599	5580	0.0427	0.0393	0	0.7978
Nb patients sous APP	1860	76.26	70.5	4	327	5580	38.58	34	0	221
Part de patients sous APP	1860	0.0654	0.0613	0.0033	0.2164	5580	0.0610	0.0552	0	0.4019

2.5. Les performants ne sont pas toujours gagnants

Le tableau 5.14. et le graphique 5.1. résument la répartition des médecins entre les différentes catégories que nous avons construites. Sur l'échantillon total de 7 440 médecins, 1 869 sont déclarés performants (leur taux de réalisation global est supérieur ou égal à 428 points sur 785) et 1890 sont déclarés gagnants (leur rémunération est supérieure ou égale à 3 282 €). Si, d'une part, 62.27 % des médecins sont à la fois non performants et perdants et, d'autre part, 12.39 % des médecins sont performants et gagnants, 25.4 % des médecins – soit 1 885 – sont « entre deux » : soit performants mais perdants (pour 12.73 %), soit non performants mais gagnants (pour 12.61 %). L'existence de ces deux dernières catégories semble écorner l'un des objectifs du ROSP, c'est-à-dire la rémunération au mérite. Il est en effet possible d'être performant sans être rémunéré et d'être rémunéré sans être performant. La diagonale sud-ouest/nord-est peut s'interpréter comme la diagonale des « anomalies » du ROSP au regard de la critique d'identification des gagnants et des performants.

Tableau 5.14. Les performants ne sont pas toujours gagnants

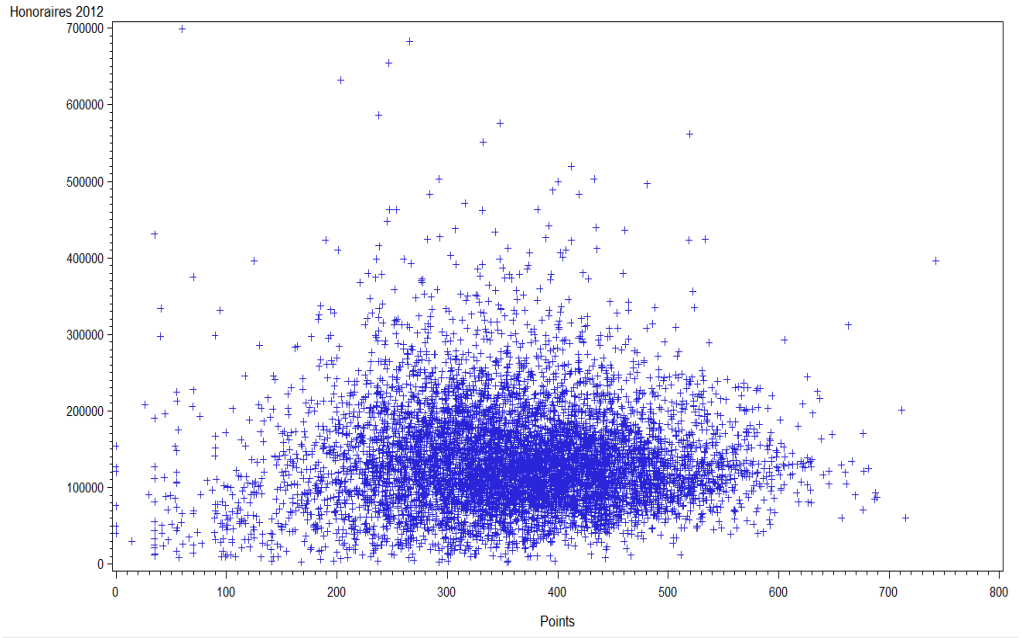
Performants	Gagnants		
	Non	**Oui**	**Total**
Non	4633	938	5571
% Total	62.27	12.61	74.88
% En ligne	83.16	16.84	
% En colonne	83.03	50.43	
Oui	947	922	1869
% Total	12.73	12.39	25.12
% En ligne	50.67	49.33	
% En colonne	16.97	49.57	
Total	5580	1860	7440
	75	25	100

Graphique 5.1. Les performants ne sont pas toujours gagnants

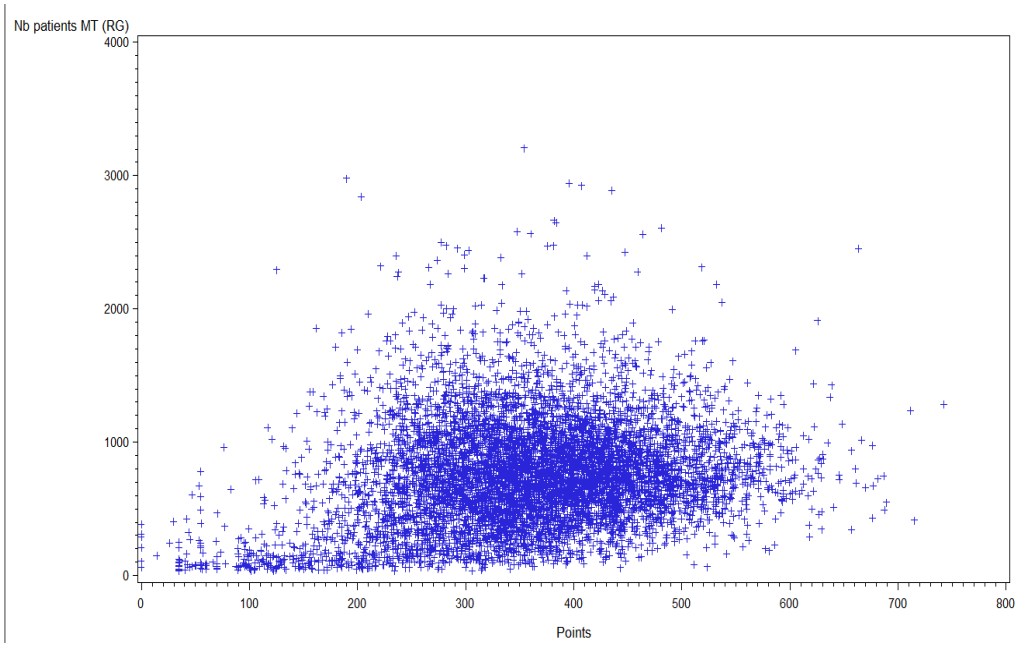
	Les gagnants	
Les performants	922 (12.39 %)	938 (12.61 %)
Les non performants	947 (12.73 %)	4633 (62.27 %)
	Les perdants	

Avant de passer à l'analyse économétrique, il est intéressant de commenter quelques nuages de points permettant de valider plus amplement les premiers enseignements issus des statistiques descriptives. Les deux premiers nuages de points (Graphiques 5.2. et 5.3.) présentent respectivement le croisement des variables honoraires/taux de réalisation et nombre de patients MT RG/taux de réalisation. Dans l'attente d'un traitement statistique plus poussé, on observe l'absence de corrélation claire entre le taux de réalisation et, d'une part, le montant des honoraires et, d'autre part, le nombre de consultations. On peut ainsi questionner la validité de deux hypothèses implicites de la théorie économique. Premièrement, la multiplication du nombre de patients ne semble pas avoir pour incidence une augmentation du taux de réalisation par effet d'apprentissage. On pourrait en effet s'attendre à ce que les médecins soient d'autant plus habiles qu'ils s'occupent d'une patientèle large. Ce n'est pas le cas. Deuxièmement, le prix ne semble pas signaler la qualité. En effet, le graphique 5.3. montre une absence de corrélation manifeste entre le montant des honoraires perçus par les médecins et leur taux de réalisation global. Se pose alors la question de la légitimité de la variation des honoraires : s'agit-il de rémunérer une médecine qui « prend son temps » de meilleure qualité comme le prétendent les défenseurs de la médecine libérale ou bien, s'agit-il simplement d'extraire un revenu plus élevé ?

Graphique 5.2. Nuage de points Honoraires/Taux de réalisation



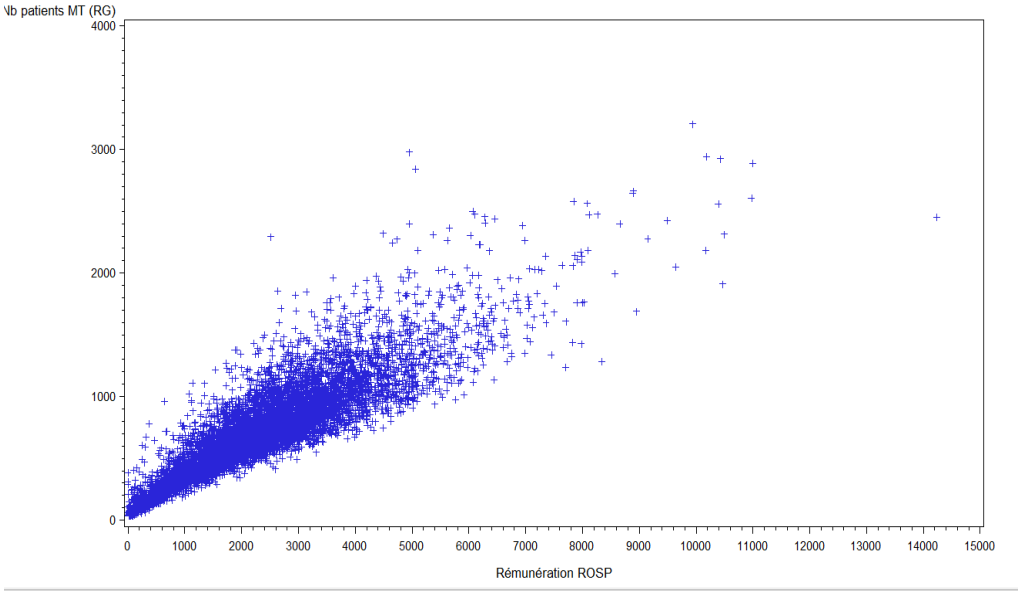
Graphique 5.3. Nuage de points Nombre de patients MT RG/taux de réalisation



Les trois derniers nuages de points (graphique 4, 5 et 6) s'intéressent à la variable de prime ROSP. Le graphique 5.4. montre une relation parfaitement linéaire – quoi qu'en entonnoir inversé – entre le niveau de la prime et le nombre de patient MT RG. Comme nous l'avons déjà suffisamment répété, la taille de la patientèle est un élément essentiel pour comprendre le niveau de la rémunération ROSP. Plus le médecin a de patient, plus de chance

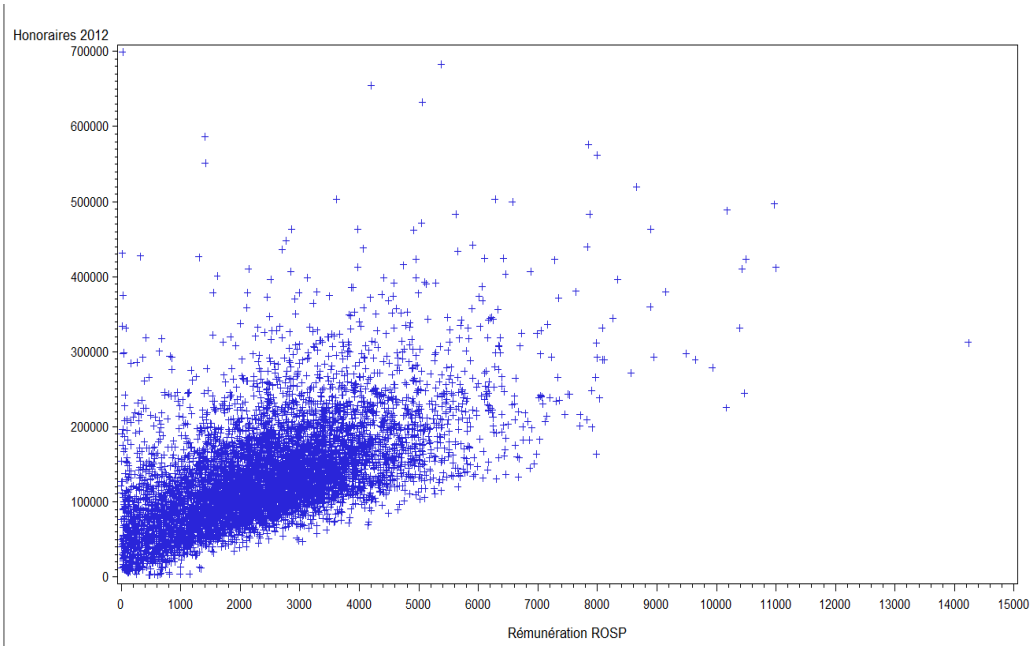
il a de recevoir un fort paiement à la performance. Sa performance, c'est la taille de sa patientèle et donc son pouvoir de marché.

Graphique 5.4. Nuage de points Nombre de patients MT RG/Rémunération ROSP

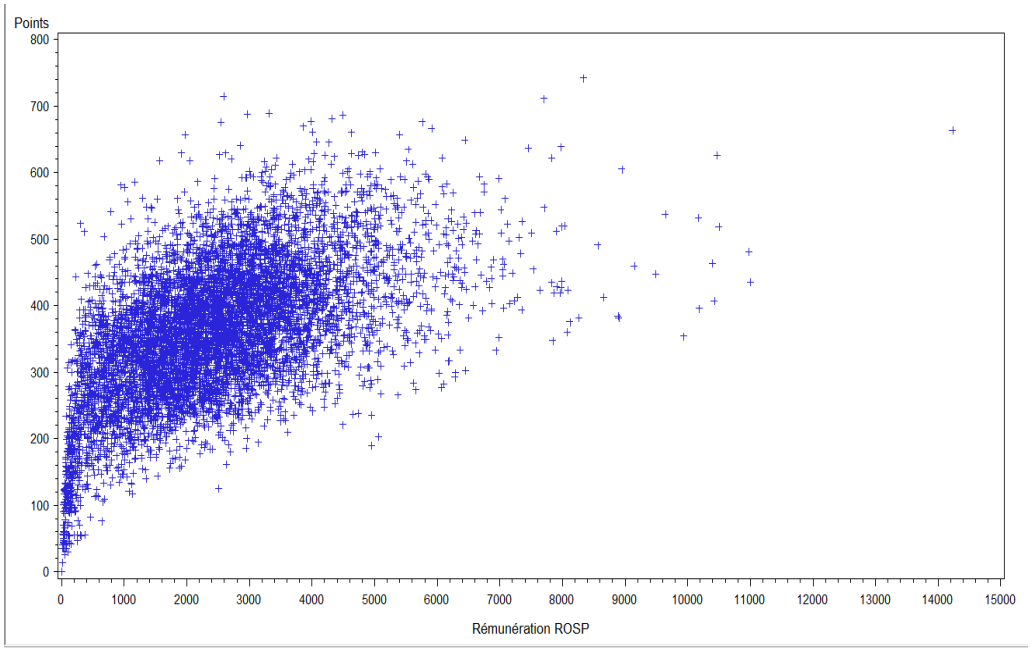


Le graphique 5.5. confirme cette relation mais, en utilisant le montant des honoraires annuels, on s'aperçoit à nouveau de la forte variation des rémunérations à patientèle égale. Or, si à patientèle égale et à taux de réalisation égal, on observe un niveau de rémunération différent, cela confirme l'idée selon laquelle le prix ne signale pas la qualité. Enfin, le graphique 5.6. décrit la relation entre le taux de réalisation global du médecin et sa rémunération à la performance. La relation semble logarithmique. Dans un premier temps – jusqu'à une rémunération d'environ 500 € et un nombre de points de 280 –, le nombre de points et la rémunération à la performance sont parfaitement croissants l'un de l'autre. Mais, dans un second temps, tandis que la rémunération continue de croître relativement rapidement, la vitesse d'augmentation du nombre de points s'essouffle. C'est certainement encore une fois l'effet de la taille de la patientèle. Bien qu'il devienne plus difficile d'améliorer son taux de réalisation global au-dessus d'un certain point (t_i=280 ?), la rémunération continue de croître grâce au nombre de patients.

Graphique 5.5. Nuage de points Honoraires/Rémunération ROSP



Graphique 5.6. Nuage de points Taux de réalisation/Rémunération ROSP



3 La variable cachée de la performance : l'effet de composition de la patientèle

Les deux sections précédentes ont mis en évidence un écart entre la performance mesurée et la performance théorique. Ici, nous réduisons drastiquement l'ampleur de la discussion afin de mener une étude économétrique. Nous testons l'hypothèse selon laquelle le taux de réalisation et la rémunération obtenus dépendent de variables indépendantes de l'implication et de la responsabilité du médecin : le taux de réalisation et la rémunération à la performance dépendraient d'un effet de composition de la patientèle. Nous nous intéressons particulièrement à l'état de santé de la patientèle ainsi qu'au niveau de ses revenus. Nous présentons d'abord la méthode (3.1.), puis les résultats (3.2.) que nous discutons en conclusion (3.3.).

3.1. Méthode

Nous cherchons à expliquer la rémunération du médecin (Y_i) et son taux de réalisation global (t_i). Deux grands types de variables explicatives sont à prendre en compte. D'une part, la rémunération du médecin est déterminée par le taux de réalisation global et la taille de patientèle totale :

$$Y_i = Y_i(t_i, l_i)$$

Cette première relation ne nous apprend rien : il est tautologique de trouver une corrélation entre ces variables et la rémunération à la performance, puisque celles-ci entrent dans la formule de calcul de la rémunération[83]. D'autre part, et de façon plus intéressante, il est probable que le montant de la rémunération du médecin i dépende de variables qui ne sont pas dans la fonction de prime. Les sections précédentes ont mis en évidence qu'il existe un écart entre la performance théorique et la performance réelle. Des éléments étrangers à la responsabilité du médecin et à son implication peuvent avoir un impact sur sa performance réelle. L'hypothèse que nous souhaitons tester ici est que la rémunération à la performance et le taux de réalisation global sont corrélés au profil de la patientèle du médecin. Le niveau de revenu ainsi que l'état de santé de la patientèle peuvent être corrélés avec la performance médicale. Nous testons alors les relations suivantes :

[83] Nous avons programmé plusieurs modèles de ce type et – sans surprise – ils sont fortement explicatifs. Nous ne présentons pas ces modèles dans la mesure où ils ne nous apprennent rien.

$$Y_i = Y_i(p_i)\ ;\ t_i = t_i(\ p_i)$$

Où, Y_i est la rémunération à la performance du médecin i, t_i le taux de réalisation du médecin i et p_i les caractéristiques individuelles de la patientèle du médecin i. On peut s'attendre à ce que les variables expliquées soient dépendantes de la composition de la patientèle. Nous nous intéressons particulièrement à deux caractéristiques de la patientèle. D'une part, il est probable que les médecins ayant une patientèle plus pauvre aient de moins bons taux de réalisation global que les autres si l'on considère que ces patients sont les moins éduqués à la santé. Or, comme certains indicateurs dépendent directement de l'implication du patient, une mauvaise implication peut engendrer de moins bons résultats pour les médecins. D'autre part, il est probable que les médecins ayant une forte patientèle de très malades aient de meilleurs taux de réalisation que les autres. En effet, ce sont ces médecins là qui sont ciblés par le ROSP notamment via les indicateurs de suivi des maladies chroniques. On peut alors s'attendre à ce qu'ils aient déjà développés une spécialisation particulière en raison de la composition de leur patientèle.

Afin de tester notre hypothèse nous utilisons 5 variables de composition de la patientèle ainsi que 3 variables portant sur les caractéristiques de médecins (variables de contrôle) toutes obtenues par la méthode pas à pas (*stepwise*)[84].

Nous proposons d'approcher le niveau de revenu de la patientèle par la part des patients titulaires de la Couverture Maladie Universelle Complémentaire (CMUC). Ce dispositif permet aux patients les plus pauvres d'être exonérés totalement du coût des soins ainsi que de l'avance de frais (tiers-payant)[85]. Concernant l'état de santé, les bases de données nous renseignent sur l'obtention par les patients du dispositif Affection Longue Durée (ALD) – dont fait partie le diabète de type 2 explicitement ciblé par les indicateurs ROSP. Cette variable peut nous aider à approcher l'état de santé de la patientèle. De façon plus mineure, d'autres variables caractéristiques de la patientèle sont retenues car elles ont un rapport avec les indicateurs ROSP : les patientes âgées entre 50 et 74 ans sont la cible de l'indicateur sur le dépistage du cancer du sein ; la part de boites d'antibiotiques prescrites risque d'être corrélé

[84] Le problème majeur de l'approche économétrique est celui de la sélection des variables, l'objectif étant de minimiser le nombre de variables utilisées tout en maximisant la portée explicative du modèle. La méthode pas à pas inclut d'abord la variable proposant le meilleur coefficient de détermination, puis elle inclut celle qui améliore le plus le coefficient de détermination, puis la suivante et ainsi de suite. Toutes les variables sont entrées dans le modèle et, enfin, les variables qui contribuent le moins au modèle sont exclu.

[85] Rappelons que pour une personne seule le plafond de rémunération mensuel ouvrant les droits à la CMUC est de 720 € environ (www.cmu.fr). Si l'on retient la définition du seuil de pauvreté à 60 % du revenu médian, celui-ci s'établissait en 2012 à 828 € mensuels.

avec l'indicateur de prescription de générique ; la part de patients sous antiagrégants plaquettaires (APP) risque d'être corrélée à l'indicateur de prescription de générique. L'ensemble des variables sont décrites dans le tableau 5.15.

Tableau 5.15. Les variables retenues

Caractéristiques	Nom de variable	Description
du médecin (m_i)	Hon. par patient	Montant des honoraires moyens perçus par patient déclarés par le médecin traitant
	Age du MT	Age du médecin traitant
de la patientèle (p_i)	Part ALD	Part de patients déclarés par le médecin traitant et bénéficiaires du dispositif ALD
	Part CMUC	Part de patients déclarés médecin traitant titulaires du dispositif CMUC
	Part 50/74 ans	Part de patientes déclarées médecin traitant âgées entre 50 et 74 ans
	Part AB	Part de boîtes d'antibiotiques prescrites par patient
	Part APP	Part de patients sous antiagrégant plaquettaire (APP)

Afin de tester notre hypothèse nous utilisons deux types de modèles : une régression par moindres carrés ordinaires et une régression par quantiles. L'idée est que plus la composition de la patientèle est constituée de tel type de patient plus il est difficile (ou facile) d'obtenir de bons résultats en termes de taux de réalisation global et de rémunération à la performance. Le modèle par quantile permet d'affiner l'analyse en comparant les médecins entre eux selon leurs résultats.

3.2. Résultats

Le tableau 5.16. présente les résultats des régressions par quantile et par moindres carrés ordinaires (MCO) pour la variable expliquée taux de réalisation (t_i).

Tableau 5.16. Régression par quantiles et MCO sur taux de réalisation

Régression par quantiles du taux de réalisation - Valeurs estimées des paramètres						
Variables	**MCO**	**10%**	**25%**	**50%**	**75%**	**90%**
Constante	435.4192	290.4596***	380.2258***	466.0469***	523.7429***	618.1212***
	(7.99206)	(11.5976)	(11.0340)	(10.4328)	(10.1997)	(14.0995)
Hon. par patient	-0.3856***	-0.6143***	-0.4992***	-0.4248***	-0.2973***	-0.2375***
	(0.0234)	(0.0691)	(0.0538)	(0.0376)	(0.0262)	(0.0423)
Age du MT	-0.6608***	-0.4909**	-0.6764***	-0.9021***	-0.7415***	-1.3004***
	(0.1400)	(0.1982)	(0.1964)	(0.1926)	(0.1714)	(0.2487)
Part ALD	67.8582**	132.5658***	79.6780 **	65.2042*	15.2434	-39.7824
	(22.6640)	(26.8109)	(34.1150)	(35.3647)	(35.3248)	(47.6069)
Part CMUC	-192.623***	-73.3074	-158.924***	-198.837***	-233.623***	-169.087***
	(29.0603)	(52.7488)	(47.1798)	(36.3763)	(44.0678)	(38.3434)
Part 50/74 ans	-97.8197***	5.3753	-51.5840*	-92.1213***	-172.661***	-231.921***
	(20.0026)	(27.8226)	(27.6764)	(25.6300)	(25.3470)	(38.1593)
Part AB	-41.0070***	-23.2073***	-37.2402***	-48.8141***	-52.1262***	-59.3232***
	(2.4449)	2.9947	(3.2899)	(2.7377)	(2.5508)	(3.2133)
Part APP	917.6172***	804.4675***	830.1158***	845.4491***	975.9273***	1125.073***
	(43.4254)	51.8541	(51.8817)	(63.4200)	(64.2007)	(95.4112)
R^2	0.1554					
R^2 ajusté	0.1546					

Erreur type entre parenthèses
*** p<0,01, ** p<0,05, * p<0,1

Concernant les honoraires par patient MT, l'effet négatif du montant des honoraires est d'autant plus fort que le médecin obtient un taux de réalisation faible. Pour les médecins appartenant au premier décile du taux de réalisation, l'effet négatif des honoraires par patient est le plus fort – comparativement aux autres classes de médecins. La corrélation entre l'âge du médecin et le taux de réalisation est toujours négative le long de la distribution mais, elle est plus forte pour les médecins ayant un taux de réalisation compris dans le dernier décile. L'effet négatif de l'âge est le plus fort pour les « meilleurs ». La corrélation entre le taux de réalisation et la part des patients titulaires du dispositif ALD est positive pour les médecins obtenant un taux de réalisation inférieur à la médiane – et d'autant plus forte que le taux de réalisation est faible. Pour le reste de la distribution, la part de patients titulaires du dispositif ALD n'a pas d'effet significatif sur le taux de réalisation. La corrélation entre le taux de réalisation et la part de patients titulaires du dispositif CMUC est négative sur toute la distribution mais la corrélation est plus forte pour les médecins situés sur le troisième quartile. La part de patientes âgées de 50 à 74 ans est corrélée négativement avec le taux de réalisation d'autant plus fortement que le médecin obtient un meilleur taux de réalisation. La part de patients sous antibiotiques est corrélée d'autant plus négativement au taux de réalisation que le médecin obtient un taux de réalisation élevé. Concernant la part de patients sous APP, la corrélation est positive avec le taux de réalisation et d'autant plus forte que le taux de réalisation du médecin est élevé.

Le tableau 5.17. présente les résultats des régressions par quantile et par moindres carrés ordinaires pour la variable expliquée rémunération à la performance (Y_i).

Tableau 5.17. Régression par quantiles et MCO sur la rémunération à la performance

Régression par quantiles de la Rémunération à la performance - Valeurs estimées des paramètres						
Variables	**MCO**	**10%**	**25%**	**50%**	**75%**	**90%**
Constante	3228.06368*** (115.22621)	1037.657* (117.6255)	2208.041*** (132.4431)	3350.509*** (123.0399)	4513.039*** (148.5260)	5151.933*** (266.0333)
Hon. par patient	-7.69082*** (0.33794)	-5.2345*** (0.3450)	-7.2508*** (0.3884)	-8.4148*** (0.3609)	-7.6766*** (0.4356)	-6.3254*** (0.7802)
Age du MT	-2.47132 (2.01837)	-1.7551 (2.0604)	-5.4254** (2.3199)	-5.7140*** (2.1552)	-4.6631* (2.6017)	3.4411 (4.6600)
Part ALD	-505.47540 (326.75951)	426.9321 (333.5634)	347.3647 (375.5832)	-119.084 (348.9176)	-1734.34*** (421.1914)	-3812.76*** (754.4197)
Part CMUC	2310.60605*** (418.97904)	265.5477 (427.7032)	894.7050* (481.5820)	1571.732*** (447.3908)	3091.428*** (540.0619)	5872.522*** (967.3354)
Part 50/74 ans	- 2389.86806*** (288.38931)	-539.788* (294.3943)	-1387.96*** (331.4798)	-2660.82*** (307.9455)	-4046.57*** (371.7324)	-5094.64*** (665.8309)
Part AB	208.97320*** (35.24940)	280.1452*** (35.9834)	207.8360*** (40.5163)	204.5993*** (37.6397)	95.8844** (45.4363)	93.5078 (81.3835)
Part APP	8275.21854*** (626.08953)	5151.589*** (639.1262)	6985.141*** (719.6384)	8284.028*** (668.5458)	11105.72*** (807.0263)	14360.08*** (1445.510)
R2	0.1240					
R ajusté	0.1231					

Erreur type entre parenthèses
*** p<0,01, ** p<0,05, * p<0,1

La régression par MCO sur la rémunération à la performance, montre une corrélation clairement négative pour les variables honoraires par patient et part de patientes âgés entre 50 et 74 ans. À l'inverse, la corrélation est positive pour les variables part de patients bénéficiaires de la CMUC, part de boîtes d'antibiotiques prescrites par patient et part de patients sous APP. L'âge du médecin traitant et la part de patient titulaire du dispositif ALD ne sont pas corrélés significativement avec la rémunération ROSP.

Concernant les honoraires par patient, la régression par quantile ne montre pas de différences très importantes entre les différentes catégories de médecins. La force de la corrélation négative entre rémunération ROSP et honoraires par patient semble toutefois plus importante pour les médecins ayant obtenu une rémunération ROSP inférieure à la médiane de la distribution. Si la régression par MCO ne montre pas de résultat significatif, la régression par quantile permet d'observer une corrélation négative entre l'âge du médecin et la rémunération à la performance pour les médecins appartenant au deuxième et troisième quartile de la distribution des rémunérations. Par ailleurs, la corrélation négative entre la part de femmes âgées entre 50 et 74 ans et la rémunération à la performance est d'autant plus forte que le médecin obtient une rémunération importante. La corrélation positive entre la part de patients sous APP et la rémunération ROSP est elle aussi d'autant plus forte que le médecin se situe parmi les plus rémunérés. En ce qui concerne la prescription d'antibiotiques, la corrélation positive mise en évidence dans la régression par MCO est d'autant plus forte que le médecin obtient une faible rémunération à la performance.

Concernant la part de patients bénéficiaires du dispositif ALD, si la régression par MCO n'est pas significative, la régression par quantiles montre que, pour les médecins les mieux rémunérés (Q3 et D9), la corrélation est négative et fortement significative. Enfin, concernant les patients titulaires du dispositif CMUC, la corrélation positive mise en évidence avec la méthode des MCO n'est plus significative pour les médecins les moins bien rémunérés (D1 et Q1) mais positive et significative pour les médecins les mieux rémunérés. Pour ceux-là, la corrélation est d'autant plus forte qu'ils sont les mieux rémunérés.

3.3. Discussion

Notre hypothèse de départ était que les caractéristiques de la patientèle – indépendante de l'implication et de la responsabilité du médecin – sont corrélées avec le taux de réalisation et la rémunération à la performance du médecin. L'hypothèse semble confirmée.

3.3.1. Le rôle de la patientèle ALD dans la performance

Plus le médecin dispose d'une part de patientèle ALD importante dans sa patientèle totale, plus il a de chance d'avoir un fort taux de réalisation. Si le fait d'avoir une forte patientèle ALD est corrélé à l'obtention de bons résultats sur le critère du taux de réalisation, la corrélation est, dans certains cas, négative avec la rémunération à la performance. En effet, pour les médecins obtenant les plus fortes primes à la performance, le fait d'avoir une forte patientèle ALD est négativement corrélé à la rémunération ROSP.

La patientèle ALD est une patientèle plus malade – donc plus difficile à traiter – qui est en partie ciblée par le ROSP – notamment sur les indicateurs concernant le suivi du diabète. Il est alors raisonnable de penser que si les médecins ayant une plus forte patientèle ALD sont meilleurs que les autres… c'est qu'ils sont ciblés par le dispositif ROSP (à travers leurs patients). Puisqu'ils ont déjà davantage de patients ALD, alors ces médecins ont été attentifs à la politique publique de ROSP qui les concerne, pris des habitudes de pratique, suivi des formations, etc. qui les ont préparées à être meilleurs que les autres. En ce sens, il est possible que le dispositif, en révélant la patientèle des médecins, révèle une pratique « vertueuse » déjà existante.

Les régressions par quantile permettent de compléter l'analyse. Concernant le taux de réalisation, il est particulièrement intéressant de remarquer que l'effet positif d'une forte patientèle ALD n'est pas significatif lorsque le taux de réalisation est supérieur à la médiane de la distribution. Le fait d'avoir une forte patientèle ALD aide le médecin à améliorer son taux de réalisation lorsque celui-ci est bas, mais, au-dessus de la médiane, le taux de réalisation n'est pas amélioré par l'existence d'une forte patientèle ALD. Soit il y a comme un plafond de l'expérience acquise avec la croissance de la patientèle ALD, soit avec la croissance du nombre de patients, l'effet d'expérience serait noyé par l'augmentation du nombre de patients à traiter.

La corrélation négative entre la part de patient ALD et la rémunération à la performance est plus problématique. Une hypothèse raisonnable serait de considérer qu'avec l'augmentation de la taille de la patientèle (elle-même positivement corrélée à la prime), les médecins ont moins de temps à consacrer à cette population particulière – ce qui réduit leur taux de réalisation global (lui aussi corrélé positivement à la prime). La corrélation finale négative entre la part de patient ALD et la rémunération à la performance peut alors

s'interpréter comme la supériorité de l'effet positif de l'augmentation de la taille de la patientèle face à l'effet négatif de détérioration du taux de réalisation.

3.3.2. La part des patients CMUC négativement corrélé à la performance

Le taux de réalisation est négativement corrélé avec le fait d'avoir une forte patientèle de bénéficiaires du dispositif CMUC. L'analyse par quantile montre que pour avoir une plus forte probabilité d'être « performant », le médecin doit disposer d'une patientèle CMUC inférieure au troisième quartile. En d'autres mots, plus les patients du médecin sont démunis – raison pour laquelle ils peuvent prétendre au dispositif CMUC –, plus le taux de réalisation des médecins est faible. Si les patients ALD sont des cas difficiles parce que ce sont les patients les plus malades, les patients CMUC peuvent également être jugés difficiles dans la mesure où l'on sait que les individus les plus pauvres sont aussi les moins éduqués à la santé. Ainsi, il est raisonnable de penser que les patients CMUC peuvent détériorer les indicateurs ROSP du médecin parce qu'ils sont moins observants (*compliants*)[86]. Les inégalités de revenus étant géographiquement déterminées, il est également raisonnable de penser que certains médecins – exerçant dans certaines zones – peuvent être amenés à se spécialiser dans ce type de patients ce qui implique une plus grande difficulté à obtenir un taux de réalisation plus élevé.

Le sens de la corrélation s'inverse lorsque l'on s'intéresse à la relation entre patientèle CMUC et rémunération à la performance. Les médecins ayant une forte part de patientèle CMUC voient leur chance d'obtenir une forte rémunération augmenter. On pourrait alors penser que le ROSP serait une récompense pour une forme de « médecine sociale ». Cependant, l'analyse de la régression par quantile montre que la corrélation positive entre patientèle CMUC et rémunération ROSP est avant tout tirée par les médecins obtenant une forte rémunération ROSP (Q2, Q3 et D9). La force de la relation positive étant très fortement croissante de la rémunération ROSP, on peut suspecter soit une spécialisation de certains médecins sur les patients les plus pauvres, soit un effet de la taille de la patientèle sur la probabilité d'avoir une forte part de sa patientèle CMUC.

[86] L'observance est le respect par le patient des prescriptions et des stratégies thérapeutiques du médecin. Il s'agit bien évidemment d'une condition essentielle d'efficacité des soins.

3.3.3. La part de patientes âgées entre 50 et 74 ans négativement corrélée à la performance

Le taux de réalisation et la rémunération à la performance sont négativement corrélés à la part de patientes âgées entre 50 et 74 ans. Ce résultat est assez contre-intuitif dans la mesure où la part de patientes âgées entre 50 et 74 ans permet d'approcher la part de patientes par médecin concernée par le dépistage du cancer du sein. Si un médecin a accès à une forte part de ces patientes, on peut s'attendre à ce qu'il obtienne un meilleur taux de réalisation – il « possède » les patients cible. Or, la corrélation est négative : ce résultat peut néanmoins s'expliquer. En effet, le mode de calcul du taux de réalisation incite théoriquement à ne prendre qu'un patient cible par indicateur. Pour l'indicateur de dépistage du cancer du sein, par exemple, il est intéressant de n'avoir qu'une patiente âgée entre 50 et 74 ans et de lui prescrire le dépistage de sorte à obtenir un taux de réalisation de… 100 %. La pondération du taux de réalisation par la patientèle étant faite à partir de la patientèle MT totale et non à partir de la patientèle cible – dans notre exemple, une patiente –, avoir une forte patientèle non cible permet d'obtenir une rémunération très forte. Dès lors, il n'est pas étonnant que la part de patientes âgées entre 50 et 74 ans soit négativement corrélée avec la rémunération à la performance. Plus le nombre de patients cible est fort, plus il est probable de rencontrer des patients non observants et plus l'effort à réaliser est lourd.

3.3.4. La prescription d'antibiotiques : une corrélation négative avec la performance mais pas avec sa rémunération

Avec la variable de prescription d'antibiotiques par patient, on retrouve un cas de corrélation contradictoire selon la variable expliquée. Ainsi, la prescription d'antibiotiques est négativement corrélée avec le taux de réalisation. On peut supposer que les médecins ayant une pratique de prescription vertueuse (moins d'antibiotiques) ont plus de chance d'obtenir un fort taux de réalisation. Cependant, le signe de la corrélation change lorsque l'on s'intéresse à la relation entre la prescription d'antibiotiques et la rémunération à la performance. Les médecins qui prescrivent le plus d'antibiotiques par patient sont aussi ceux qui reçoivent les primes de performance les plus élevées. Paradoxalement, avec le dispositif de ROSP, si prescrire peu c'est prescrire bien, prescrire beaucoup c'est gagner plus.

3.3.5. La part de patients sous antiagrégants plaquettaires : une corrélation positive avec la performance

La part de patient sous APP est positivement corrélée au taux de réalisation et à la rémunération à la performance. Les médecins ayant une part de patient sous APP supérieure à la moyenne, ont plus de chance d'être performants et gagnants. Contrairement aux cas des patientes âgées entre 50 et 74 ans, ici, le fait d'avoir un certain type de patient ciblé par le ROSP est un avantage pour améliorer sa performance et sa rémunération à la performance. Cette différence peut s'expliquer par la plus grande facilité à prescrire des génériques qu'à participer au dépistage du cancer du sein. Alors que dans un cas, la participation du patient et l'investissement en temps du médecin sont indispensables, dans l'autre, il suffit d'ajuster la prescription : il n'y a pas de problème d'observance. On sait par ailleurs, que, d'une part, les pharmaciens sont désormais habilités à substituer les médicaments pour le compte du médecin et que, d'autre part, les médecins sont rémunérés par le ROSP pour se procurer les logiciels informatiques destinés à prescrire des médicaments génériques (indicateurs d'organisation du cabinet et de qualité des services).

3.3.6. La performance et les honoraires : le prix ne signale pas la qualité

Concernant les honoraires par patient, plus ils sont élevés, plus le taux de réalisation est faible et plus la rémunération à la performance est faible. La régression par quantile montre que l'effet négatif des honoraires par patient est d'autant plus fort que les médecins sont en bas de distribution (faible taux de réalisation et faible rémunération à la performance). Ceci suggère que pour les plus faibles (en termes de taux de réalisation et de rémunération à performance), une pratique de prix élevé renforce cette corrélation négative.

Le montant des honoraires par patient est le produit entre le prix de l'acte et le nombre d'actes réalisés par le médecin pour chaque patient. Pour que les honoraires soient faibles, il convient que le montant de l'acte soit faible (sans dépassement d'honoraire) ou que le médecin voit peu de fois le patient dans l'année. En conséquence, ce sont les médecins adoptant l'une de ces deux pratiques qui ont le plus de chance d'être performants et/ou gagnants.

Si l'on privilégie une interprétation en termes de prix de l'acte, la corrélation entre les honoraires par patient et le taux de réalisation produit un résultat contre-intuitif pour la théorie économique : le prix ne signale pas la qualité. En effet, si le taux de réalisation est, pour

l'assurance maladie, un indicateur de qualité, alors, les médecins qui pratiquent les prix les plus élevés sont aussi ceux qui ont les taux de réalisation les plus faibles. Si dans la théorie économique *mainstream* le prix est censé refléter la qualité, dans le secteur des soins on peut supposer que ce n'est pas avéré.

La corrélation entre l'âge du MT et le taux de réalisation est négative. Il est probable qu'avec l'ancienneté, les médecins ne sont plus informés des pratiques contemporaines – d'où sont issus les indicateurs ROSP. Il est également probable que les générations plus anciennes soient plus réticentes à cette forme de médecine quantifiée. La régression par MCO montre que l'âge du médecin n'a pas d'effet significatif sur sa prime à la performance. L'analyse par quantile montre néanmoins que pour les médecins obtenant une rémunération à la performance relativement élevée (Q2 et Q3), la corrélation entre âge et prime est négative. Il est possible que l'effet de l'âge soit brouillé par une corrélation entre cette variable et la taille de la patientèle – bien que l'analyse statistique préliminaire ne le montre pas.

Conclusion

La question de la capacité du P4P médicale à atteindre ses objectifs quantifiés n'est pas le seul problème pertinent lorsqu'il s'agit d'évaluer ce type de dispositif. S'attarder trop longuement sur cette problématique occulte le plus important : que veut dire « performance » dans paiement à la « performance » médicale ? Cet impensé est d'ailleurs peut-être une réponse à la difficulté du P4P à faire la preuve de son utilité. Peut-être est-ce parce qu'il y a un débat étouffé sur ce qu'est la performance que le paiement à la performance est si peu performant ? Tout ce passe comme si la définition de performance médicale allait de soi.

Or, ce chapitre a mis en évidence deux problèmes majeurs en mobilisant l'analyse du dispositif français de ROSP.

D'une part, le premier problème posé est celui de l'identification de la « performance » à une amélioration de l'« état du monde » (au sens de Salais, 2010). Pour toute une série de raison, il n'est pas possible de présenter le dispositif de P4P médicale introduit en France en 2011 comme un outil promouvant l'amélioration de la qualité des soins. Par exemple, contrairement à ce que peut laisser croire la CNAM, il existe un débat dans le milieu médical sur la pertinence même des indicateurs définissant la ROSP. Dès lors, nous ne sommes plus

sûrs qu'une hypothétique amélioration de la performance soit synonyme d'une amélioration de l'« état du monde ».

Dans le cas des médecins généralistes de la région Ile-de-France, nous avons montré que l'identification du taux de réalisation global à la rémunération à la performance implique un écart non négligeable entre médecins « gagnants » et médecins « performants »[87]. En raison de l'effet de taille de la patientèle, les performants ne sont pas forcément gagnants : 12.61 % des gagnants ne sont pas performants et 12.73 % des performant ne sont pas des gagnants. Cet effet taille de la patientèle invite alors à considérer le ROSP comme une rémunération à l'acte déguisée.

D'autre part, le second problème posé par les mécanismes de P4P porte sur l'attribution de la performance : est-il pertinent d'attribuer les « performances » – c'est-à-dire les scores obtenus dans les différents indicateurs ROSP – à la seule responsabilité du médecin ? L'analyse économétrique s'est intéressée à ce versant du problème. Nous avons apporté des éléments de réponse à cette problématique générale en testant l'hypothèse selon laquelle la composition de la patientèle est corrélée aux résultats ROSP des médecins. Nos résultats montrent que les médecins disposant d'une patientèle plus malade obtiennent de meilleurs taux de réalisation mais… de moins bonnes primes à la performance que les autres médecins. Par ailleurs, les médecins ayant une patientèle plus pauvre obtiennent de moins bons taux de réalisation mais… de meilleurs primes que les autres médecins. Dans la mesure où le médecin n'est pas « responsable » du profil de la patientèle qui vient le consulter, il apparaît suspect de lui attribuer toute la responsabilité ou le mérite d'une amélioration de ses indicateurs.

Au total, il est possible de lister brièvement les principaux résultats empiriques obtenus:

- Le taux de réalisation global (45.99 %), c'est-à-dire l'indice de performance en nombre de points, est tiré vers le bas par celui sur le suivi des pathologies chroniques (29.47 %) et celui sur la prévention (39.06 %). Ce sont donc les points réalisés sur l'efficience de la prescription (57.47 % de taux de réalisation) qui contribuent le plus au taux de réalisation total. Or, cette pratique est rémunérée une deuxième fois (indicateur d'organisation du cabinet) et elle est celle qui correspond le moins à la notion profane de performance.

[87] Rappelons la définition de performant et de gagnant. Est considéré performant un médecin qui obtient un taux de réalisation global supérieur à 75 % de la distribution des médecins. Est considéré gagnant un médecin qui obtient une rémunération à la performance supérieure à 75 % de la distribution des médecins.

- Le coefficient de taille de la patientèle ($800/l_i$) pèse de façon majeure dans le calcul de la rémunération à la performance. Compris entre 0.04 et 4.01, ce coefficient est à l'origine d'un écart important sur le nombre de points mis en jeu pour chaque médecin : entre 36 et 3 609 points ! Cela explique en grande partie la « diagonale des anomalies » d'après laquelle 25.34 % sont soit performants mais non gagnants, soit gagnants mais non performants.

- Le fait d'avoir une forte patientèle ALD est corrélé positivement avec le taux de réalisation et est corrélé négativement avec la rémunération à la performance. On peut supposer que la ROSP vient récompenser des médecins déjà « engagés » parce que spécialisés dans la patientèle la plus malade. La corrélation négative avec la rémunération ROSP peut indiquer une sorte d'arbitrage entre quantité de patients ALD et qualité des soins.

- Le fait d'avoir une forte patientèle CMUC est corrélé négativement avec le taux de réalisation mais positivement avec la rémunération ROSP. Il est probable qu'en raison d'inégalités d'accès aux soins et de différences d'éducation à la santé, les patients les plus pauvres sont aussi ceux qui respectent le moins les prescriptions du médecin.

- Le taux de réalisation et la rémunération à la performance sont négativement corrélés à la part de patientes âgées entre 50 et 74 ans. Plus la part de patientes cibles pour l'indicateur de prévention du cancer du sein est élevé, plus il semble difficile d'avoir de bons résultats ROSP. Cela s'explique certainement par le fait que cet indicateur nécessite l'observance des patients (il est co-produit).

- La part de patients sous antiagrégants plaquettaires est positivement corrélée aux résultats ROSP. Contrairement au cas du dépistage du cancer du sein, avoir une part de « patient cible » élevé permet d'améliorer ses indicateurs. Cette différence peut s'expliquer par le fait que l'indicateur dont il est question ici est un indicateur de prescription qui ne suppose pas l'observance des patients.

- Plus les médecins prescrivent d'antibiotiques, plus ils sont rémunérés à la performance mais moins ils réalisent un taux de réalisation élevé. En d'autres termes, si peu prescrire est associé à de meilleures pratiques, c'est paradoxalement le fait de beaucoup prescrire d'antibiotiques qui est récompensé par le ROSP.

L'ensemble de résultats inédits présentés ici sont particulièrement importants mais ils doivent être contextualisés dans le cadre plus général de cette thèse. À un premier niveau, notre étude est une évaluation brute du ROSP qui peut être utilisée par la Caisse nationale d'assurance maladie pour améliorer le dispositif, par exemple, en faisant en sorte que les

gagnants soit partout les performants. La formule de rémunération comme les indicateurs choisis peuvent être reconstruits dans ce sens. Cependant, ils soulèvent d'autres types de questions plus structurelles. À un second niveau, en effet, l'ambition de cette thèse est bien plus large notamment lorsque nous soutenons que la performance mesurée diffère de la performance réelle, ou encore de la définition profane de la performance. Dans la mesure où l'argument n'est pas démontrable à partir de l'analyse économétrique, ce chapitre ne constitue qu'un élément de la démonstration illustrant (mais ne prétendant pas prouver entièrement) l'idée défendue.

Au final, la ROSP est une construction sociale bien plus complexe que l'abstraction théorique de « performance médicale » – entendue comme une amélioration de l'« état du monde » liée exclusivement au mérite individuel du médecin. Dans le cas du travail médical, non seulement l'incitation n'incite pas[88] mais celle-ci porte sur une construction sociale bien fragile. En conséquence, il est pertinent de se demander s'il n'est pas heureux que l'incitation n'incite pas.

[88] On pense à la revue de littérature internationale réalisée au chapitre 4. Le ROSP n'a pas encore fait l'objet d'une évaluation de son efficacité (qualité des soins et économie des ressources).

Troisième partie :

Du paiement à la performance médicale à la régulation par les pratiques

Vers le néolibéralisme sanitaire ?

Introduction de la troisième partie

La théorie économique *mainstream* produit des recommandations de politique publique qui ont le plus grand mal à faire la preuve de leur efficacité. La deuxième partie de cette thèse a montré que le paiement à la performance médicale ne permet pas d'améliorer la qualité des soins ni de maîtriser les dépenses de santé. Il produit même des effets pervers qui remettent en cause la qualité des soins. Pour autant, le paiement à la performance médicale se diffuse dans le monde. On peut alors supposer que, derrière les discours du *mainstream* sur ce mode de rémunération, il existe une rationalité cachée du dispositif. Dans cette hypothèse, le paiement à la performance médicale sert des objectifs qui ne sont pas visibles avec les outils théoriques du *mainstream*.

Cette troisième partie explore l'hypothèse d'une rationalité cachée du paiement à la performance médicale. Pour ce faire nous mobilisons le cadre théorique institutionnaliste. L'idée est alors de chercher dans les institutions – et non dans les individus – les causes de cette forme particulière de la politique publique. De cette façon, nous pourrons également comprendre les résultats médiocres du paiement à la performance.

Plutôt que de penser les outils de politique publique de façon décontextualisée, nous proposons dans le chapitre 6 une analyse historique de la régulation de la médecine libérale française, en nous inspirant de l'Ecole de la Régulation. Cette montée en généralité nous permettra de voir que la Rémunération sur objectif de santé publique (ROSP) n'est qu'un instrument d'un nouveau mode de régulation de la médecine. L'histoire de la régulation de la médecine libérale est en effet celle d'un conflit entre socialisation et libéralisation du financement des soins. Or, la politique publique récente est caractérisée par le passage d'un régime d'accumulation fordiste à un régime d'accumulation postfordiste dans lequel la légitimité de la dépense publique de santé est radicalement remise en cause. Dans cette perspective, l'introduction de la ROSP est un marqueur parmi d'autres de l'évolution de la stratégie du régulateur face à la médecine libérale : alors qu'historiquement la régulation était exclusivement une régulation par les prix (le tarif opposable) dans l'optique d'étendre l'accès aux soins, depuis les années 1990, la régulation se concentre sur les pratiques (la définition et le contrôle de l'activité médicale elle-même) dans le but de contrôler les dépenses de santé. Le contrôle des pratiques est alors ce qu'il reste au régulateur lorsqu'il abandonne l'idée de contrôler les tarifs.

Le chapitre 7 propose d'analyser la régulation par les pratiques comme une construction sociale qui ne peut pas se réduire à une « rationalisation » de l'usage des ressources. Si l'émergence de la régulation par les pratiques s'inscrit pleinement dans les transformations impulsées par le *New public management* (NPM), nous montrons que la condition nécessaire de cette évolution a été l'avènement dans le champ de la santé de la Médecine fondée sur les preuves (*Evidence based medicine*, EBM). Ce n'est que parce que l'EBM fournit des normes chiffrées de qualité des soins qu'il a été possible de penser le contrôle de l'activité médicale dans le cadre de la nouvelle gestion publique. La conjugaison de l'EBM et du NMP institue une nouvelle représentation de la qualité des soins. Nous montrons que cette nouvelle définition de la qualité des soins est contestable car la qualité n'est pas soluble dans les quantités. Parce qu'il n'est pas possible de désencastrer les méthodes de l'EBM de ses conditions sociales de production et parce que le soin est un service (et non un produit), la standardisation de la production des soins parait hautement problématique – ceci expliquant certainement en grande partie les difficultés du paiement à la performance à faire la preuve de son utilité.

L'institution d'une nouvelle représentation de la qualité des soins ne va pas de soi. Elle rencontre de nombreuses résistances menées au non d'une autre définition de la qualité. Le chapitre 8 prend au sérieux l'affrontement entre plusieurs conceptions légitimes de la qualité des soins. À partir du cadre théorique de l'Économie des conventions, nous montrons que la régulation par les pratiques repose sur une évolution de la convention de qualité des soins : d'une convention « domestique/ inspirée » – concevant le soin de qualité comme le fruit d'une relation singulière dans laquelle le médecin doit exercer un jugement – à une convention « industrielle » – concevant le soin comme le respect impersonnel d'un standard. Cette évolution de la convention de qualité a un impact sur l'éthique médicale puisque l'on n'attend plus du médecin une compétence morale mais une compétence technique. À travers la mutation de la convention de qualité des soins et de l'éthique médicale, ce sont les représentations sur le système de santé en général qui évoluent. La standardisation des biens appelle la désingularisation de la relation de soin. Or, dans le contexte plus général du néolibéralisme sanitaire, cette désingularisation est la condition nécessaire pour laisser s'exprimer la rationalité concurrentielle des offreurs et des demandeurs de soin. Il s'agit de faire de la santé un bien comme un autre.

Chapitre 6

La médecine libérale française contre l'État social.

De la régulation par les prix à la régulation par les pratiques

---

Introduction

Le développement à échelle internationale des politiques publiques de paiement à la performance (P4P) médicale est étonnant. Tant théoriquement qu'empiriquement, les deux premières parties de cette thèse ont montré qu'il ne manque pas d'éléments pour juger contestable la diffusion de ce type de dispositif. Ces mécanismes sont bâtis sur des fondements théoriques douteux voire paradoxaux, ils sont d'une efficacité limitée et d'une efficience problématique – sans compter les nombreux effets pervers qu'ils produisent.

La France n'échappe pas à ce constat. Dès 2009, avec le Contrat d'Amélioration des Pratiques Individuelles (CAPI), puis en 2011, avec la Rémunération sur Objectif de Santé Publique (ROSP), les médecins libéraux français sont éligibles à la rémunération à la performance. Si peu de choses sont connues sur l'évaluation du P4P à la française, les maigres informations disponibles invitent à la prudence. Notre chapitre empirique[89] a montré que les « gagnants » ne sont pas forcément les « performants » – ceux qui obtiennent de bon scores à leurs indicateurs de qualité ne sont pas forcément ceux qui obtiennent la plus forte prime à la performance. Les résultats de la ROSP sont par ailleurs corrélés à la composition de la patientèle : les taux de réalisation et la rémunération à la performance ne sont pas uniquement dépendants de l'implication et de la responsabilité du médecin. Plus largement, l'étude des indicateurs de performance révèle la plus grande difficulté à définir précisément ce qu'est la performance, d'autant plus quand une amélioration de la performance aux sens des indicateurs conduit manifestement à une situation préjudiciable au patient et à sa santé.

[89] Cf. chapitre 5.

D'après la théorie *mainstream* et une grande majorité des travaux économétriques sur le P4P[90], toutes les difficultés que ce type de dispositif rencontre proviennent d'une mauvaise adaptation de la théorie au réel. La théorie n'est pas à remettre en cause – il suffirait d'améliorer la conception (*design*) des mécanismes de P4P et de prendre en compte les contextes locaux.

Les critiques que l'on peut formuler au paiement à la performance sont donc imputables à l'effet de contexte. L'histoire des systèmes de santé, les rapports de force, les alternances politiques, les institutions, etc. viennent brouiller l'efficacité d'un mécanisme qui reste pertinent pour la théorie. L'argument du contexte se présente comme un moyen de justifier la théorie quand elle peine à faire la preuve empirique de sa justesse. Le contexte est ainsi l'écart entre le monde idéal de la théorie et celui de la réalité qui est enlaidie par l'existence d'institutions.

Ce chapitre a précisément pour objet de prendre au sérieux ces effets de contexte en considérant que les dispositifs de paiement à la performance s'insèrent dans une histoire longue des systèmes de santé. Le P4P est un moment historique des politiques de santé. Il n'est en rien naturel ni mécanique. Il est le fruit des arbitrages politiques qui ont présidé à son institution. Pour l'illustrer, nous mobilisons les approches institutionnalistes parce qu'elles cherchent à resituer les dispositifs institutionnels dans le contexte de leur émergence.

Pour comprendre les échecs du P4P médicale il est alors nécessaire d'analyser les institutions médicales et leurs évolutions à travers l'histoire. Peut-être qu'il sera alors possible de comprendre l'intérêt unanime des puissances publiques pour le P4P ? En effet, si le P4P ne sert pas à améliorer la performance, à quoi sert-il ? L'objectif de ce chapitre 6 est donc de replacer la politique publique contemporaine dans l'histoire longue de la régulation de la médecine libérale.

La politique de santé en France s'inscrit dans une trajectoire qui reflète l'émergence et l'évolution de l'État social (Castel, 1995, Billaudot, 2008, Ramaux, 2012), en rapport avec les transformations de l'ordre économique et social. On ne peut alors comprendre la situation contemporaine qu'en la replaçant dans une perspective historique, qui constitue le fil directeur de notre analyse. Nous montrerons que la transformation de la médecine libérale tient à la fois

[90] Voir respectivement la première partie et le chapitre 4.

« du hasard et de la nécessité » (Monod, 1970), de la volonté politique et des contraintes économiques.

Cette transformation s'inscrit dans le cadre de la régulation du système de santé, partie prenante des transformations qui affectent l'ordre économique et social. Ces transformations mettent en œuvre un ensemble de procédures et de comportements individuels qui, simultanément, affectent les rapports sociaux par le biais de formes institutionnelles historiquement déterminées (dont le dispositif de santé et de protection sociale), ainsi que par le régime d'accumulation en vigueur (Boyer, 1986, Boyer et Saillard, 1995). Les agents économiques interagissent alors à partir d'une série d'institutions, de règles du jeu et de conventions qui impliquent autant de rationalités situées (Orléan, 1994, Favereau, 1995). Il n'y a pas de transformation des structures sans action transformatrice c'est-à-dire sans action politique (Théret, 1990, Lordon, 2013).

Le capitalisme concurrentiel d'avant la Seconde Guerre mondiale, caractérisé par la diversité des acteurs, va de pair avec une régulation concurrentielle puis débouche sur la construction du monopole des médecins. Le capitalisme fordiste des « trente glorieuses » initie une régulation monopolistique où l'extension de la couverture sociale apparaît comme le fruit du dialogue social et politique, et de la complémentarité entre progrès économique et progrès médical. Le capitalisme postfordiste, concomitant à la crise économique contemporaine, amorce un retour à la régulation marchande avec un recul de la solidarité nationale mais, aussi la coexistence de plusieurs logiques peu compatibles entre elles.

Dans un premier temps (1.), dans les contextes successifs du capitalisme concurrentiel puis fordiste, et en référence aux principales étapes juridiques qui jalonnent la construction de la profession, nous verrons que, dès l'origine, la profession se tient à égale distance du marché et de la solidarité. Avec la Sécurité sociale, le conventionnement soumet les conditions d'exercice de la médecine libérale, notamment les tarifs des actes médicaux, à des accords issus de la négociation entre les représentants des médecins et ceux de l'assurance maladie. Sa mise en place progressive est la pierre angulaire d'une politique d'accès aux soins par solvabilisation de la demande. Mais elle fait apparaître un hiatus entre un choix politique audacieux, dans l'esprit qui anime le Conseil National de la Résistance, et les hésitations et pérégrinations qui président à la concrétisation de ce conventionnement dans la mesure où il porte atteinte au principe de l'entente directe. Selon ce principe, fondateur de l'identité libérale de la médecine française, le prix de l'acte doit être fixé uniquement par un accord

direct et sans intermédiaire entre le patient et le médecin. Pendant cette période, la politique de santé est avant tout une régulation des prix et la régulation des pratiques est inexistante.

La crise économique (et la « nécessité » d'une maîtrise des dépenses de santé) couplée avec le retour d'une conception libérale de la régulation (le « hasard » des choix politiques) conduisent à une certaine remise en cause de la solidarité nationale caractéristique du capitalisme postfordiste.

Dans un second temps (2.), ciblé sur ce capitalisme postfordiste contemporain, nous analyserons comment le tournant du *New Public Management* se traduit dans le champ de la médecine libérale. La politique visant à agir par les prix sur la demande relève d'une démarche incitative de « responsabilisation » implicite ou explicite selon les modalités mises en œuvre, avec pour corollaire une extension de la complémentaire santé porteuse d'inégalités d'accès aux soins renforcées. La politique de l'offre prend la forme d'une régulation par les pratiques qui s'appuie sur l'édiction de normes et sur la rémunération à la performance, qui font peser la même logique incitative sur le médecin, sans que leur efficacité ne soit avérée quant à la modération des dépenses.

Selon nous, depuis le début des années 1990, le passage du contrôle par les prix au contrôle par les pratiques constitue le fait majeur de la dynamique de régulation de la médecine libérale. Ce mouvement, qui peut se définir comme le second grand tournant de la médecine libérale (Bloy et Rigal, 2012), est alors essentiel pour comprendre ce qui se joue autour du P4P, qui n'est qu'un outil d'un objectif plus général de la politique publique : reprendre le contrôle sur la profession médicale.

1 Médecine libérale et solidarité nationale. Un apprentissage difficile

Historiquement, la médecine libérale s'est organisée avant même que naisse l'idée de solidarité nationale. Alors que la grande pauvreté, les guerres et l'échec du *statu quo* libéral vont rendre nécessaire la socialisation des soins, la profession médicale va se construire contre les assurances sociales. La médecine libérale va d'abord freiner l'émergence de la solidarité nationale par son opposition farouche au tarif opposable et à la procédure de conventionnement (1.1.).

La date de 1945 marque un tournant historique avec la création de la Sécurité sociale. Cette institution ambitieuse a pour objectif d'assurer à tous les citoyens une couverture contre les grands risques sociaux. Si les médecins mènent un dernier combat de retardement contre la dynamique de socialisation, à partir de 1960, le principe de solidarité nationale va enfin se concrétiser en médecine ambulatoire. En 1971, le conventionnement national des médecins introduit même un tarif opposable uniforme sur tout le territoire. C'est la période d'affermissement de la solidarité nationale (1.2).

1.1. L'émergence de la solidarité nationale freinée par la profession médicale

Alors que sous l'Ancien Régime les médecins disposaient du monopole de la dispense des soins, la période qui s'ouvre après la Révolution est marquée par une lutte constante de ces derniers visant à retrouver leur situation monopolistique (1.1.1). Si ce combat est couronné de succès en 1892, une nouvelle menace fait jour en cette fin de XIXème siècle. En effet, face au paupérisme, la puissance publique va chercher à organiser collectivement le financement et la dispense des soins. C'est la naissance de l'idée de solidarité nationale. Cependant, les médecins étant hostiles à l'introduction d'un tiers dans la relation de soin, commence alors un combat visant à repousser la socialisation des soins au profit de l'entente directe (1.1.2.).

1.1.1 La profession se construit contre la concurrence des officiers de santé…

La fin du XVIIIème siècle et le XIXème siècle voient se réorganiser la profession médicale après la suppression des institutions conçues sous l'Ancien Régime. Si le décret de Marly (1707) avait donné le monopole des soins aux docteurs en médecine, la révolution de 1789 engagea une période où l'ordre médical ancien fut aboli. Ainsi, la suppression des universités en août 1792, après celle des corporations induite par la loi Le Chapelier, met sur le même pied tous ceux qui se prévalent de la capacité à soigner. Les nouvelles conditions d'accès à la profession permettent de former plus rapidement les professionnels de santé dans un contexte de lourdes pertes dues à la guerre contre l'Autriche.

La loi du 19 ventôse an XI (10 mars 1803) établit les principes de l'organisation de l'activité de soignant en créant le statut des officiers de santé et en instaurant une hiérarchie entre les docteurs en médecine, les sages-femmes et les officiers de santé (Rabier, 2010). Tout en autorisant l'exercice des soins à ces différents acteurs, la loi fait une distinction nette entre

les médecins et les officiers de santé. Les docteurs en médecine doivent effectuer des études longues (quatre ans minimum) et coûteuses (1 000 francs pour passer les examens) alors que les officiers de santé n'ont pas l'obligation d'étudier dans une école de médecine. Ils doivent simplement pouvoir justifier d'avoir exercé quelques années sous l'autorité d'un docteur en médecine ou d'avoir pratiqué dans des hôpitaux civils ou militaires. Mais l'essentiel est ailleurs : la profession médicale ne dispose plus du monopole des soins depuis la Révolution et la loi de 1803 confirme cette situation : les médecins sont en concurrence avec les officiers – les premiers qualifiant les seconds de charlatans[91].

Pour les représentants des médecins, les difficultés économiques de la profession proviennent du trop grand nombre de praticiens. L'offre de soins étant trop élevée, les forces du marché ne pourraient que conduire à exercer une pression à la baisse sur le prix des actes. Les conséquences en termes de politique économique sont alors tout à fait limpides : il serait nécessaire de réduire le nombre d'offreurs en commençant par interdire l'exercice aux officiers de santé (Domin, 2013). De cette manière, les prix de la consultation pourront devenir suffisamment élevés pour assurer à tous la dispense des soins médicaux. En effet, dans la tradition libérale de la médecine, c'est le médecin qui permet d'assurer l'accès aux soins pour tous. L'entente directe autorise l'application d'un prix différencié selon la catégorie sociale du malade. En prélevant des prix élevés sur les malades les plus aisés, les médecins peuvent pratiquer des soins gratuits pour les plus pauvres. Les médecins assurent ainsi un rôle de péréquation des revenus – rendant les institutions telles que les assurances sociales totalement inutiles et même nuisibles.

La volonté de limiter le nombre d'offreur de soins peut se lire dans l'évolution et le développement du syndicalisme médical (Hassenteufel, 1997). Si des premières manifestations collectives de la profession médicale voient le jour dès le début du XIXème siècle, ce n'est que bien plus tard, en 1881, que sont créés les premiers syndicats de médecins (en Vendée et au Havre). Ceux-ci avaient initialement pour objectifs, d'une part, de renforcer la lutte contre le charlatanisme et l'officiat – de « moraliser la profession » – et, d'autre part, de gérer les relations avec les organisations publiques et privées de santé (Damamme, 1991).

Les mutuelles sont vues comme une potentielle menace pour l'exercice libéral. Afin de défendre les intérêts des médecins, le nombre de syndicats progresse rapidement à cette

[91] Cette situation de concurrence peut être nuancée dans la mesure où la patientèle des médecins appartenait à la bourgeoisie tandis que les officiers de santé soignaient exclusivement les patients les plus pauvres.

époque : entre 1881 et 1892, le nombre de syndicats médicaux est passé de 15 à 127. En 1884, certains de ces syndicats se réunissent au sein de l'Union des syndicats de médecins Français (USMF, 74 syndicats, 3 500 médecins pour environ 20 % profession). Mais, la Cour de Cassation ne reconnait pas ces différents syndicats en 1885 car les intérêts du médecin seraient « d'une autre nature » (Hassenteufel, p. 88) que les intérêts habituellement défendus par les syndicats professionnels.

Les médecins doivent attendre la loi Chevandier du 30 novembre 1892 pour que soit légalisé le syndicalisme médical. Le promoteur de la loi, Antoine-Daniel Chevandier, est lui-même médecin et a pu s'appuyer sur la forte présence de députés médecins à l'Assemblée Nationale. En effet, pendant la législature où a été votée la loi (1893-1898), 12.5 % des députes étaient des médecins ce qui en faisait la deuxième profession la plus représentée derrière les avocats et hommes de loi (Pierru, 2007, p. 78). Cette loi rétablit le monopole des soins aux médecins et supprime la profession d'officier de santé[92]. La légalisation des syndicats médicaux constitue par ailleurs un outil pour assurer l'éradication des « charlatans » : les syndicats de médecins sont désormais autorisés à poursuivre en justice toute personne qui exerce sans le titre de docteur en médecine.

Cette forme particulaire d'organisation médicale est censée être un levier pour améliorer le fonctionnement du système de soins. Sans les officiers de santé, les médecins, maintenant mieux rémunérés devraient pouvoir assurer un meilleur accès aux soins pour tous. Cette position se verra remise en cause par les faits. Face à la grande pauvreté, les médecins – qu'ils le veuillent ou non – n'assurent pas la péréquation des revenus promise. Les frais de santé des patients ne sont pas couverts par quelque institution de prévoyance que ce soit : en 1890, seulement 3 % de la population est couverte par un système socialisé (Domin, 2014). Ces dernières sont très rares et ne couvrent généralement que les accidents du travail et le risque vieillesse. Même lorsque le risque santé est couvert les primes sont rarement attribuées et plutôt faibles.

Au point de vue macroéconomique, cette période se caractérise par ce que la Théorie de la Régulation nomme le régime d'accumulation concurrentiel (Encadré 6.1.). Les entreprises capitalistes sont dans une phase primaire d'accumulation du capital et le salariat est une quantité négligeable – ce qui induit des salaires bas et des besoins de santé importants. Il est

[92] Les officiers de santé peuvent poursuivre leur activité (certains termineront leur carrière dans les années 1920) mais, le métier est amené à disparaître.

une quantité négligeable à double titre. D'une part, dans ce régime d'accumulation, les salariés sont très peu nombreux, comparativement à la paysannerie par exemple, et très désorganisés. En conséquence, ils sont incapables de négocier des salaires leur permettant de financer leurs besoins de santé. Le salariat n'est qu'une « armée de réserve » permettant de stabiliser les emballées successives de l'industrie, laissant les ouvriers dans une incertitude croissante quant à leur avenir. D'autre part, l'industrie naissante broie les corps des ouvriers dans les mines comme à l'usine. Les accidents du travail sont légion et il est difficile de faire prendre en charge les réparations par les employeurs qui refusent presque systématiquement le remboursement.

Plus généralement, la dissociation entre le capital (les moyens de production) et le travail (les producteurs), déjà entrevue par Jean Sismondi, plonge la grande masse des producteurs dans le paupérisme. En dépit de la liberté sur le marché du travail, à certains égards, la condition ouvrière est inférieure à l'esclavage puisque le maître nourrit toujours son esclave alors que le capitaliste peut rompre toute relation avec l'ouvrier en le licenciant (Hatzfeld, 1971). Ce dernier n'étant plus producteur mais seulement travailleur, il n'est plus assuré de pouvoir vivre. Le paupérisme peut alors se définir comme cet état d'insécurité grandissant propre au capitalisme concurrentiel.

Dans ce contexte de paupérisation, l'insécurité est grande et l'accès aux soins particulièrement limité. C'est en voyant les atrocités du capitalisme du XIXème siècle que les « républicains de progrès » (Léon Bourgois, Jean Jaurès, Alexandre Millerand, Daniel Vincent, etc.) vont entamer la construction d'une organisation collective et publique permettant d'assurer l'accès aux soins pour le plus grand nombre – et plus généralement de pallier l'« incertitude de l'existence » (Hatzfeld, 1971). Ainsi, si l'État a d'abord activement participé à la construction du monopole de la profession médicale, il va par la suite étendre son action en mettant en place peu à peu des institutions favorisant l'accès aux soins.

Encadré 6.1. Les régimes d'accumulation dans la théorie de la régulation

Boyer (2004)

La Théorie de la régulation définit le régime d'accumulation comme un « ensemble des régularités assurant une progression générale et relativement cohérente de l'accumulation du capital, c'est-à-dire permettant d'étaler et de résorber dans le temps les distorsions et déséquilibres qui naissent en permanence du processus lui-même » (Boyer, 2004, p. 54).

Les recherches, croisant histoire et économie, soulignent le fait que le capitalisme est caractérisé par un changement relativement régulier de régime d'accumulation. Pour distinguer ces différents types de régimes d'accumulation, il faut s'intéresser à deux paramètres importants : le caractère de l'accumulation et le mode de consommation. Le caractère de l'accumulation peut être à dominante extensive – la configuration productive s'étend sans changement significatif des techniques de production – ou à dominante intensive – les gains de productivité sont dus à la transformation permanente des techniques de production. Le mode de consommation importe également pour qualifier un régime d'accumulation dans la mesure où il détermine les caractéristiques de la demande. Soit la consommation est faiblement intégrée à la production manufacturière capitaliste, dans ce cas, la demande est tirée par le secteur agricole, la petite production marchande et la rente – tous en dehors de l'industrie capitaliste. Soit, avec le hausse du nombre de salariés et la transformation du mode de vie des salariés, la demande dépend des salariés de l'industrie, la production manufacturière capitaliste est ainsi dépendante de la main d'œuvre qu'elle emploie pour ses débouchés.

A partir de ces différentes modes de consommation et des caractères d'accumulation, il est possible de dresser le portrait de quatre grands régimes d'accumulation (Tableau 6.1.) – qui ne sont pas exhaustifs de l'ensemble des régimes d'accumulation.

Tableau 6.1. Les quatre principaux régimes d'accumulation

Nature de la consommation		Nature de l'accumulation	
		À dominante extensive	À dominante intensive
Peu intégrée au capitalisme	Exemple historique	Économie Anglaise XVIIIème - XIXème siècle	Économie Américaine XIXème et entre deux guerres
	Nom du régime	Régulation concurrentielle	Accumulation intensive sans consommation de masse
Très intégrée au capitalisme	Exemple historique	Économie Américaine derniers tiers du XXème	Économie de l'OCDE après 1945
	Nom du régime	Postfordisme	Fordisme

D'après Boyer (2004, p. 55)

Régulation concurrentielle : Dans ce régime de régulation, la concurrence et le rapport salarial sont des formes institutionnelles hiérarchiquement supérieures et imposent leurs logiques à l'économie. La concurrence des entreprises capitalistes s'exerce contre la survie des anciens modes de production tels que la petite production marchande. La productivité est décuplée par l'accumulation extensive dans le secteur industriel. Le salariat industriel est minoritaire c'est pourquoi, s'il participe à la formation du profit, il est marginal dans la formation de la demande. Le bouclage de la reproduction économique est réalisé par la demande issue de la paysannerie, de la bourgeoisie et, lorsqu'elle existe, par la dépense publique. La demande est donc tirée par les profits et la stabilité du système repose sur l'armée de réserve : les capitalistes embauchent à bas coût lors des phases de croissance et licencient lors des phases de récession. Le salariat n'étant pas organisé, les entreprises capitalistes sont en position dominante et peuvent imposer leurs conditions salariales.

Accumulation intensive sans consommation de masse : Ce régime est marqué par le développement des sciences et des techniques pour renouveler l'offre de produits et pour rationaliser les méthodes de production. Le passage d'une accumulation extensive à une accumulation intensive provoque des gains de productivité sans précédent. Cependant le mode de consommation évolue peu. La hausse du nombre de salariés n'agit que très marginalement sur le niveau de la demande car leur faible pouvoir de négociation conduit à la modération salariale. Le rapport de force capital/travail est au désavantage du second. Cette situation conduit à un déséquilibre entre les capacités de production stimulées par l'accumulation intensive et le niveau stagnant de la demande. La crise de 1929 est le marqueur de la non viabilité de ce régime d'accumulation.

Fordisme : Le fordisme est rendu possible par une modification majeure dans les compromis institutionnel entre capital et travail après la Seconde Guerre mondiale. Ce nouvel arrangement est à l'origine de l'émergence d'un nouveau mode de consommation : la consommation de masse. Le rapport salarial fordiste repose sur le partage *ex ante* des gains de productivité entre salariés et capitalistes. Ainsi, la poursuite des gains de productivité, par la systématisation de l'application de la science et des avancées technologiques, trouve un débouché dans la demande de la population salariée. L'État dispose également d'une place centrale dans ce régime d'accumulation puisqu'il encourage l'investissement productif, il réalise les infrastructures nécessaires au développement des entreprises et il développe toute une série de protections à destination des salariés – comme la Sécurité sociale.

Postfordisme : Le fordisme s'éteint à partir des années 1970 suite à l'épuisement des gains de productivité pour des raisons technologiques et/ou sociales. On assiste alors au retour d'une accumulation à dominante extensive. Par ailleurs, la période connaît une décomposition du rapport salarial avec la perte de pouvoir de négociation des syndicats. La hausse du chômage et l'individualisation des contrats accroît la position dominante du capital sur le travail. Les inégalités se creusent et l'industrie répond par une différenciation des produits selon les catégories de salariés. L'ouverture des économies à l'international accroît la pression sur le salariat, réduisant à la fois la part salariée de la demande et les protections organisées par les États fordistes.

1.1.2 … tout en refusant la remise en cause de l'entente directe

La fin du XIXème siècle est le point de départ d'une action publique volontariste et ambitieuse en matière de protection sociale[93]. Dans le champ de la santé, c'est la loi du 15 juillet 1893 sur l'Assistance Médicale Gratuite (AMG) qui bouleverse l'organisation des soins. L'État, le département et la commune s'obligent financièrement à mettre en œuvre l'accès aux soins des plus pauvres. Les résultats semblent assez rapidement plutôt bons. Sans financement étatique avant la promulgation de la loi sur l'AMG de 1893, l'assistance médicale dépendait du bon vouloir des départements et des communes – les autorités locales devant financer et organiser la dispense des soins. Avec le temps, ce système a conduit à la décroissance des effectifs de malades pris en charge par l'échelon local : en 1887, 548 119 personnes bénéficiaient du dispositif contre 823 785 en 1885 (Faure, 1984). La tendance à la baisse du nombre d'inscrits et de soignés ne sera inversée qu'à partir de la mise en œuvre de la loi de 1893. À la fin du siècle, le nombre d'inscrits approche les deux millions[94]. Cette loi sera suivie d'autres initiatives parlementaires ayant pour objectif d'améliorer les soins et l'état de santé de la population (loi de protection de la Santé Publique en 1902, loi sur l'assistance

[93] Il faut noter que ce volontarisme est dicté par des impératifs économiques : le solde naturel est négatif et il manque des bras pour produire.

[94] Le nombre exacte de bénéficiaires est de 1 845 911 (Faure, 1984, p.604).

aux vieillards, infirmes et incurables en 1905, etc.) et, au total, malgré la fragmentation des dispositifs et la faiblesse des moyens mis à disposition, les progrès en matière de santé publique sont réellement encourageant.

Cependant, il faut déjà noter les difficultés à faire appliquer la loi du fait de l'opposition de la médecine libérale. L'AMG implique que les médecins soient rémunérés par un tiers à la relation de soin. Par définition, les patients en cause étant démunis, c'est la collectivité – le tiers – qui doit payer le médecin. Or, il est clair que la puissance publique n'est pas prête à rémunérer l'activité médicale à n'importe quel prix : celui-ci, fixé à l'avance, doit être relativement bas afin de ne pas grever les comptes publics. Voilà des exigences insupportables pour la médecine libérale à la française. Dès la fin du XIXème siècle, la solidarité nationale s'oppose à la médecine libérale. Si les lois sur l'assistance ne concernent qu'une catégorie assez bornée de bénéficiaires (les enfants, les vieillards, les anciens militaires, etc.), les lois sur les assurances sociales – beaucoup plus ambitieuses – vont amplifier les confits et accélérer l'organisation de la défense des intérêts médicaux.

En effet, à côté des lois d'assistance, la fin du siècle va voir se généraliser peu à peu les systèmes d'assurance sociale, d'abord mis au point par le patronat. Les lois du 9 Avril 1898 sur les accidents du travail et du 5 avril 1910 sur les retraites ouvrière et paysannes engagent la dynamique de généralisation de l'assurance sociale[95]. Le mouvement de socialisation va s'accélérer au sortir de la Grande Guerre avec le recouvrement de l'Alsace-Lorraine en 1919. Afin de faciliter la réintégration de ces territoires, il ne pouvait pas être question de réduire la qualité des assurances sociales mises en place sous le régime de Bismarck[96].

Un premier projet de loi d'assurance santé est déposé en 1921. En dépit d'une volonté de se distinguer, il reprend certaines caractéristiques du modèle allemand. Les patients sont assurés partiellement contre le risque santé. Pour contrer l'éventualité d'une fraude, ils doivent payer un ticket modérateur. La rémunération des médecins, sous forme de paiement à l'acte, se ferait par l'intermédiaire de leur groupement professionnel. Dans un premier temps, une rémunération globale se verrait attribuée par les différentes caisses aux associations professionnelles travaillant pour elle. Dans un second temps, les associations professionnelles

[95] Notons que, bien avant, la Charte de la Mutualité du 1er avril 1898 légalise les mutuelles et favorise la croissance de la socialisation de la médecine. Si en 1890 3 % de la population est couverte, en 1913 c'est 17 % (Domin, 2014).

[96] La loi sur l'assurance maladie allemande date de 1883 (Palier, 1997).

se répartiraient la manne en fonction du nombre d'actes. Les médecins français voient dans ce projet la fin programmée de la libre entente et vont s'y opposer farouchement.

Il faut pourtant se garder de l'illusion d'une unité sans faille du syndicalisme médical. Si la question du tiers-payant fait consensus au sein de l''Union des syndicats médicaux français (USMF) – tous les syndicats le refusent –, la position vis-à-vis des caisses d'assurance santé fait l'objet d'une vive mésentente. L'assemblée générale de l'USMF en décembre 1925 en est le témoin (Hassenteufel, 1997). Deux ordres du jour sont proposés, l'un en faveur des assurances sociales, l'autre refusant la socialisation et promouvant des rapports individuels entre médecins et patients. À l'issue du vote, le premier ordre du jour est majoritaire (8 316 mandats contre 5 207) mais la minorité fait scission pour créer la Fédération Nationale des Syndicats de Médecins de France (par la suite, la Fédération).

Par la suite, la Fédération va se construire autour des principes libéraux de la médecine où l'entente directe a une bonne place. Minoritaire, elle va au fil du temps gagner en audience auprès des médecins notamment grâce à la légitimité de l'élite médicale fermement accrochée à la défense des principes libéraux. Le congrès de médecins du 30 novembre 1927 viendra confirmer cette audience par l'accord des différents syndicats sur la Charte de la médecine libérale. Les principes énumérés sont le symbole du rejet des assurances sociales et de la défense de l'identité libérale : libre choix, droit à des honoraires pour tout malade soigné, paiement direct par l'assuré en prenant pour base minimale les tarifs syndicaux, liberté thérapeutique, liberté de prescription, etc.

L'organisation des médecins porte ses fruits dès la loi du 5 avril 1928. Cette dernière est en recul par rapport aux projets précédents (notamment celui de 1921) et prévoit le retrait de la médiation des associations professionnelles concernant la rémunération. À la place, le paiement à l'acte est préféré dans le cadre de contrats collectifs locaux (conventions) entre représentant de médecins et caisses qui doivent en outre décider si la rémunération se fait directement par l'assuré ou par l'intermédiaire de la caisse (Hatzfeld, 1963). En d'autres mots, le tiers-payant n'est pas exclu par la loi. Il y a même pire puisque la notion de tarif opposable est introduite. Cela signifierait, en cas d'application, que les médecins seraient contraints d'appliquer le tarif négocié dans le cadre conventionnel – sinon les patients venant consulter ne seraient pas remboursés et seraient incités à consulter d'autres médecins.

Au vu des entraves que fait peser la loi du 5 avril 1928 sur la liberté du médecin, la Confédération des Syndicats Médicaux Français (CSMF), créée dans le courant de l'année (6

décembre 1928) par fusion entre l'UNSMF et la Fédération, refuse d'appliquer la loi telle qu'elle. La reconnaissance du libre choix du patient n'est pas suffisante : les médecins veulent que la loi sanctuarise l'entente directe. Or, la possibilité est laissée d'organiser le tiers-payant. Sous l'autorité du nouveau ministre du travail, Pierre Laval, une nouvelle loi donnant entière satisfaction aux médecins sera votée le 30 avril 1930. Les tarifs opposables deviennent des tarifs de responsabilité non garantits par l'assurance sociale. C'est le montant du remboursement qui est décidé et non le montant reçu par le médecin. Les caisses sont engagées envers les patients à rembourser un certain montant des soins mais les médecins ne s'engagent pas sur le prix des soins. En deux mots, le principe de la convention collective est vidé de son sens. Sans surprise, lors de l'assemblée générale du 27 juillet 1930, la CSMF accepte de participer à la mise en œuvre de la loi.

Du point de vue des cotisants, les lois sur les assurances sociales sont une avancée considérable dans la mesure où sont systématisées l'assurance invalidité et l'assurance vieillesse. Pour le patient, l'assurance contre le risque maladie est créée de toute pièce. Le nombre total de personnes assujetties aux assurances sociales en 1935 est d'environ 12 360 000 soit 56 % de la population active totale (Dreyfus et al. 2006). Ajoutant aux assujettis les ayants-droits, les bénéficiaires des assurances sociales représentent plus de 56 % de la population française (soit près de 23 millions de personnes)[97].

Pourtant, les difficultés d'application ne tardèrent pas. En 1931, des conventions seront signées dans seulement 17 départements. Mais même où elles sont signées, leur fonctionnement ne permet pas d'assurer un taux de remboursement correct aux patients. En effet, les tarifs de responsabilité sont des tarifs minimums fixés sur base syndicale qui ne sont pas opposables. L'existence d'une convention n'empêche donc pas la possibilité pour un médecin de pratiquer des prix supérieurs. Les efforts des caisses signataires s'engageant à rembourser les assurés à un tarif relativement haut (tarif syndical) sont ruinés par le droit des médecins d'appliquer des tarifs encore plus élevés. Les médecins n'étant pas obligés de respecter des tarifs déterminés à l'avance, les assurés sociaux voient l'écart entre le prix payé et le niveau des remboursements périodiquement se réduire puis s'élever. L'émergence de la solidarité nationale apparaît ainsi freinée par la profession médicale.

[97] D'autres estimations trouvent un taux de couverture plus proche de 46 % (Domin, 2014).

1.2. L'affirmation de la solidarité nationale par la mise en place progressive du conventionnement

L'esprit de renouveau issu de la fin de la Seconde Guerre mondiale appelle une nécessaire reconstruction des assurances sociales. L'ambition affichée par les textes fondant la Sécurité sociale reflète un choix de société audacieux tourné vers la solidarité nationale. Il s'agit de garantir à chacun une vie décente par la couverture des grands risques sociaux. Pour le malade, la Sécurité sociale s'engage à garantir une continuité du salaire ainsi que le remboursement des soins et des honoraires médicaux à hauteur de 80 %. Cependant, toujours hostiles à la prise en charge collective du risque maladie, les médecins vont engager un ultime combat de retardement face à la Sécurité sociale (1.2.1.). Il est alors possible de dire que, plutôt qu'une fracture avec la période précédente, la création de la Sécurité sociale est en continuité par rapport aux lois d'Assurance Sociale de 1928-1930. L'esprit des textes ne prenant vie qu'à partir de l'avènement de la Vème république, la rupture est d'abord d'ordre symbolique. Il faut en effet attendre 1971 pour que soit établit un tarif opposable uniforme sur l'ensemble du territoire. C'est l'affirmation du principe de solidarité nationale (1.2.2.).

1.2.1. Des textes qui reflètent un choix de société audacieux…

À la Libération la volonté de rompre avec la violence de la guerre et le paupérisme est unanime. C'est l'esprit qui anime le Conseil National de la Résistance (CNR) dans son programme pour la France libre intitulé *Les jours heureux*.

Cette ambition n'est pourtant pas « gratuite ». Elle s'inscrit dans l'évolution des rapports de force au sein de la société capitaliste. Si les progrès sociaux s'accomplissent à échelle mondiale (plan Beveridge en 1942, Déclaration de Philadelphie en 1944, Charte des Nations Unies en 1945, etc.), la France se distingue néanmoins des autres pays frappés par la guerre sur au moins trois points. D'une part, le traumatisme de la défaite se double de l'occupation du vainqueur et de la collaboration d'une grande partie des classes dirigeantes. D'autre part, la gauche – révolutionnaire ou non – est animée, en 1945, d'une volonté de « châtiment » pour les traitres, qui se composent essentiellement de divers partis de droite et du grand patronat industriel. Les temps sont à la table rase. Enfin, la période est marquée en France par la puissance du Parti Communiste Français et de ses « satellites » (syndicats comme la Confédération Générale du Travail et « intellectuels »). En plus de ces traits distinctifs, la France se retrouve au centre d'une lutte de pouvoirs dans le cadre plus général

de la guerre froide – le modèle capitalise étant alors modéré par l'expérience soviétique commencée en 1917. Tous ces éléments mis bout-à-bout permettent de comprendre l'incroyable basculement des idées et des rapports de force qui rendent possible l'expérience du CNR.

L'effet majeur de la recomposition des pouvoirs à la suite de la guerre porte sur le rôle central accordé au salariat dans ce qui deviendra le régime fordiste. L'assurance maladie n'est alors qu'un élément d'une entreprise plus vaste de modification des rapports entre capital et travail. La Sécurité sociale est un pilier de l'État social[98] lui-même n'étant qu'un rouage du régime fordiste. Dans ce cadre, les gains de productivité sont partagés entre capital et travail de telle sorte que la demande autrefois uniquement bourgeoise et paysanne devient aussi salariée. De nouveaux débouchés s'offrent alors aux entreprises capitalistes ceci permettant d'entretenir un cycle de croissance des salaires et de croissance des profits. L'intervention étatique a par ailleurs un rôle de stabilisateur notamment en octroyant un revenu à ceux qui sont dépourvu de travail. De façon assez inattendue, la protection contre l'« incertitude de l'existence » et la croissance économique capitaliste vont de pair.

Dans cet esprit mais sans anticiper ce que sera le régime fordiste, le CNR prévoit la création de la Sécurité sociale comme institution ayant pour but d'« assurer à tous les citoyens les moyens d'existence dans tous les cas où ils sont incapables de se les procurer par le travail » (CNR, 1944). La faiblesse des prestations des Assurances Sociales de 1930 est, en toute logique, un point noir qu'il convient de corriger de toute urgence (Valat, 2001). Les ordonnances des 4 et 19 octobre 1945, fondant la Sécurité sociale, affichent l'ambition d'un remboursement à 80 % des soins médicaux (20 % de ticket modérateur[99]). Par ailleurs, le remboursement des médicaments est amélioré, les indemnités journalières sont déplafonnées, l'ouverture des droits devient effective à partir de 60 heures de travail (au lieu de 60 jours précédemment), etc.

Mais, le changement central concerne l'identité des gestionnaires de l'institution. Alors que les Assurances Sociales d'avant-guerre sont gérées par les seuls mutualistes, ou que les

[98] Selon Christophe Ramaux (2012), la protection sociale n'est qu'une partie de ce que l'on nomme l'État social. Il distingue en effet quatre piliers : la protection sociale, la régulation des rapports de travail (par l'intermédiaire du droit notamment), les services publics (comme l'éducation) et les politiques économiques de soutien à l'activité et à l'emploi (politique de relance, politique industrielle, etc.).

[99] Le ticket modérateur est la part de la consommation de soins qui est payée directement par le malade. Il est destiné à modérer la consommation de soins – autrement dit, il sert à lutter contre l'opportunisme naturel des acteurs.

systèmes d'assistance universelle, comme le *National Health Service* britannique, sont dirigés par l'État, la Sécurité sociale est gérée par les représentants de salariés (deux tiers) et de patrons (un tiers). Cette structure promeut la démocratie sociale à côté de la démocratie parlementaire – la première étant jugée plus fiable que la seconde, coupable de collaboration. Cette donnée est fondamentale pour comprendre l'évolution de la médecine libérale après 1945.

L'organisation de la profession est fondée sur le principe de conventions départementales à tarif opposable entre les caisses et les syndicats de médecins avec une commission nationale des tarifs (CNT) tripartite (médecins, caisses et administration). Les négociations entre les médecins et les caisses de Sécurité sociale vont rapidement s'organiser à l'échelle nationale entre la CSMF et la Fédération Nationale des Organisations de Sécurité Sociale (FNOSS). La CNT a pour rôle d'autoriser, ou non, les tarifs conventionnels décidés localement. Si elle les juge trop élevés, elle peut demander aux représentants du département en question de renégocier. Si aucun accord n'est trouvé, c'est la CNT qui fixe les tarifs à appliquer. L'issue des débats au sein de la CNT dépend donc du type d'alliances effectuées entre la CSMF, la FNOSS et l'administration.

La situation semble considérablement s'améliorer par rapport aux lois de 1928-1930. Cependant, la mise en application des principes de 1945 ne va pas se réaliser dans l'immédiat. En effet, l'objectif de remboursement à hauteur de 80 % suppose préalablement la coopération des médecins. Il faudrait que les médecins se mettent d'accord pour signer des conventions avec les caisses gestionnaires. Or, le corps médical profondément anti-conventionniste refusera de céder sur le principe de l'entente directe : il est hors de question pour les médecins de s'engager sur des tarifs opposables. En réaction, la CNT fixe des tarifs en dessous des tarifs syndicaux (ceux pratiqués par les médecins) et le ticket modérateur reste proche de 40-50 %, voire plus. Les caisses ayant peur de voir les tarifs syndicaux bondir à chaque revalorisation des tarifs d'autorité, ces derniers ne sont que rarement revalorisés provoquant par là le mécontentement des assurés. Comme le résume Henri Hatzfeld (1963, p. 78), « le premier acte est joué. Les médecins conservent en fait leur liberté en matière d'honoraires : les assurés sociaux ne seront pas remboursés à 80 % ».

Par la suite, il est possible de distinguer deux périodes entre 1946 et 1953 en analysant les stratégies d'alliance entre les trois parties intéressées à la CNT (Hatzfeld, 1963, p. 81). Dans une première période courant de 1946 à décembre 1950, la FNOSS et l'administration

centrale sont en accord relatif sur la politique à adopter vis-à-vis du corps médical. L'équilibre financier des caisses est un enjeu important qu'il n'est pas possible de remettre en cause en donnant satisfaction aux médecins sur leurs revendications tarifaires. Cet enjeu est présent dès les premières années de fonctionnement de la Sécurité sociale. Les dépenses de santé doublent entre 1946 et 1947 et elles doublent à nouveau entre 1947 et 1948. Cette déflagration n'est pourtant pas jugée préoccupante. L'augmentation des dépenses de santé est un effet recherché : il faut améliorer l'état de santé de la population. Afin d'atteindre cet objectif, les cotisations de Sécurité sociales sont doublées (de 8 % à 16 %). Mais les dépenses de santé vont croître beaucoup plus vite que prévu. L'objectif de dépense de 5 % des salaires à 10 ans (pour 1955) est dépassé en 1947 (Valat, 2001).

Les premiers déficits arrivent à la fin de la décennie 1940 si bien qu'en 1947 la presse s'enflamme contre la toute nouvelle institution et qu'en juillet 1949 un débat parlementaire est organisé sur le principe même de la Sécurité sociale. Deux arguments majeurs sont mobilisés pour convaincre de l'inanité de l'institution : d'une part, elle encouragerait de trop nombreux abus des bénéficiaires et des professionnels et, d'autre part, elle serait un surcoût du travail grevant les prix et nuisant aux exportations. Malgré ces griefs, l'Assemblée Nationale décide de ne pas réformer la Sécurité sociale et d'accroître, si nécessaire, les cotisations sociales. L'intervention du Parlement démontre néanmoins que l'autonomie politique des « intéressés » est dépendante de leur capacité à stabiliser les comptes. Dans ce contexte, il est hors de question de fixer des tarifs au niveau des tarifs syndicaux jugés trop élevés et susceptibles de mettre en péril le budget des Caisses. Cette position peut-être paradoxalement vue comme conforme aux revendications du syndicalisme médical qui ne souhaite pas réellement que soit posé le principe des tarifs opposables. En janvier 1950, 55 conventions sont signées, couvrant uniquement un tiers des assurés sociaux.

La seconde période mise en évidence par Henri Hatzfeld concernant les rapports des parties au sein de la CNT débute le 5 décembre 1950 lorsque le bureau de la FNOSS, par l'intermédiaire de son Secrétaire général Aimé Texier, redéfinit radicalement la stratégie de son institution. D'après lui, puisque, d'une part, il y a dans la société des besoins sociaux que la Sécurité sociale doit couvrir et que, d'autre part, c'est au gouvernement de dégager les fonds nécessaire à ce travail, il n'est pas nécessaire de se battre pour les honoraires médicaux. D'ailleurs, toujours selon la FNOSS, la question des déficits dépasse assez largement la question de la rémunération du médecin : c'est tout le système de financement qui est à revoir.

Pour ces raisons, la FNOSS donne mandat à ses représentants à la CNT pour ratifier toutes les conventions quels que soient les prix (n'excédant pas les tarifs syndicaux). Ce faisant la FNOSS rompait « l'alliance implicite » qui la liait avec l'administration lors des votes sur les tarifs à la CNT. Dans ces conditions, la prochaine réunion de la CNT ne pouvait que déboucher sur une revalorisation substantielle des honoraires médicaux.

Mais, le gouvernement de René Pleven ne laissera pas les choses se passer ainsi. Le décret du 20 décembre 1950 donne au ministre du Travail et de la Sécurité et de la Sécurité sociale, Paul Bacon, le pouvoir de suspendre les décisions de la CNT jugées incompatibles avec l'équilibre des finances. En dépit des protestations de la FNOSS et de la CSMF, l'administration reprenait alors la main sur les décisions tarifaires sans toutefois remettre en cause le retournement d'alliance. Plus tard, entre 1953 et 1960 malgré beaucoup d'efforts consentit, la situation ne s'améliore pas. L'échec du projet du ministre des Affaires Sociales, Albert Gazier, est une illustration de plus de la difficulté à trouver un terrain d'entente avec la CSMF. Les médecins ne veulent pas sacrifier leurs libertés et l'administration ne veut pas accélérer plus que ce n'est déjà le cas la croissance des dépenses de santé. Ce n'est qu'avec la proclamation de Vème République et le renforcement du pouvoir exécutif que la situation du système de santé va évoluer. Inspirée par Robert Debré, l'ordonnance du 30 décembre 1958 relative à la création des Centres Hospitaliers et Universitaires est exemplaire de ce nouvel élan donné au système de santé Français. Pour l'instant confiné à l'hôpital, c'est au printemps 1960 que l'esprit des ordonnances de 1945 va « prendre vie »[100].

1.2.2. … mais qui ne se concrétisent que tardivement

Après 15 ans de négociations entre la FNOSS, la CSMF et l'administration aucun accord substantiel sur des tarifs opposables n'est signé et le remboursement des soins médicaux, s'il est en hausse, reste médiocre. Ce n'est que le 12 mai 1960 que la situation va se débloquer. Avant de prendre connaissance dans les détails de l'évolution qui s'est jouée en mai 1960, il faut noter que cette évolution s'inscrit, encore une fois, dans une transformation de plus grande ampleur. En effet, ce ne n'est qu'avec le changement de constitution et le passage à la Vème République que le pouvoir médical devra s'incliner. Evidemment, ce n'est

[100] Selon l'expression d'Henri Hatzfeld (1963, p. 122) au sujet de l'hostilité des médecins à mettre en application les textes de 1945 : « Sans doute aurait-il mieux valu pour le corps médical lui-même qu'il consentît à quelques concessions qui auraient permis aux textes de "prendre vie" ».

pas la résistance des médecins au conventionnement qui a provoqué le renversement de la IVème République. Mais, il est intéressant d'observer que le secteur médical comme les autres est inextricablement lié à l'évolution des rapports de force au sein de la communauté politique. C'est ce qu'avait montré la création de la Sécurité sociale, c'est ce que montre la création d'une procédure de conventionnement fiable en médecine ambulatoire.

Grace à un pouvoir exécutif renforcé, l'ordonnance n°60-461 remplace la CNT par la Commission Interministérielle des Tarifs (CIT), où sont exclus les médecins et les caisses. Les plafonds des tarifs applicables sont désormais décidés par un arrêté ministériel et une convention type est annexée à l'ordonnance afin de cadrer les négociations. Dans les départements où les syndicats ne veulent pas s'engager sur des tarifs opposables, deux tarifs sont fixés. Pour les médecins qui décident d'adhérer individuellement à la convention type, un tarif préférentiel est appliqué, pour ceux qui s'y refusent un tarif moins intéressant est appliqué.

Cette stratégie est « une arme de guerre redoutable » (Hatzfeld, 1963) contre le syndicalisme médical libéral. En effet, dans les départements où des conventions sont signées individuellement, la CIT créée deux catégories de médecins pour les patients : ceux qui sont conventionnés et qui appliquent des tarifs opposables en grande partie remboursés et ceux qui appliquent des tarifs variables dont le taux de remboursement est plus faible. Consulter chez ces derniers implique dès lors deux désavantages : payer plus cher l'acte et, en plus, être proportionnellement moins bien remboursé. Par ailleurs, l'ordonnance limite sévèrement les possibilités de dépassements d'honoraires lorsqu'une convention est signée. Dans le cas où des praticiens exerçaient un comportement tarifaire extraordinaire, il devenait même possible de les poursuivre pour un « comportement général » estimé répréhensible[101]. Si cette disposition a été invalidée par le Conseil d'État en juillet 1962, elle n'en est pas moins symboliquement importante : l'exécutif cherche désormais à contrôler durement les pratiques tarifaires des médecins.

Face à cette attaque évidente au principe de l'entente directe, la CSMF, d'abord, combat vainement, puis, entre en crise. En plus de la menace sur l'un des principes fondateur de la médecine libérale, l'ordonnance du 12 mai 1960 représente une menace contre l'unité syndicale. Le décret semblait en effet être dirigé contre les départements urbains les plus

[101] Ordonnance n° 60-461, Titre II article 9 (JO, 13 mai 1960, p.4357) .

hostiles au principe de conventionnement (Seine et Rhône notamment). À l'inverse, les départements plus ruraux n'y voyaient rien d'inquiétant d'autant plus que ce sont eux qui signaient auparavant le plus de conventions. Il faut insister sur cette fracture entre la médecine urbaine et la médecine de campagne : alors que la première peut s'appuyer sur une population plus importante et plus aisée, la seconde doit exercer dans des conditions plus difficiles et voit d'un bon œil la solvabilisation de la demande de soins.

Le syndicalisme médical connaît dans cette période une forte diversification, qui est le symptôme d'une remise en cause du caractère libéral de la médecine. Le caractère « ultra » de certains contestataires va, comme en 1925, conduire en 1961 à la scission syndicale avec la formation de l'Union syndicale pour la réforme des décrets du 12 mai. Ce syndicat deviendra en 1967 la Fédération des Médecins de France (FMF) dont l'objectif prioritaire est de défendre l'identité libérale de la médecine face à sa socialisation. Cette frange du syndicalisme représente une catégorie de médecins spécifique en pleine expansion : les médecins spécialistes (Hassenteufel, 1992). En plus de la division entre médecine urbaine et médecine rurale, la division du corps médical s'accentue par l'opposition entre spécialistes et généralistes.

La naissance du Groupe Infirmation Santé, d'inspiration trotskyste et maoïste, est un autre exemple de la fragmentation de la représentation médicale. Il s'intéresse aux relations entre l'exploitation de la classe ouvrière et les questions de santé à travers des thèmes comme les conditions de travail, le rôle des médecins du travail ou encore celui de la pression patronale sur les arrêts de travail. Cette branche de la contestation aboutira en 1975 à la création du Syndicat de la Médecine Générale (SMG). En définitive, la CSMF est maintenant divisée et dépassée sur sa gauche par le SMG et sur sa droite par la FMF.

Malgré cette fracture du syndicalisme médical, une chose est sûre : l'accès aux soins a été grandement facilité par l'ordonnance du 12 mai 1960. Un nombre encore jamais atteint de conventions a été signé même dans des départements de longue date anti-conventionnistes : 74 départements signataires en novembre 1960 soit 27 000 médecins pour 9.5 millions d'assujettis. En plus de ces conventions départementales, près de 5 000 conventions individuelles furent signées dans les départements encore réticents. Pour s'assurer de la réussite de la réforme, les pouvoirs publics ont augmenté le niveau des tarifs conventionnels au-dessus des anciens tarifs plafond. Selon les cas, il en résulte une hausse de 30 à 60 % du prix de la consultation. Après une stagnation des dépenses dans les années cinquante, place au

rattrapage : la part des honoraires médicaux, chirurgicaux et dentaires dans le total des dépenses de santé passe entre 1958 et 1967 de 25.1 à 31.3 % (Valat, 2001).

Avec l'augmentation rapide des dépenses de santé, un besoin de financement se fait sentir en 1962 et 1963. Si des mesures de retour à l'équilibre sont décidées dans le cadre du plan Giscard d'Estaing, une nouvelle crise financière se produit en 1966, obligeant à l'élévation des cotisations patronales à hauteur 21 %. La question de l'équilibre financier de la Sécurité sociale devait trouver une réponse dans les ordonnances du 22 Aout 1967. Cette date marque la première grande réforme de l'institution de 1945. D'abord, du point de vue de la gestion, la position majoritaire des syndicats de salariés est mise à mal par la promulgation du paritarisme. Alors que les syndicats salariés disposaient de deux tiers des sièges dans le modèle précédent, ils sont désormais à égalité avec la représentation patronale. Par ailleurs, l'unité de la Sécurité sociale est brisée par la séparation des risques en trois « branches », la Caisse Nationale d'Assurance Maladie des travailleurs salariés (CNAMts) s'occupant du risque maladie. Enfin, les ordonnances de 1967 prévoient la « responsabilisation » des assurés notamment par le durcissement de l'ouverture au droit (120 heures de travail) et l'élévation du ticket modérateur pour les honoraires médicaux de 20 % à 30 %. L'ambition de 1945 est (déjà) revue à la baisse.

En dépit de ces premières mesures tendant vers un infléchissement de la politique publique, les effets en termes de dépense publique et d'accès aux soins de la réforme de 1967 sont assez faibles. L'objectif de solidarité nationale reste prioritaire. Le meilleur témoin de la persistance de cet objectif est le vote, le 3 février 1971, de la loi n°71-525 relative aux rapports entre les caisses d'assurance maladie et les praticiens et auxiliaires médicaux, qui introduit dans le Code de la Sécurité sociale le principe de convention nationale liant les syndicats médicaux représentatifs et la CNAMts. L'idée centrale de la loi est d'organiser la médecine libérale par une convention nationale au lieu des multiples conventions départementales. Il s'agit de généraliser et d'uniformiser sur le territoire le principe de tarif opposable. Le 29 octobre 1971, est publié un arrêté entérinant la première convention nationale signée par la CSMF et la FMF. Le tarif opposable est une réalité sur l'ensemble du territoire. De Brest à Strasbourg et de Lille à Perpignan, en banlieue, en ville ou en campagne, les malades ont la garantie de payer le même prix l'acte médical en consultation ambulatoire et d'être remboursés une part fixe de ce montant. Rétrospectivement, il est possible de dire

que c'est, dans le cadre du remboursement des soins de ville, l'apogée de la solidarité nationale.

Si le conventionnement est une avancée indiscutable pour les patients, elle représente un coût supplémentaire pour ce qui est désormais la CNAMts. Plus de personnes se font soigner et elles sont mieux remboursées. Il est important de rappeler que c'est l'effet recherché. La couverture du risque maladie est l'un des objectifs prioritaire de politique publique depuis les travaux du CNR. Les difficultés financières de la Sécurité sociale, puis de ces différentes branches, ont toujours existé. Mais, jusqu'à cette date, elles ne résistent pas à l'impératif de solidarité. Avec le passage du régime d'accumulation fordiste au régime d'accumulation postfordiste, un basculement va s'opérer au cours de la décennie 1970. Les deux chocs pétroliers de 1973 et de 1979 et la crise qui s'en suit vont être à l'origine d'un renversement des priorités.

2. La régulation par les pratiques : un effet de la remise en cause de la solidarité nationale

À partir de 1973, la crise économique initie un changement de régime d'accumulation du capitalisme en passant du fordisme au postfordisme. Le compromis capital/travail est remis en cause dans l'ensemble de la sphère économique et l'objectif pour les politiques publiques en général et la politique de santé en particulier s'en trouve considérablement modifié. L'État social amorce un tournant néolibéral (Dardot et Laval, 2009). La rationalité individuelle substantielle prend le pas sur la rationalité collective de service public. L'État devient un État régulateur qui « fait son marché » (Batifoulier et al., 2007) en organisant la concurrence entre producteurs dans le champ de la santé et de la protection sociale comme ailleurs. L'incitation financière devient l'instrument privilégié de cette politique.

En médecine de ville, cette politique de *New Public Management* prend la forme traditionnelle d'une régulation par les prix qui pèse principalement sur les usagers et la demande de soins. Le renoncement public à un contrôle des honoraires rend le prix de l'acte plus cher. Pour certains patients, la maladie n'est plus seulement une épreuve physique et mentale mais aussi une épreuve financière (Batifoulier, 2014). C'est dorénavant moins l'offre médicale qui est soumis à une politique des prix que la demande (2.1.). Plus récemment il

semble que cette politique de prix, impopulaire de par ses effets évidents sur l'égalité dans l'accès aux soins, soit supplantée par un renouveau de la politique de l'offre prenant la forme d'une régulation par les pratiques, selon une logique similaire de marchandisation (2.2). C'est la naissance du second grand tournant de la médecine libérale. Tout se passe comme si la lutte traditionnelle entre l'État social et la médecine libérale avait trouvé un nouveau terrain d'expression : les pratiques médicales.

2.1.Une politique de prix au détriment des patients

La politique de prix, caractéristique du « grand bond en arrière » des années 1980 (Halimi, 2004), vise à sensibiliser les acteurs du système de soins au coût de la santé. Le retour de l'entente directe, inhérente à la création du « secteur 2 » (dit à dépassement d'honoraires), relève d'une incitation involontaire : les pouvoirs publics redonnent une place à l'entente directe pour échapper à la revalorisation des tarifs opposables. Le résultat indirect est évidemment la hausse du prix des soins pour les patients. L'objectif de maîtrise des dépenses est bien là… au détriment de l'accès aux soins (2.1.1). Par ailleurs une politique délibérée de reste à charge se développe : hausse des tickets modérateurs, introduction de forfaits hospitaliers, etc. Il s'agit de « responsabiliser » le patient en le faisant participer au coût des soins, au prix de fortes inégalités d'accès aux soins (2.1.2.).

2.1.1. La revanche de l'entente directe…

La crise économique, qui débute dans les premières années de la décennie 1970, est à l'origine d'un retournement d'orientation de politique publique majeur. Autrefois louée pour ses effets stabilisateurs, la dépense publique est dès lors considérée comme la principale cause de la crise. Elle n'est plus un remède mais le mal à soigner. Ce changement profond de référentiel (Muller, 2000) doit être replacé dans le cadre plus général de transformation de régime d'accumulation du capitalisme.

Avec les crises des années 1970, s'opère l'entrée progressive dans le postfordisme. Les gains de productivité dans l'industrie s'amenuisent ce qui provoque une tension sur le compris fordiste entre capital et travail. Or, le poids de la Seconde Guerre mondiale s'est estompé et le rapport de force n'est plus ce qu'il était. Malgré les grandes luttes ouvrières et étudiantes de la fin des années 1960, le conflit de répartition de la valeur est tranché en faveur du capital (Husson, 2010). L'individualisation de la relation salariale – notamment par l'intermédiaire de

dispositifs de rémunération à la performance – désolidarise la classe ouvrière (au sens large). La fin de siècle se caractérise alors par la modération salariale, le retrait progressif de l'État social et la hausse des inégalités de tous types (économique, sociale, état de santé, etc.).

C'est à l'aune de ces transformations du capitalisme que Raymond Barre, économiste de profession, est nommé Premier ministre en aout 1976 par le Président Valéry Giscard d'Estain. Il a pour objectif prioritaire de juguler un phénomène nouveau : la croissance simultanée du chômage et de l'inflation. Alors que longtemps le paradigme keynésien voyait dans ces deux maux une alternative, la crise consécutive aux chocs pétroliers brouille la compréhension des dynamiques économiques[102].

La solution apportée par l'exécutif à ce problème réside dans la maîtrise des dépenses publiques destinée à favoriser le secteur marchand. Les maigres gains de productivité doivent désormais servir à soutenir l'exportation ceci favorisant en retour la croissance. C'est la politique de « désinflation compétitive » – qui, notons-le, marque le début de l'acception généralisée de la part de la classe politique dominante de la « contrainte extérieure », c'est-à-dire, des conséquences logiques du libre-échange. La politique de désinflation compétitive marque ainsi une rupture par rapport aux politiques publiques précédentes qui voyaient dans la relance et le soutien à la consommation le meilleur moyen d'encourager la croissance.

Dans le champ de la santé, l'ordre du jour est le même : contenir l'augmentation des dépenses dans une situation où il n'est pas question d'augmenter des recettes – qui stagnent en raison de la crise économique. Renaît alors de ses cendres le débat autour du déficit de la Sécurité sociale, qui avait été clos par l'Assemblée Nationale en 1949. En ce sens, le problème du déficit de la sécurité social n'est pas un problème neuf, ce qui est nouveau c'est l'attention qu'on lui accorde. L'augmentation des cotisations n'est plus à l'ordre du jour et, encore aujourd'hui, cette présentation mystificatrice du problème reste dominante (Duval, 2007). Le rapport de force et les idées ayant changé, la question du déficit de la Sécurité sociale tend à devenir un problème « naturalisé », relevant de l'ordre des choses physiques et

[102] On pense ici notamment aux controverses concernant la courbe de Phillips laquelle prédit, à partir d'une lecture possible des travaux d'inspiration keynésienne, une relation décroissante entre inflation et chômage. De nombreux auteurs (dont les prestigieux Milton Friedman et Robert Lucas) vont s'efforcer de démontrer à partir des années 1960 les limites de ce type de travaux en insistant sur la notion d'anticipation et d'absence d'illusion monétaire.

non d'une construction sociale et, de la même façon, la maîtrise des dépenses apparaît être une solution naturelle à ce problème naturel[103].

En juillet 1979, le Plan Barrot, du nom du ministre de la Santé et de la Sécurité sociale, Jacques Barrot, prévoit d'encadrer rigoureusement les honoraires médicaux dans l'optique de la discussion sur la convention nationale prévue pour 1980. L'objectif est de borner mécaniquement l'évolution des dépenses au-dessous de l'évolution du PIB par l'instrument d'enveloppe globale (Pierru, 2011). Calqué sur le modèle Allemand, cet instrument permet de fixer *a priori* les dépenses socialisées de santé et de répartir *a posteriori* ce budget (quelle que soit l'activité des médecins). Ce mécanisme repose sur la flexibilité du prix des actes. Au cas où l'activité soit telle que l'enveloppe globale est dépassée, le prix de l'acte est réduit d'autant. Après avoir imposé aux médecins un tarif opposable, il s'agit maintenant de leur imposer une limite macroéconomique de dépenses[104].

Comme ce fut le cas durant la longue marche vers le tarif opposable, les médecins vont s'arc-bouter sur la défense des principes de la médecine libérale et refuser toute maîtrise comptable de leur activité – qui serait, par définition, nuisible à la qualité des soins, donc nuisible aux patients. En 1980, la CSMF appel à la grève et des affrontements violents à Paris entre Compagnies Républicaines de Sécurité et médecins manifestants accroîssent les tensions déjà vives (Pierru, 2011). La défiance à cette période est telle que le projet d'enveloppe globale est rapidement abandonné alors même qu'une contrepartie a déjà été cédée aux médecins : le retour partiel à l'entente directe.

En effet, en 1980, c'est-à-dire en pleine contestation du projet d'enveloppe globale, la frange la plus libérale de la médecine, représentée par la FMF, signe seule une convention nationale prévoyant une réorganisation de l'offre ambulatoire en deux secteurs de conventionnement. Pour les médecins de secteur 1, rien ne change, ils sont astreints au tarif opposable. En revanche, pour les médecins qui décident de s'installer en secteur 2, ils ont la possibilité de pratiquer le tarif qu'ils jugent opportun. Initialement les médecins de secteur 1

[103] Bien évidemment le déficit chronique de la Sécurité sociale et de sa branche maladie n'a rien de naturel. Plusieurs explications ayant un lien avec la modification du régime d'accumulation permettent de remettre cause la fausse évidence du besoin de financement de la Sécurité sociale et la fausse nécessité de maîtriser les dépenses. Ainsi, le déficit de la Sécurité sociale s'explique avant tout par la pénurie organisée des recettes (notamment avec les politiques de réduction des cotisations sociales) et par l'ampleur du chômage et de l'emploi de faible qualité.

[104] Encore une fois, cette idée n'est pas neuve. On se souvient en effet que le projet de loi d'assurance santé de 1921 prévoyait déjà, pour la France, de copier le modèle Allemand en médecine ambulatoire (voir supra).

devaient à eux seuls supporter les limitations d'honoraires prévues par le mécanisme d'enveloppe globale – les médecins de secteur 2 étant « libérés » de toute obligation tarifaire. L'ensemble des contraintes liées à l'enveloppe globale ayant été abandonnées à cause des manifestations, seule la possibilité de s'installer en secteur 2 est restée ouverte. Cet épisode marque la revanche de l'entente directe.

Pour les soins réalisés par des médecins en secteur 2, la différence entre le tarif pratiqué et le tarif opposable deviendra ce que l'on appelle aujourd'hui le dépassement d'honoraire. Notons que le concept de dépassement naît à ce moment précis car, pour qu'il y ait dépassement, il faut une norme – le tarif opposable. La thématique du dépassement d'honoraire n'est rien d'autre qu'un renouvellement de celle de l'entente directe. En ce sens, le secteur 2 n'a rien de nouveau, au contraire, il représente la figure historique de la médecine libérale française. Son intérêt pour la politique publique est lui aussi bien connu : l'existence d'un secteur à honoraires libres permet de ne pas revaloriser régulièrement le tarif opposable. Les médecins peuvent ainsi pratiquer des prix croissants sans que les budgets publics ne doivent dans le même temps augmenter le volume total de remboursement – le montant forfaitaire du remboursement étant un pourcentage fixe de la norme de secteur 1.

La nouveauté avec l'ouverture du secteur 2 vient du fait que les patients consultant de tels médecins sont remboursés comme s'il consultait un médecin de secteur 1 : le remboursement nominal est le même quel que soit le médecin consulté. Alors qu'avant 1960 les médecins ne voulant pas pratiquer de tarifs opposables étaient hors convention et leurs patients étaient peu ou pas remboursés, à partir de 1980 même les médecins ne voulant pas pratiquer de tarifs opposables sont conventionnés (en secteur 2) et voient leurs patients remboursés par la Sécurité sociale sur la même base que pour les médecins de secteur 1. Ainsi, l'ordonnance du 12 mai 1960 est totalement vidée de sa substance et de sa puissance disciplinaire. Avec la convention nationale de 1980, la puissance publique finance partiellement l'entente directe… en introduisant de nouvelles barrières financières d'accès aux soins.

Pour autant, la hausse des dépenses socialisées de santé ne s'est pas arrêtée et la question de la maîtrise de ces dépenses s'est reposée lors des négociations concernant la convention de 1990 (Hassenteufel, 1997). En 1989, la CSMF et la FMF signent un accord cadre s'engageant à limiter les dépenses de santé. Mais le Syndicat de la Médecine Libérale (SML), créé en 1982 par des médecins jugeant la FMF trop conciliante avec les pouvoirs

publics, dénonce cet accord ce qui le rend inapplicable. Les pouvoirs publics se sont alors tournés vers le syndicat de médecins généralistes MG-France – où est représenté le SMG et d'anciens membres de la CSMF – négociant sur la base du gel du secteur 2 : fermeture du secteur 2 pour les nouveaux entrants et limitation de la concentration géographique des médecins en secteur 2. Encore une fois la contestation oblige le pouvoir à la reculade et à la signature avec la FMF (soutenue par le SML non représentatif) d'une convention en 1990 qui limite tout en pérennisant le secteur 2 : le gel n'est pas total puisque les médecins ayant occupé le poste de chefs de clinique peuvent s'installer en secteur 2. Cependant, aucune contrainte sur l'évolution des dépenses n'est prévue et les honoraires sont revalorisés.

Le principe de l'entente directe reprend une dimension centrale dans les négociations entre pouvoirs publics et syndicats de médecins. À certains égards, nous voilà revenu en 1945 – l'objectif de développement inconditionnel de l'accès aux soins en moins et l'objectif de maîtrise des dépenses en plus[105]. Dans ce cadre général, il est de moins en moins acceptable d'augmenter les recettes de l'assurance maladie obligatoire – au risque de perdre de nouvelles parts de marché – et, parallèlement, il n'est pas question d'autoriser la croissance des déficits publics. Il n'y a donc pas d'alternative à la réduction des dépenses.

En conséquence, le contrôle des pratiques tarifaires des médecins libéraux revient à l'ordre du jour. Et les difficultés d'hier vont se renouveler aujourd'hui. Le témoin de ces difficultés est le plan Juppé de 1995. Deux lignes directrices organisent cette réforme. D'une part, il s'agit d'affermir le rôle du parlement dans la régulation des dépenses d'assurance maladie par la création de l'Objectif National des Dépenses d'Assurance Maladie (ONDAM), voté par l'assemblée nationale. Et, d'autre part, ce plan a pour objectif de renforcer les contraintes économiques pesant sur les médecins par l'introduction de l'instrument d'enveloppe globale. Mais, les contestations habituelles sont de mise et les mobilisations, lancées par les syndicats représentatifs, viennent rapidement à bout de la réforme. Le gouvernement renonce une nouvelle fois au mécanisme d'enveloppe globale pour la médecine libérale (Pierru, 2011 ; Hassenteufel et Palier, 2005).

[105] Une explication supplémentaire de ce retournement provient de deux tendances lourdes de l'économie française qui s'affirment clairement au début des années 1990. D'abord, la construction européenne se poursuit et la création de la monnaie unique décidée en 1992 se fonde sur des principes d'économie politique très orthodoxes. Les critères de convergence (déficit public, endettement public et inflation) limitent drastiquement la marge de manœuvre financière des différents gouvernements signataires. Ensuite, les effets de la concurrence internationale se font sentir de plus en plus durement. L'émergence des « dragons asiatiques » ainsi que la place grandissante de la Chine fait peser une « menace » sur l'économie française.

L'échec est d'abord politique. La faillite du plan Juppé, dans son volet santé comme dans son volet retraite, est un élément important pour expliquer l'impopularité et la crise politique du gouvernement à la fin 1996. La dissolution du 21 avril 1997 s'accompagne la même année de la perte des élections législatives au profit de la « gauche plurielle ». Une explication courante de cette défaite repose sur l'opposition de la médecine libérale au pouvoir sortant. Il y aurait en quelque sorte une sanction politique de la part du corps médical à l'encontre de ses représentants traditionnels. Mais, l'échec est aussi judiciaire. Saisi par les syndicats de médecins, le Conseil d'État annule en juin et en juillet 1998 les arrêtés portant approbation des conventions séparées de spécialistes et de généralistes. Les quelques contraintes financières qui y figurent sont jugées illégales car entrainant une rupture d'égalité[106].

La régulation de la médecine ambulatoire paraît vaine. L'expérience de 1996-1997 semble même créer un précédent justifiant l'absence de régulation. C'est certainement ainsi qu'il faut comprendre les différentes revalorisations du tarif de la consultation au début des années 2000. MG France signe, en février 2002, une convention prévoyant la revalorisation de la consultation à 18.5 € (contre 17.5 € depuis 1998). Mais dans la foulée, la contestation des autres syndicats oblige le syndicat MG France à contester lui aussi cette convention en réclamant une nouvelle revalorisation. Dans le contexte d'élection présidentielle, le tarif de la consultation passe en juin à 20 € : « la droite a compris » et le secteur de la médecine ambulatoire n'est plus du tout contrôlé (Pierru, 2007). Puisque les médecins ne sont pas ouverts à la négociation concernant leurs pratiques tarifaires, c'est du côté de la demande de soins qu'il faut trouver une solution à la maîtrise des dépenses.

2.1.2. … et une politique délibérée de reste à charge

En situation de pénurie organisée des ressources (Duval, 2007) et dans la tradition de la théorie conventionnelle de l'aléa moral (Encadré 6.2.), le nouvel objectif de politique de santé est de rendre les patients sensibles aux prix pour qu'ils n'« abusent » pas de la couverture généreuse de l'assurance maladie (Batifoulier, 2013). Il suffirait alors d'augmenter le ticket modérateur et plus généralement tous les restes à charge, pour soulager les finances publiques

[106] Quelques années plus tard, le nouveau gouvernement (dit de gauche plurielle) essaiera lui aussi d'introduire dans la convention de généralistes des contraintes financières mais là encore les décisions seront annulées par le Conseil Constitutionnel.

et discipliner les patients consommateurs de soin. Dans cette perspective, le prix aurait pour effet d'inciter les patients à adopter le « bon » comportement.

Là encore, il est intéressant de noter que cette rationalité de politique publique n'est pas neuve. Depuis longtemps le principe même d'assurance obligatoire publique est violemment critiqué car susceptible d'encourager les individus à céder à leurs penchants naturellement opportunistes. Dans cette perspective, la charité publique encourage les vices privés. La position d'Adolphe Thiers à la fin du XIXème siècle est un bon exemple de ce courant de pensée :

> *« Il est en effet très important pour le bien général des pauvres qu'aucun individu ne puisse considérer la charité comme un fonds sur lequel il a le droit de compter. Le pauvre doit apprendre qu'il doit mettre uniquement sa confiance dans ses propres efforts, dans son travail, dans sa prévoyance : si tout cela lui manque, il faut qu'il sache que sa seule espérance réside dans un secours, et que cette espérance même dépend pour beaucoup de sa conduite et de la certitude qu'il ne s'est pas plongé lui-même dans les difficultés par paresse ou imprudence » (cité dans Hatzfeld, 1971, p. 68.).*

Pour mériter la charité, la pauvreté ne suffit pas : il faut faire la preuve d'être un pauvre vertueux ; un « bon pauvre ». Tout comme la charité entrainerait la mendicité, l'assurance santé entrainerait l'abus. C'est déjà ce que défendait le docteur Liek (1929) à propos du système de santé allemand au début du XXème siècle :

> *« L'imprévoyance, la veulerie, l'irresponsabilité que l'obligation apporte avec elle sont en quelque sorte aggravées par l'"esprit sportulaire", c'est-à-dire le désir des assurés d'"en avoir pour leur argent", de récupérer en prestations ce qu'il ont versé en cotisations. D'où la tendance à quémander toujours d'avantage, à simuler et à frauder. [L'assurance maladie favorise] la mollesse du corps et de l'esprit et conduit à une éternisassions de la maladie » (cité dans Hatzfeld, 1971, p. 92).*

C'est dans cet esprit que les lois d'Assurances Sociales puis la Sécurité sociale prévoient systématiquement un ticket modérateur destiné à limiter la consommation de soins, c'est-à-dire à modérer l'opportunisme présumé des malades. Cette problématique est transversale à l'histoire de l'assurance santé publique. L'abus que font les patients de la générosité de l'assurance maladie ne constitue donc pas une problématique nouvelle, mais la nouveauté réside dans l'attention qu'on lui accorde.

Encadré 6.2. La théorie de l'aléa moral « machine de guerre contre le patient ». Batifoulier (2014)

La théorie de l'aléa moral du patient est une émanation de la Théorie Standard Etendue et de son analyse de la relation de soin comme relation d'agence. Cette fois-ci, ce n'est pas la relation patient-médecin qui intéresse le théoricien mais la relation patient-assureur. Le centre du problème est l'existence d'une asymétrie d'information entre les deux acteurs du contrat. Deux types d'aléa moral peuvent alors être mis en évidence :

- Aléa moral *ex ante* : avant que la maladie ne se déclare, sachant que l'assureur va rembourser ses dépenses de soins en cas de maladie, le patient a intérêt à ne pas faire d'effort de prévention. Par exemple, le fumeur n'a pas besoin de se soucier de sa pratique puisque le jour où on lui diagnostique la maladie, un cancer de la gorge par exemple, il sait qu'il sera généreusement pris en charge par son assureur.
- Aléa moral *ex post* : une fois la maladie déclarée, sachant que l'assureur va rembourser ses dépenses de santé, le patient a intérêt à consommer plus de soins que ce que son état ne le nécessite ou que sa couverture n'est censée couvrir. Pour reprendre notre exemple, le patient malade d'un cancer de la gorge a intérêt à faire payer à son assureur des prestations sans lien avec sa maladie et à se faire prescrire des soins très coûteux – autant de prestations qu'il n'aurait pas acheté s'il avait payé lui-même l'intégralité du coût des soins.

La théorie de l'aléa moral est construite autour de l'idée que l'assurance induit de la consommation de soins. Dans une situation optimale de marché, le patient arbitrerait entre les différents biens en fonction du rapport coût/utilité. Or, à cause de l'assurance, le coût du bien santé est artificiellement réduit ce qui induit nécessairement une surconsommation de soins. La dépense affectée à la santé aurait alors pu être affectée à un autre poste de dépense, octroyant une plus grande utilité. L'assurance santé est un problème et la réduction de l'assurance devient une solution. Sans supprimer totalement l'assurance, il serait souhaitable de multiplier les politiques de ticket modérateur et/ou de co-paiement afin que les patients prennent conscience du coût des soins.

Cette théorie économique s'inscrit totalement dans le paradigme du *New Public Management* et constitue même, selon Philippe Batifoulier, une machine de guerre contre le patient. En effet, cette théorie légitime la réduction de l'État social, alors que les résultats empiriques montrent les effets pervers dévastateurs de ce type de politique publique. Parmi ces effets pervers on peut citer la mise en place par les patients de la « stratégie des 3 R » : les patients Renoncent aux soins, ils Retardent le moment des soins dans le temps et ils Reportent leurs soins vers des lieux mieux pris en charge par l'assurance publique (par exemple, de l'ambulatoire à l'hôpital).

Dans ce cadre, la revanche de l'entente directe a eu pour effet immédiat la hausse du prix de la consultation pour les patients. La politique publique autorise et légitime des dépassements d'honoraires (DH) de plus en plus élevés et de plus en plus fréquents, même chez les médecins de secteur 1, astreints théoriquement au tarif conventionnel. Les DH, tous secteurs de conventionnement confondus, ont augmenté de plus 40 % depuis la fin des années 1990 (Aballea et al., 2008). Ils sont concentrés chez les spécialistes de secteur 2 qui ont vu le montant global de leurs DH passer de 763 à 1 778 millions d'euros (sur près de 18 milliards d'honoraires totaux). Le gel du secteur 2 en 1990 n'est qu'une illusion en ce qui concerne la médecine spécialisée. En effet, en 2004, la majorité des installations de spécialistes (environ 51 %) se sont faites en secteur 2 (dont 86 % des chirurgiens, 90 % des ORL, 75 % des gynécologues).

Outre ces DH constituant une part croissante du reste à charge, la politique publique s'est caractérisée par l'augmentation des tickets modérateurs (TM). Si le TM en ambulatoire

était fixé à 20 % en 1945 puis à 30 % en 1967, la loi de 2004 relative au parcours de soin approfondit la logique de déremboursement. Les patients, auparavant supposés libres de consulter qui bon leur semble, sont désormais incités à déclarer un médecin traitant – seul autorisé à les rediriger vers la médecine spécialisée. La loi prévoit qu'une augmentation du TM est autorisée pour tout patient qui ne respecte pas le parcours de soin. Ainsi, au premier janvier 2006, un patient ne respectant pas le parcours de soin voyait son TM majoré de 10 % soit un total de 40 %. La majoration passait à 20 % en 2007 puis à 40 % en 2009 soit un TM total de 70 %. Entre temps, en 2005, un forfait d'un euro par acte réalisé a été instauré afin de « responsabiliser » les patients devant le coût de la santé.

La légitimation des DH semble être à l'origine du phénomène de refus de soins pour les personnes titulaires de la Couverture Maladie Universelle Complémentaire (CMUC). Il s'agit d'un dispositif destiné aux catégories les plus pauvres de la population leur permettant légalement de ne pas faire l'avance des soins et d'être remboursés intégralement (à 100 %). C'est un dispositif de tiers-payant sans ticket modérateur. Une étude statistique de l'IRDES (Desprès et al., 2009) montre que le refus de soins aux titulaires de la CMU-C à Paris pour les omnipraticiens de secteur 1 s'élève à 9.2 % contre 32.6 % pour les omnipraticiens de secteur 2. Ceci laisse entendre que le principe de tiers-payant est d'autant plus mal accepté par les médecins que ceux-ci pratiquent les DH. Si la relation de soins est avant tout une relation marchande, alors un patient titulaire de la CMUC est doublement handicapant : d'abord, les DH ne sont pas possibles (parce que l'assurance maladie ne les rembourse pas) et ensuite, la rémunération est différée dans le temps (en raison du tiers-payant).

La mise à contribution financière du patient ne s'arrête pas aux frontières de la médecine ambulatoire. À l'hôpital, si le TM est fixé à 20 % et n'est pas remis en cause, il faut y ajouter les « suppléments pour confort personnel » (chambre individuelle, télévision, téléphone, etc.) et les dépassements d'honoraires médicaux au sein même de l'hôpital. Par ailleurs, le forfait hospitalier journalier (FH), crée en 1983, pour faire participer le patient aux coûts d'hébergements, initialement fixé à 20 francs, a connu de nombreuses revalorisations pour atteindre 18 euros en 2010 (13,5 € en service psychiatrique).

La politique de remboursement des médicaments subit une évolution similaire. En 1945, l'objectif était de rembourser 80 % des dépenses de soins, médicaments compris. En 1977, alors que le TM est de 30 %, le plan Veil introduit une différentiation dans le niveau des remboursements. Les médicaments à vignette bleue, jugés moins efficaces, voient leur TM

passer à 60 % et ceux jugés irremplaçables sont intégralement remboursés. La réforme de 2004 développe cette logique, le remboursement dépendant du service médical rendu (SMR) : un TM à 35 % pour un SMR important ; 70 % pour un SMR modéré ; 85 % pour un SMR faible ; 100 % pour un SMR insuffisant. Au 1er janvier 2008 une franchise médicale est introduite sous forme d'une participation de 0,5 euro par boîte de médicament et 2 euros par transport sanitaire.

Globalement le reste à charge a plus que doublé en 30 ans : 547 euros en 2008 contre 217 en 1980 (Perronnin et al., 2011). Dans ce laps de temps, la part de Consommation de Soins et Biens Médicaux (CBSM) couverte par l'assurance maladie obligatoire est passée de 80 % à 75.5 % en 2012 (Commission des comptes de la Sécurité sociale, 2012). La part financée par les organismes complémentaires est de 13.7 % en 2012 pour 25.2 milliards d'euros. Le reste à charge des ménages s'est lui élevé pour atteindre 17.6 milliards d'euros (9.6 % de la CSBM).

L'explosion des restes à charge constitue une opportunité pour l'assurance complémentaire privée. Encouragée par les pouvoirs publics pour masquer le désengagement de l'assurance maladie obligatoire, l'assurance complémentaire est cependant fortement inégalitaire. En effet, ceux qui ont le plus besoin d'une forte couverture complémentaire sont aussi ceux qui ne peuvent pas se la payer. L'extension des complémentaires s'inscrit dans le processus de marchandisation de la santé. La récente signature de l'Accord national interprofessionnel « pour un nouveau modèle économique et social au service de la compétitivité des entreprises et de la sécurisation de l'emploi et des parcours professionnels des salariés »[107] par des syndicats salariés et les syndicats patronaux poursuit cette logique en rendant obligatoire les assurances santé d'entreprise.

Selon une enquête de l'IRDES (Kambia-Chopin et al., 2008), 7 % des personnes interrogées déclarent ne pas avoir de complémentaires santé, et parmi celles qui déclarent en avoir une, 7.5 % sont titulaires de la CMU-C, signe d'un effet pervers : en faisant payer davantage les patients, la puissance publique est obligée de créer des dispositifs de type « filets de sécurité » potentiellement plus coûteux pour le contribuable (Batifoulier et al., 2010). Cette enquête montre également que le taux d'effort, c'est-à-dire le poids du contrat d'assurance complémentaire dans le budget du ménage, est décroissant en fonction du

[107] Accord du 11 janvier 2013.

revenu : 2.9 % pour les ménages les plus riches (1867 euros et plus par unité de consommation dans le ménage), 10.3 % pour les plus pauvres (0 à 799 euros), avec des sommes absolues également décroissantes ce qui implique des contrats de moins bonne qualité. Autrement dit, pour un effort supérieur, les garanties couvertes sont plus faibles. En outre, la hausse du coût des soins, consécutive au désengagement de l'assurance maladie obligatoire, se traduit par un renoncement aux soins pour raisons financières (un français sur sept en 2006, sur base déclarative).

Le renforcement de la place de la complémentaire santé pour la couverture des soins engendre ainsi une « insécurité tarifaire » (Tabuteau, 2013), qui débouche sur des inégalités d'accès aux soins. Ces inégalités sont par ailleurs renforcées par le fait qu'un renoncement aux soins augmente le risque de dégradation de l'état de santé (Dourgnon et al., 2012) : ne pas se soigner rend encore plus malade. Si l'on ajoute à cet état des lieux que ce sont les personnes aux revenus les plus faibles qui ont le plus besoin de soins médicaux (parce que ce sont statistiquement les plus malades), la tendance est au cumul des inégalités sociales et des inégalités de santé (Jusot et al. 2010).

2.2. Le second grand tournant de la médecine libérale : le contrôle de la pratique médicale

La politique publique dans le champ de la médecine libérale ne peut pas se résumer à la politique de prix. S'il est vrai que le régulateur a organisé la liberté tarifaire pour les médecins libéraux, il a également modifié substantiellement la nature de la contrainte qu'il veut imposer à la profession médicale. La régulation par les prix s'est transformée en régulation par les pratiques. De plus en plus, la prolifération de normes chiffrées de bonnes pratiques encadre le contenu du travail médical en amenuisant l'espace de l'« art médical ». Désintéressée du contrôle des prix, la puissance publique cherche aujourd'hui à prescrire les prescriptions – au sens large de prescription médicamenteuse et d'actes thérapeutiques. C'est le second grand tournant de la médecine libérale (2.2.1.). La Rémunération sur Objectif de Santé Publique (ROSP) n'est dans ce cadre qu'un outil de la reprise de contrôle de la profession médicale par la puissance publique (2.2.2.).

2.2.1. Du premier au second grand tournant de la médecine libérale[108]

Dans les années 1960, le sociologue Henri Hatzfeld analysait la méthode de régulation de la médecine libérale par convention comme un élément majeur d'une dynamique plus générale venant remettre en cause le statut de la profession médicale. La socialisation croissante et inéluctable du risque maladie impliquait une collaboration des médecins avec l'assurance maladie, bien qu'historiquement cette profession ce soit construite contre les assurances sociales. Dans la mesure où les médecins commençaient à négocier des accords avec la toute nouvelle Sécurité sociale, nous assistions selon lui au « grand tournant de la médecine libérale » (Hatzfeld, 1963). Ce grand tournant avait alors pour conséquence une réduction sans commune mesure de l'autonomie de la médecine (remettant en cause son caractère libéral si vigoureusement promu et défendu) :

> *« Constatons-le, les deux raisons qui déterminent l'évolution de la médecine agissent dans le même sens. D'une part, les modes financiers de protection contre la maladie mis en place par la société indépendamment des médecins. D'autre part, les exigences scientifiques et techniques de la médecine elle-même. Sous cette double pression, le médecin semble condamné à perdre une part de son autonomie professionnelle : son activité et ses revenus ne serons plus déterminés par son libre vouloir et sa réussite ; il va devenir l'homme d'un statut ou l'homme d'un contrat ; sa place dans l'organisation sociale va se préciser, se fixer » (Hatzfeld, 1963, p. 297).*

Si le pronostic d'Henri Hatzfeld sur le mode de financement du risque maladie et sur la rémunération de la médecine libérale n'est pas gagné, celui selon lequel « les exigences scientifiques et techniques de la médecine elle-même » seraient la source d'une perte d'autonomie semble bien plus proche d'être réalisé.

En effet, depuis les années 1980, il est possible d'observer une tendance générale, issue d'acteurs non nécessairement liés entre eux, cherchant à remettre en cause l'autonomie du médecin dans sa pratique : d'une part, les patients cherchent à s'émanciper du paternalisme médical, d'autre part, la puissance publique veut réduire l'augmentation des dépenses en contrôlant l'activité des médecins. La médecine elle-même est alors transformée de l'intérieur par l'hégémonie de la médecine fondée sur les preuves[109], la nouvelle gestion publique insiste sur la « nécessaire » régulation d'une profession jugée privilégiée, etc. En résumé, tous en appellent à une forme de rationalisation de la médecine afin de poursuivre l'amélioration de la

[108] Le contenu de cette section fait l'objet de développements plus étendus dans le chapitre suivant. L'idée ici est de garder une hauteur de vue permettant d'insérer les évènements contemporains dans le cadre de l'histoire longue de la médecine libérale.

[109] Nous définissons plus bas.

qualité des soins et la maîtrise des dépenses. Les pratiques hétérogènes étant synonyme de plus faible qualité (Kerleau, 1998), la normalisation – c'est-à-dire l'organisation de la convergence des pratiques autour d'une norme (dont le statut reste à préciser) – doit permettre d'améliorer la qualité des soins (Polton, 2010). En tant que science, la médecine devrait pouvoir produire des normes objectives – en cela opposées aux archaïsmes traditionnels sources d'inefficacités.

La décennie 1990 est capitale pour comprendre l'orientation actuelle de la politique de santé. Elle est le témoin d'un tâtonnement entre deux types d'interventions tournées vers l'objectif de maîtrise des dépenses. Nous avons vu que la volonté de contrôler les tarifs médicaux après la libéralisation de 1980 s'est soldée par un échec complet. Mais, parallèlement, la politique publique va entamer pendant cette période une nouvelle forme de régulation de la médecine libérale : le contrôle des pratiques médicales – définissant ce que l'on peut appeler le second grand tournant de la médecine libérale (Bloy et Rigal, 2012).

L'idée centrale justifiant le contrôle de la pratique médicale est de contester le pouvoir médical jugé à l'origine de l'inflation des dépenses de santé. C'est une thématique classique en économie de la santé[110]. La théorie de la demande induite énonce que le pouvoir discrétionnaire du médecin lui permet de formuler lui-même la demande de soins (Evans, 1974). L'offre de soins induit la demande et l'opportunisme médical ne peut conduire qu'à l'inflation des coûts. La théorie de l'agence arrive essentiellement aux mêmes conclusions (Rochaix, 1997). Les asymétries d'information entre le médecin et le couple patient-assureur permettent au médecin d'extraire une rente indue sur le système de santé. Ces théories ont l'ambition d'expliquer non seulement la hausse des coûts mais aussi le niveau de la qualité des soins. Seul acteur informé de la relation de soins, le médecin peut dispenser des soins de piètre qualité sans craindre une sanction marchande. La qualité devient alors une nouvelle source de légitimation pour la politique publique de contrôle de la profession médicale. Ce n'est plus seulement le gestionnaire contre le médecin, ou encore les comptes contre la santé. Le contrôle du médecin est aussi une source d'amélioration de la qualité des soins.

Si cette thématique se fonde sur la théorie économique et sur l'émergence du *New Public Management*, elle est également rendue légitime par l'évolution du rapport de force au sein de la profession médicale elle-même. En effet, la fin du XXème siècle voit le régime

[110] Cf. Première partie.

d'administration de la preuve dans la recherche médicale évoluer (Marks, 1999). Traditionnellement, la recherche médicale était fondée sur l'expérience en laboratoire dont les résultats étaient par la suite testés à l'hôpital. La pratique médicale légitime était alors basée sur l'expérience du praticien ainsi que son autorité. Cette forme d'organisation de la profession s'est vue fortement critiquée pour de nombreuses raisons. En résumé, le mandarin charismatique n'a pas forcément raison et la multitude des points de vue et la complexité des savoirs, décuplés par les progrès médicaux et l'amélioration de l'accessibilité aux données, rendent nécessaire une mise en ordre des connaissances.

Ces raisons sont à l'origine du développement spectaculaire de l'*Evidence Based Medicine* (EBM ; médecine fondée sur les preuves). L'EBM est une méthodologie de recherche fondée sur l'essai clinique randomisé. Selon une définition usuelle (Laurent et al., 2009, Löwy, 1998), ce dernier repose sur trois critères : (i) l'existence d'un traitement de contrôle en plus du groupe test (généralement un autre traitement ou un placebo) (ii) la mise au point d'indicateurs chiffrés permettant de mesurer et d'évaluer les résultats, ainsi que (iii) l'assignation aléatoire pour l'inclusion des patients dans l'essai. On parle d'essai clinique randomisé en « double aveugle » lorsque les informations de l'expérimentation sont cachées à la fois aux patients et aux médecins.

Nous reviendrons longuement au chapitre suivant sur l'influence de l'EBM sur le paiement à la performance. Nous nous contentons de noter ici que l'EBM vient contester et/ou de compléter les connaissances venues de la pratique individuelle par des données externes ayant fait leur preuves statistiquement (Sackett et al., 1996). Ceci permet de réduire la variabilité des pratiques notamment en choisissant les traitements les plus efficaces et en écartant les traitements dangereux. En quelques années l'EBM va devenir le *gold standard* de la recherche médicale et modifier la hiérarchie de la preuve au profit de l'essai clinique randomisé.

Cette modification interne au champ médical aurait pu y rester confinée. Mais, le poids de la contrainte budgétaire grandissant, la puissance publique va s'approprier peu à peu cette vision de l'évaluation médicale (au détriment d'autres). En effet, s'il est possible d'améliorer la qualité des soins et de réduire les dépenses en utilisant des normes de bonnes pratiques scientifiquement prouvées, alors l'autorégulation de la profession n'a plus de raisons d'être. Les normes fournies par le champ médical peuvent être utilisées en retour contre lui. Cette évolution stratégique de la politique publique peut s'illustrer à travers le rôle conféré aux

différentes agences de santé créées à partir de 1989 (Robelet, 1999). À cette date est fondée l'Agence Nationale pour le Développement de l'Evaluation Médicale (ANDEM) qui conçoit l'évaluation en termes de formation continue. La création de l'Agence Nationale d'Accréditation et d'Evaluation en Santé (ANAES), en 1997, montre déjà une évolution de la politique publique. Il ne s'agit plus seulement d'encourager à la formation médicale mais d'autoriser ou non certains établissement à exercer par la procédure d'accréditation. La création de l'Haute Autorité de Santé (HAS, en 2004) participe de la même dynamique : procédure d'accréditation de médecins exerçant certaines spécialités ; diffusion des normes de bonnes pratique issues de l'EBM, etc.

La volonté publique de rompre avec l'autorégulation de la profession médicale s'enracine dans la promulgation de la loi Teulade de 1993. C'est à cette date qu'est introduit le principe de Références Médicales Opposables (RMO). La loi prévoit que les médecins soient contraints d'appliquer certaines normes professionnelles sous peine de sanction financières. Si les RMO ne seront jamais véritablement opposables (à cause de la contestation médicale couronnée de succès par les décisions judiciaires), le plus important est ailleurs. La politique publique change radicalement le mode de régulation de la médecine. Pour la première fois, l'encadrement de la médecine par la puissance publique ne porte plus sur les tarifs mais sur les pratiques.

Malgré l'échec des RMO, l'avènement de l'EBM et son utilisation dans le cadre de la politique publique tendent à renouveler le contenu de la responsabilité médicale. Historiquement, les médecins étaient contraints à respecter les « données acquises de la science ». La question qui se pose aujourd'hui est de savoir s'il faut introduire les résultats de l'EBM dans les données acquises de la science. Si la jurisprudence était frileuse dans les années 1990 (par exemple en vidant le principe de RMO de son contenu), elle tend désormais à reconnaître à ce « droit mou » un statut contraignant.

Le mouvement qui vient d'être décrit n'est rien d'autre qu'une normalisation du travail médical. Un cadre nouveau vient dire ce que doit être l'activité médicale. Ce qui est nouveau du point de vue de l'analyse du système de santé, c'est que ce cadre, provenant d'abord du champ médical, s'est peu à peu imposé par l'activisme de la puissance publique venant briser l'autorégulation des pratiques. L'une des explications de cette appropriation provient de la forme de la normalisation : il s'agit avant tout d'une quantification. Cette quantification permet d'objectiver le travail médical auparavant subjectif. Il devient alors possible pour

n'importe quel profane d'exercer une forme ou une autre de contrôle du travail médical. Doté d'une définition objectivée du travail médical, il ne reste plus qu'à ouvrir la boite noire de la relation médicale pour évaluer individuellement les médecins.

Afin de lever le voile sur le « colloque singulier », une série de mesures en apparence décousues vont être mises en place au cours des années 1990-2000. Il s'agit, d'une part, de créer des systèmes d'information capables de collecter et d'exploiter les informations sur le travail des médecins et, d'autre part, de construire une patientèle à partir de laquelle évaluer le médecin.

Tout contrôle de la main d'œuvre nécessite un système d'information permettant de d'avoir une idée de l'activité de celle-ci. Le système de santé connait cette problématique depuis de nombreuses années. À l'hôpital, le Programme de Médicalisation des Systèmes d'Information (PMSI), destiné à rassembler des malades à caractéristiques communes, est créé dès 1982 et mis en place par le plan Juppé de 1996. Ces informations rendent possibles de pratiquer, à l'heure actuelle, la tarification à l'activité (Domin, 2013). Pour la médecine ambulatoire, la situation est plus difficile. Les rares informations disponibles sont réparties essentiellement dans deux bases de données : le Système National d'Information Inter Régimes de l'Assurance Maladie (SNIIR-AM) et le Medic'AM. Les informations portent principalement sur l'assuré, le bénéficiaire et sur le professionnel prescripteur. Il est par exemple possible de savoir quel médecin prescrit quel médicament à quel patient. Mais la médicalisation des informations n'est pas encore totalement opérationnelle : l'abondance de données administratives contraste avec la faiblesse des données médicales. Le Dossier Médical Personnel (DMP), en cours de développement depuis 2005, est un élément qui viendra bientôt combler les lacunes des systèmes existants (voir sur le sujet les rapports de l'IRDES (2006) et de l'IGAS (2002)). Si les difficultés à extraire une information opérationnelle sont encore importantes, il est déjà possible d'avoir une idée de la « production » des médecins libéraux.

S'appuyant sur ces données informatisées, le P4P médicale n'est alors qu'un instrument parmi d'autres du contrôle que souhaite imposer la puissance publique aux pratiques médicales.

2.2.2. Le P4P médicale comme instrument du contrôle des pratiques

L'introduction des dispositifs de P4P s'est faite, en France, à petit pas. Avant sa généralisation en 2011, on trouve une première tentative de mise en place de P4P à la toute fin des années 1980. Quelques semaines avant d'acter l'ouverture du secteur 2 (en 1980), l'assurance maladie proposa aux médecins libéraux un mécanisme de maîtrise des dépenses ciblant 18 thèmes prioritaires – les médecins obtenant les meilleurs résultats se voyant attribuer une prime forfaitaire (Hassenteufel, 1997). Les représentants de médecins refusèrent cette proposition et il ne fut plus question de ce type de rémunération jusqu'au tout début des années 2000.

Il faudra en effet attendre la Loi de Financement de la Sécurité Sociale (LFSS) de 2000 pour renouer avec le principe de paiement à la performance en France. La LFSS prévoit ainsi la création d'Accords de Bon Usage des Soins (AcBUS) et de Contrat de Bonne Pratiques (CBP). Plus tard, avec la LFSS de 2002, le Contrat de Santé Publique (CSP) est venu compléter le renouveau du système incitatif Français (Encadré 6.3.). L'idée centrale pour l'assurance maladie est d'organiser une sorte de réallocation des ressources en proposant aux médecins de les intéresser (collectivement ou individuellement) à la réduction des dépenses de santé. La sur-rémunération est en ce sens une contrepartie aux dépenses évitées grâce à l'amélioration des pratiques – en termes de qualité et de gaspillages.

Encadré 6.3. AcBUS, CBP et CSP :

D'après la Loi de Financement de la Sécurité Sociale (JO, 2000 et 2002)

L'Accord de Bon Usage des Soins (AcBUS) vise à améliorer les pratiques des professionnels conventionnés et des centres de santé par l'intermédiaire d'un intéressement collectif. Des objectifs chiffrés d'évolution des pratiques et des dépenses permettent d'orienter les médecins et, le cas échéant, de les rémunérer.

Les Contrats de Bonnes Pratiques (CBP), fondés sur le volontariat, se différencient des AcBUS par le fait qu'ils sont individuels. L'objectif de la Caisse nationale d'assurance maladie est cependant le même : faire évoluer les pratiques des médecins vers le bon usage des soins. Le médecin s'engage dans l'évaluation de ses pratiques professionnelles, dans des actions de formation continue, dans le contrôle de son activité médicale, dans la prescription de médicaments génériques, dans l'application de références médicales opposables et de références de bonnes pratiques, dans la coordination des soins, etc. En contrepartie, le médecin est éligible à une rémunération forfaitaire annuelle et à une majoration de la participation de la Caisse Nationale d'Assurance Maladie à ses cotisations sociales.

Le Contrat de Santé Publique (CSP) engage le professionnel à des actions sur la permanence des soins, à la coordination des soins et à des actions de prévention. Si le champ d'application du CSP est plus restreint que les CBP, il s'accompagne également d'une rémunération forfaitaire.

Ces trois dispositifs ont été supprimés à partir du 1 avril 2010 dans le cadre de la mise en œuvre de la loi portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires (HPST) publiée au Journal officiel le 22 juillet 2009.

En 2004, la réforme du « parcours de soin » est une étape supplémentaire vers la définition de la « production » médicale – étape nécessaire afin de mettre en place la rémunération sur objectifs chiffrés. Cette réforme vise à ce que chaque patient désigne un médecin comme son médecin traitant sous peine de sanctions financières (hausse du ticket modérateur). Elle permet de résoudre le problème de la constitution de la patientèle d'un médecin. En effet, comme le système de santé français repose sur la liberté des acteurs, théoriquement, un patient peut consulter n'importe quel médecin. Il est alors impossible de lier un patient à un médecin. Grâce à la réforme du parcours de soin, il devient possible d'associer à des informations concernant les patients un médecin (traitant) « responsable ».

La politique publique dispose désormais d'informations sur l'activité des médecins et de normes de bonnes pratiques permettant de contrôler la qualité des prestations. C'est grâce à ces deux outils que, depuis 2009, les médecins libéraux français ont la possibilité d'être payés à la performance. Dans cette conception, la norme étant une cible à atteindre, établir un lien arithmétique entre la rémunération et la production est un moyen d'inciter les médecins à améliorer la qualité des pratiques tout en réduisant les dépenses de santé. C'est l'objectif du Contrat d'Amélioration des Pratiques Individuelles (CAPI) de 2009.

Le CAPI est un contrat individuel d'une durée minimale de 3 ans signé entre un médecin libéral conventionné et la Caisse Primaire d'Assurance Maladie dont il dépend. Le choix d'adhésion (ou non) est volontaire. Le contrat s'adresse à des médecins libéraux conventionnés ayant une patientèle suffisante l'ayant déclaré comme médecin traitant (MT) quel que soit leur secteur tarifaire (avec ou sans dépassement d'honoraires) – le calcul des indicateurs reposant uniquement sur la patientèle déclarée MT. D'après les indicateurs construits par l'assurance maladie, le médecin signataire s'engage à améliorer ses pratiques de prévention (vaccination contre la grippe, dépistage du cancer du sein et iatrogénie médicamenteuse) sa qualité de prise en charge des patients atteints de diabète ou d'hypertension artérielle et ses prescriptions – qui doivent être « moins onéreuse à efficacité comparable conformément aux données actuelles de la science ». De son côté, l'assurance maladie s'engage à fournir aux médecins, chaque trimestre, les données nécessaires au suivi du contrat et une contrepartie financière pour les médecins atteignant les résultats prédéfinis (Encadré 6.4.).

Encadré 6.4. Les conditions de rémunération dans le cadre du CAPI

D'après Journal officiel (2009)

Pour chaque indicateur sont calculés un niveau initial (celui qui est observé avant la signature du contrat), un objectif intermédiaire et un objectif cible (ces deux derniers étant communs à tous les médecins). À chaque anniversaire du contrat, le niveau de l'indicateur atteint par le médecin est constaté pour l'année écoulée. De là un taux de réalisation (TR) de l'objectif est calculé :

- Si le niveau constaté est inférieur au niveau intermédiaire, le taux de réalisation est proportionnel au progrès réalisé sans toutefois pouvoir dépassé 50 % (Tr < 50 %).
- Si le niveau constaté est compris entre l'objectif intermédiaire et l'objectif cible, le taux de réalisation est inversement proportionnel à l'écart à l'objectif cible, majoré de 50 % (50 % < TR < 100 %).
- Si ce le niveau constaté est égal ou supérieur à l'objectif cible, le taux de réalisation est égal à 100 % (TR = 100 %).

Une moyenne des taux de réalisation par indicateur est calculé pour chaque champ : « dépistage et prévention – suivi des pathologies chroniques » et « optimisation des prescriptions ». La contrepartie financière est versée chaque année dès lors que la moyenne des taux de réalisation des deux champs est supérieure à 25 % (TR > 25 %). La rémunération est égale à 7 € par patient MT multiplié par taux de réalisation moyen (le taux moyen est la moyenne entre les deux champs avec une pondération de 40 % pour le champ « optimisation des prescriptions » et 60 % pour l'autre).

Par exemple, l'un des indicateurs concerne la part des patients MT de plus de 65 ans vaccinés contre le virus de la grippe (seuil minimal de 20 patients). Si cette part est d'au moins 71 % (objectif intermédiaire), alors le médecin obtient pour cet objectif un taux de réalisation supérieur à 50 %. En conséquence, pour cet indicateur, il est réputé performant et est éligible au supplément de rémunération. En deçà de 71 %, si toutefois sa progression n'est pas suffisante entre le niveau initial et le niveau constaté (TR < 25 %), il ne reçoit rien pour cause de performance insuffisante. Il est cependant possible d'être performant sans atteindre l'objectif intermédiaire si toutefois la progression est suffisamment grande (25 % < TR < 50 %). Au-delà de 75 % (objectif cible) de patients de plus de 65 ans vaccinés contre le virus de la grippe, le médecin en question obtient un taux de réalisation de 100 % de l'objectif. Pour cet indicateur, il est réputé performant et est éligible à la rémunération maximale.

Par les sommes allouées et les objectifs affichés, le CAPI présente une évolution majeure par rapports aux dispositifs antérieurs. Mais, cette évolution ne va pas sans résistances. Il en est ainsi du Conseil National de l'Ordre des Médecins (CNOM) qui, par l'intermédiaire de son président, le Docteur Michel Legmann, s'est insurgé de la situation – jugeant le CAPI « antidéontologique », « discriminatoire » et « de nature à briser la relation de confiance entre patients et médecins » (CNOM, 2009). Les syndicats de médecins moins hostiles mais contestataires ont également refusé de donner leur aval au CAPI. Cependant, dans la mesure où ce dispositif est fondé sur le principe du volontariat individuel, ses organisations n'avaient pas le pouvoir de bloquer la mise en œuvre du dispositif.

Ainsi, en dépit de l'opposition initiale des syndicats professionnels et du conseil de l'Ordre, les médecins ont assez largement adhéré à ce dispositif (16 000 médecins traitants libéraux, soit 1 éligible sur 3 en 2010). Ce succès s'explique probablement par le fait qu'aucune revalorisation du tarif de la consultation n'avait eu lieu depuis 2007. Concernant le

CAPI, selon les chiffres publiés par l'assurance maladie, 70 % des signataires ont touché une prime, qui en moyenne s'élevait à environ 3 000 € (Commission des comptes de la Sécurité sociale, 2011). Le montant de la prime s'élevait à 1 500 € en moyenne pour les 10 % les moins rémunérés et à 4 900 € en moyenne pour les 10 % les mieux rémunérés – le maximum supérieur à 17 000 €.

Il faut noter que ces primes restent marginales compte tenu de la rémunération totale des médecins. En effet, le paiement à l'acte représente près de 92 % de la rémunération totale des médecins libéraux, le CAPI ne représentant que 2% de celle-ci – le reste étant constitué de différents forfaits non liés à l'atteinte d'objectifs chiffrés. Du point de vue de l'assurance maladie, le CAPI lui a coûté la première année près de 28 millions d'euros. D'après elle, le dispositif serait neutre budgétairement dans la mesure où les paiements au titre de la performance correspondrait aux économies réalisée grâce à l'amélioration des pratiques[111].

Le contraste entre l'opposition – plus ou moins virulente – des instances représentatives de la profession et l'adhésion au dispositif a servi de justification à l'assurance maladie pour développer le P4P : d'une logique de volontariat, le CAPI a été généralisé, à partir de 2011, à l'ensemble des médecins traitants conventionnés. Le CAPI est devenu Rémunération sur Objectif de Santé Publique (ROSP).

Dans sa nouvelle version, le P4P Français n'est plus fondé sur le volontariat – ou presque. En effet, la convention prévoit que tous les médecins traitants ayant le seuil minimal pour le calcul des indicateurs soient automatiquement bénéficiaires du dispositif. Pour refuser le dispositif, les médecins concernés doivent exprimer ce choix par lettre recommandée avec accusé de réception à la Caisse primaire d'assurance maladie dont ils dépendent dans un délai de trois mois après la parution de la convention au Journal officiel (ou trois mois après la première installation, le cas échant). Ainsi, alors que seulement un tiers des médecins éligibles avait signé volontairement le CAPI, 98 à 99 % sont inclus dans le calcul du ROSP[112].

Au-delà de ces résistances visibles, il faut noter l'important développement du P4P en France depuis 2000 : AcBUS et CBP en 2000, CSP en 2002, CAPI en 2009 et ROSP depuis 2011. Cette dynamique n'a d'ailleurs pas de raisons de s'arrêter dans la mesure où le P4P est

[111] A notre connaissance, aucune étude ne vient corroborer cette justification du CAPI.

[112]Concernant la minorité refusant le ROSP, on pourra consulter le site internet http://www.les-medecins-qui-refusent-la-prime-secu.org/, recensant, de manière non exhaustive, certaines des lettres envoyées aux Caisses primaires d'assurance maladie.

destiné à devenir le troisième pilier de la rémunération de la médecine libérale (Encadré 6.5.). Il faut ainsi noter l'évolution du nombre d'indicateurs du ROSP. Si le CAPI prévoyait 16 objectifs cibles pour les seuls médecins traitants, en 2011, le ROSP élargissait la performance à 29 indicateurs, dont 5 concernent tous les médecins. En plus des 16 indicateurs initiaux, les médecins ont désormais 5 objectifs d'organisation du cabinet – seuls indicateurs concernant tous les médecins –, et 8 indicateurs de qualité supplémentaires. Après 3 ans, la ROSP comporte aujourd'hui 9 indicateurs pour les spécialistes de cardiologie vasculaire et des maladies vasculaires et 8 indicateurs pour les spécialistes de gastroentérologie et d'hépatologie. La dynamique va donc à l'accroissement du nombre d'indicateurs ainsi qu'à l'accroissement du nombre de médecins concernés.

La dynamique du P4P médicale en France doit être rapprochée de l'évolution de la politique publique. En effet, la tendance à l'élargissement du P4P médicale s'inscrit totalement dans la régulation par les pratiques que cherche à imposer la puissance publique. En ce sens, le P4P médicale ne peut pas être analysé en dehors de ses évolutions historiques : il constitue un outil de la reprise de contrôle du régulateur sur la profession médicale.

Encadré 6.5. Le ROSP comme troisième pilier de la rémunération des médecins libéraux français

Extraits de l'article 25 de la Convention médicale (2011, p. 33-34)

« Les partenaires conventionnels conviennent de la nécessité de faire évoluer les modes de rémunération afin de valoriser la qualité des pratiques et d'optimiser les dépenses de santé. La diversification des modes de rémunération des médecins constitue un objectif majeur de la présente convention médicale. Les parties signataires souhaitent évoluer vers une rémunération mixte maintenant le caractère principal du paiement des actes cliniques ou techniques mais prenant aussi en compte les missions des médecins et leurs engagements de service, ainsi que les résultats obtenus en santé publique et en termes d'efficience du système de soins.

Dans ce cadre, elles conviennent d'organiser désormais la rémunération autour de trois piliers :

(i) La rémunération à l'acte, principe fondateur de l'exercice libéral ; dans l'attente de la mise en place de la CCAM clinique d'ici le 31 décembre 2012, des consultations à haute valeur ajoutée de santé publique sont instituées ;

(ii) La rémunération forfaitaire permettant de rémunérer certaines activités correspondant à l'engagement des professionnels dans des domaines comme la prise en charge ou l'accompagnement de pathologies chroniques ou la prise en compte de tâches administratives,

(iii) La rémunération à la performance, versée en fonction de l'atteinte d'objectifs de santé publique et d'efficience, destinée à se développer pour l'ensemble des spécialités et pour l'ensemble des médecins qui le souhaitent.

La présente convention renforce les incitations à la modernisation des cabinets médicaux, à l'utilisation des télé-services et à l'optimisation de l'accueil des patients. Elle valorise dans ce cadre l'engagement des médecins dans la télétransmission des données de facturation qui constitue un service rendu aux patients et une simplification administrative pour les assurés. Elle permet de reconnaître la mission de synthèse de la situation médicale des patients réalisée par le médecin traitant à partir des dossiers informatisés, en diversifiant la rémunération de ces médecins traitants de façon proportionnelle au nombre des patients qui les ont choisi. La rémunération à la performance valorise de plus l'engagement des médecins traitants dans la prévention, le suivi des pathologies chroniques et l'efficience des prescriptions. Ces volets sont susceptibles de s'étendre aux autres spécialités concernées. »

Conclusion

Depuis la fin du XIXe siècle, en France, le champ de la santé est le témoin d'une opposition structurelle entre deux principes : celui de solidarité nationale contre celui de la responsabilisation individuelle devant la maladie. Plus qu'un « contexte », ce processus historique est à l'origine de la création d'institutions aux objectifs contradictoires. Après quelques tentatives malheureuses ou restée marginales, c'est au début du XXème siècle que naît l'idée de solidarité nationale, dans le champ de la santé, avec le vote des lois d'Assurances sociales entre 1928 et 1930. L'État social commence à s'étendre au-delà de ses anciennes prérogatives et s'intéresse à la santé des citoyens. En voulant organiser collectivement le financement et la dispense des soins, l'État social naissant fait face à ce qui peut être considéré comme le problème majeur de l'assurance santé publique française : comment s'assurer de la coopération de la profession médicale avec l'assureur publique ?

L'originalité de la médecine est que la profession médicale française s'est construite contre le principe de socialisation de la santé – présenté comme contraire à la dispense de soins de qualité car pervertissant la liberté du médecin. La Charte de la médecine libérale de 1927 est le symbole de ce refus. Du point de vue de la CSMF, la relation de soin doit être un « colloque singulier » – c'est-à-dire une relation bilatérale entre le patient et son médecin sans l'intervention de tiers. L'État social, parce qu'il cherche à imposer des conditions financières à la relation patient/médecin, est alors un ennemi pour la profession médicale.

Evidemment, cette présentation occulte l'hétérogénéité de la profession médicale. Si les médecins exerçant dans les grandes villes ne veulent pas d'assurance sociale car ils y trouvent un patientèle riche, les médecins exerçant en campagne attendent avec impatience la généralisation de l'assurance publique – celle-ci assurant la solvabilisation de la demande. Mais, les médecins les plus prestigieux et les plus charismatiques – professeurs de médecine, médecins urbains, médecins spécialistes – sont aussi ceux qui parviennent à imposer leur vision libérale de la profession.

Cette histoire de la médecine libérale est, selon nous, le canevas général autour duquel se posent tous les problèmes dans le domaine de la santé depuis la fin du XIXe siècle jusqu'à aujourd'hui. Il faut ainsi impérativement se garder de l'illusion de modernité concernant les débats contemporains. Déficits de la Sécurité sociale, abus des malades d'une couverture sociale généreuse, dépassement d'honoraires, etc. Tous ces thèmes ne sont pas neufs, ce qui est neuf, c'est l'attention que l'on leur accorde. La notion de ticket modérateur est introduite dès 1921 – non pas afin de réduire les dépenses de santé mais dans le but de contrer l'opportunisme présumé des patients. Dès 1949, l'Assemblée Nationale débat sur ce qu'il convient de faire du déficit de la Sécurité sociale. Depuis les débuts de l'assurance publique, les médecins pratiquent ce que l'on n'appelle pas encore les dépassements d'honoraires : l'absence de tarif opposable avant 1960 implique nécessairement un écart entre ce qui est remboursé par l'assureur public et ce qui est payé par le patient – écart supérieur au ticket modérateur prévu par les textes.

Comment explique-t-on que le cadre général du débat – combat entre solidarité nationale et responsabilisation individuelle – demeure relativement stable alors que les réponses apportées évoluent dans le temps ? Qu'est-ce qui change entre le débat parlementaire de 1949 sur le déficit de la Sécurité sociale et aujourd'hui pour que dans un cas soit décidée la hausse des cotisations – donc de la socialisation – et que dans l'autre soit décidée la maîtrise

des dépenses par la désocialisation? La principale modification relève du rapport de force : les décisions ne sont pas les mêmes à la fin du XIXe siècle, au sortir de la Seconde Guerre mondiale, après la naissance de la Ve République et depuis la crise permanente des années 1970. Bien évidemment, ces repères chronologiques (imparfaits) sont des marqueurs de l'évolution du capitalisme français : concurrentiel, fordiste et postfordiste.

La couleur des décisions prises par l'État ne peut donc pas être désencastrée des rapports de force consécutifs, notamment, de l'évolution du capitalisme. On peut alors risquer de dire que lorsque l'État social veut, il peut. En effet, lorsque l'on observe le chemin parcouru entre 1893 et 1945, dans un pays moins riche qu'aujourd'hui et détruit par deux guerres, on ne peut que s'étonner des discours sur les difficultés de régulation de la médecine contemporaine.

Cependant, l'histoire ne se répète pas. Le régime d'accumulation postfordiste actuel n'est pas le régime d'accumulation concurrentiel du XIXème siècle. Le système de soins a également fortement évolué et il n'est plus question d'un retour en arrière complet dans la régulation de la médecine libérale. Si la politique publique a abandonné en partie l'ambition d'une régulation par les prix – ce qui peut s'interpréter comme un retour au principe d'entente directe –, les instruments de régulation se sont renouvelés.

À partir du début des années 1990, la puissance publique va en effet chercher à amorcer une régulation par les pratiques. L'idée est de prescrire le travail du médecin pour s'assurer d'une pratique économe en ressources et de qualité. C'est le second tournant de la médecine libérale. Les dispositifs de paiement à la performance, qui se développent en France et dans le monde, sont un outil parmi d'autres (références médicales opposables, recommandations de bonne pratique, agences indépendantes, démarche qualité, accréditation, certification, etc.) de cet objectif de politique publique dans le capitalisme postfordiste.

Cette transformation de la régulation de la médecine libérale n'en est qu'à ses débuts mais elle enseigne quelque chose d'important sur le syndicalisme médical dominant. Il semble que, de son point de vue, toutes les libertés ne se valent pas. En effet, si la Charte de la médecine libérale de 1927 défend le financement des soins par le principe d'entente directe, elle comporte également les principes de liberté thérapeutique et de liberté de prescription. Or, avec la normalisation et, plus encore, avec l'introduction de la Rémunération sur objectif de santé publique (ROSP), les médecins se voient contraindre certaines prescriptions (pour promouvoir les génériques) et certaines pratiques thérapeutiques (prévention et suivi des

maladies chroniques) – qui sont par ailleurs contestées par certains professionnels[113]. Le contraste entre le combat farouche sur la question de l'entente directe et la discrétion sur la remise en cause de la liberté de pratiques est frappant.

[113] Cf. Chapitre 5.

Chapitre 7

La qualité des soins est-elle soluble dans la quantification ?

Introduction

Le chapitre 6 a établi le cadre général dans lequel s'insère la nouvelle régulation de la médecine libérale. D'une régulation traditionnelle par les prix, la politique publique opère désormais une régulation par les pratiques – dont le paiement à la performance (P4P) médicale n'est qu'un outil parmi d'autres. Le P4P n'a pas de valeur en soi – au sens où il ne permet pas à lui seul de saisir ce qu'il se passe dans le secteur des soins ambulatoires. Le cadre d'analyse doit nécessairement être plus large. Dans le capitalisme postfordiste, les anciennes alliances se dénouent et la solidarité nationale s'érode au profit du secteur privé et de la mise à contribution du patient, qu'il convient de responsabiliser.

Du côté de l'offre de soins, la puissance publique abandonne l'instrument des prix (tarif opposable, enveloppe globale, etc.) laissant ainsi une importante liberté tarifaire pour une grande partie de la profession – mais pas toute –, et elle se concentre sur la régulation par les pratiques. Systèmes d'information, normes de bonnes pratiques, références médicales opposables, etc. viennent prescrire le travail médical, auparavant caractérisé par une large marge de liberté. Dans cette optique, le P4P n'est qu'un aboutissement de la logique visant à contrôler le travail du médecin. L'idée permettant de justifier ce mouvement normalisateur est que la faible qualité des soins et les fortes dépenses de santé pourraient être contrecarrées par le suivi de normes quantifiées.

Contrôler les médecins est un objectif légitime de la puissance publique qui est le payeur principal des soins. Faire porter l'essentiel du contrôle du médecin sur les pratiques, en oubliant les prix (qui se libéralisent du fait de la prolifération des dépassements d'honoraires), en sacrifiant l'objectif de socialisation (du fait de l'augmentation des restes à charge) et en négligeant le maillage sanitaire du territoire (du fait de l'accentuation des déserts médicaux) est critiquable et souvent critiqué comme on l'a montré dans le chapitre précédent. Le contrôle des pratiques est alors présenté comme une victoire sur la médecine à identité

libérale, dans un contexte où elle se conforte en matière de liberté d'installation et se renforce en matière de liberté tarifaire. Dans cette perspective, le contrôle des prescriptions fait l'objet d'un large consensus et est présenté comme un succès d'autant plus appréciable qu'ils sont rares. Combattre l'hétérogénéité des pratiques, c'est réduire les marges de manœuvre du médecin.

Ce chapitre questionne la qualité de cette régulation. La façon dont s'organise le contrôle des pratiques en médecine ne va pas de soi et induit une nouvelle représentation du travail médical. La régulation des pratiques fait émerger une nouvelle rationalité, à laquelle les médecins peuvent s'opposer du fait « d'*a priori* » ou, plus généralement, d'une représentation différente de ce que doit être leur travail (Clerc et al., 2011a). La littérature en économie de la santé a montré à quel point la tentative d'imposer des règles de bonnes pratiques portait un conflit sur la représentation du travail du médecin et pouvait conduire à des attitudes négatives envers la tentative publique de régulation des pratiques. Or, les attitudes des professionnels en matière de règles de bonnes pratiques peuvent affecter la connaissance de ces bonnes pratiques et plaider pour une éducation ou une formation des médecins à leur usage (Clerc et al., 2010, 2011b).

Nous proposons une lecture de ce conflit en prenant comme fil directeur la notion de quantification : les règles de bonnes pratiques sont des critères mesurables de résultat. Le conflit sur la représentation du travail des médecins n'est pas un conflit entre la puissance publique et les professionnels. Il brouille les frontières. Ce sont des groupes particuliers de médecins qui portent les bonnes pratiques et une conception protocolisée de la médecine contre d'autres médecins partisans d'un « art médical », laissant une grande place à l'autonomie. C'est une représentation particulière de l'État, venant du *New Public Management* qui soutient la tentative de standardisation de la médecine contre une autre conception de l'État social.

Notre analyse part du constat que cette régulation des pratiques prend appui sur une vaste entreprise de quantification de la qualité des soins. Or, la qualité n'est pas soluble dans la quantification. Pour le comprendre, il faut revenir aux racines de cette nouvelle rationalité dans le secteur médical. Comment s'est imposée l'idée que le secteur des soins est… un secteur comme un autre qui peut faire l'objet d'une « rationalisation » ? En effet, la normalisation d'un travail pour mieux le contrôler n'est pas une idée neuve. Le fordisme, l'organisation scientifique du travail ou plus généralement les diverses « démarches qualité »

existent depuis bien longtemps dans le secteur industriel. La nouveauté provient de la propagation de ces idées dans le champ de la médecine ambulatoire.

En prolongement du chapitre 6, il s'agit ici d'approfondir l'étude de la période postfordiste durant laquelle la nouvelle rationalité dans le champ médical a émergé. Cette période est marquée par une triple crise de l'autorité médicale. D'abord, la crise économique a pour effet de mettre les dépenses socialisées de santé au même niveau que les autres – le médecin devant alors subir, comme d'autres acteurs de la protection sociale, l'ajustement des budgets sociaux. Ensuite, la multiplication des crises sanitaires (comme celle du sang contaminé) contribuent également à la perte d'autorité médicale. Enfin, le développement du Sida renverse le paternalisme médical en montrant que le malade peut être autant expert que le médecin de sa propre maladie. Ce climat de défiance ne se suffit pas de lui-même à faire émerger un nouveau monde. Il faut un ensemble d'idées pour que l'organisation de la délivrance des soins ambulatoire évolue. Sans ces idées nouvelles, sans solution de rechange, le système de santé ne peut pas se transformer. Deux courants majeurs ont nourri la nouvelle rationalité du champ médical : le *New Public Management* (NPM) et l'*Evidence Based Medicine* (EBM).

Le courant du NPM est en effet l'une des sources contemporaines les plus importantes inspirant la « réforme » de l'État dont l'objectif est de renouveler l'action publique. Ce renouvellement est présenté comme inéluctable et salvateur tant le modèle de gestion bureaucratique antérieur semble dépassé car inefficace, dispendieux et démotivant. Dans le secteur des soins, se développe alors l'idée qu'il faut « rationnaliser » l'usage des ressources notamment en adoptant une « démarche qualité » ayant déjà fait ses preuves dans le secteur industriel. Il faut normaliser et quantifier le travail du médecin afin de contrôler au plus près l'usage des ressources publiques. Le pouvoir de contrôle du médecin est d'ailleurs retiré des mains publiques traditionnelles pour être remis à un nouveau type d'institution : les agences de contrôles dites indépendantes telles que la Haute Autorité de Santé.

Pour que la quantification de la qualité des soins soit possible, il faut par ailleurs produire les normes et les certifier. Or, traditionnellement la profession médicale s'est construite autour de la notion assez vague d'« art médical ». L'« art médical » n'est pas un ensemble de normes écrites et opposables à partir desquelles il serait possible de construire et de diffuser des normes quantifiées. L'« art médical » se rapproche plus d'un processus de production artisanal où la singularité du médecin qui agit implique nécessairement des

variations dans le traitement de mêmes pathologies. L'expérience et le charisme du médecin sont bien plus valorisés que l'application d'une norme extérieure à la relation de soin.

En conséquence, l'application du NPM en santé doit nécessairement s'accompagner d'une transformation au sein même du champ médical – ce qui s'est produit avec la prise de pouvoir de la médecine fondée sur les preuves (EBM). La médecine fondée sur les preuves constitue un changement paradigmatique dans la recherche médicale qui met désormais le critère statistique au sommet de la hiérarchie des preuves. Plutôt que de s'attarder sur la singularité des cas pour comprendre l'effet du traitement, il s'agit maintenant, par l'intermédiaire des essais cliniques randomisés, de tester la significativité statistique du traitement.

La conjonction de la réforme de l'État impulsée par le NPM et de l'évolution de la hiérarchie des preuves due au développement de l'EBM a rendu possible aujourd'hui un degré de normalisation et de quantification de la qualité des soins jamais atteint auparavant. Normes de bonnes pratiques, recommandation de bonnes pratiques, références médicales opposables, etc. viennent désormais encadrer la pratique médicale par des standards chiffrés. C'est à partir de ses standards chiffrés qu'est née la possibilité de mettre en place le P4P médicale. Pour rémunérer à la performance il faut en effet être capable de produire des indicateurs/objectifs à atteindre – ce que fait le ROSP lorsqu'il énonce qu'il faut que 80 % des patients diabétiques de type 2 aient des résultats de dosages d'HbA1c inférieurs à 7,5 %.

Cependant, la volonté de quantifier la qualité des soins ne va pas sans problèmes. Deux types de problèmes peuvent se poser selon que l'on se place du point de vue de la production de la norme médicale ou du point de vue de l'économie des services.

Du point de vue de la production de la norme médicale, c'est la définition de la maladie qui va être bousculée. Est-elle une variation quantitative ou une rupture qualitative de l'état normal ? La prétention à l'objectivité de la quantification reste vulnérable au contexte sociohistorique et aux forces politiques notamment le poids de l'industrie pharmaceutique – qui interdisent de prendre la norme pour une vérité transcendante. Du point de vue de l'économie des services, depuis longtemps nous savons que le service n'est pas un produit et que, en conséquence, c'est une erreur que de calquer aux services les méthodes de la production industrielle (Gadrey, 1991). Dans le champ médical comme dans les services, on ne peut pas identifier ce qui relève de la production d'un producteur. La qualité des soins et l'état de santé sont le résultat d'une production collective dont il est douteux de donner

l'entière responsabilité au médecin. Le rémunérer *via* le dispositif de P4P médicale sur la base d'un résultat dont il ne porte pas l'entière responsabilité atteint ses limites.

La première section présente les éléments fondamentaux de la nouvelle rationalité qui va émerger après 1970 dans le champ médical. NPM et EBM vont se nourrir mutuellement pour rendre légitime l'ambition de quantifier la qualité des soins (1.). La seconde section interroge cette nouvelle rationalité en s'appuyant sur les débats épistémologiques au sein même du champ médical ainsi que sur les travaux de l'économie hétérodoxe des services (2.).

1. La nouvelle rationalité : produire de la qualité par les quantités

Les transformations dans le secteur de la santé ne sont pas spécifiques. Le pilotage de l'action publique par des indicateurs chiffrés est une méthode de gouvernement issue de l'hégémonie des théories regroupées sous le vocable de *New Public Management* (NPM). En ce sens, l'émergence du contrôle des pratiques en médecine libérale n'est qu'une application du thème plus général du NPM (1.1.). Mais, pour parvenir à mettre en place ce mode de régulation dans la médecine, il faut être en mesure de produire des indicateurs quantifiés et certifiés de l'activité médicale (1.2.). Enfin, pour que cette nouvelle régulation ne soit pas que de l'encre sur du papier, l'évolution continue des ressorts de la responsabilité médicale face à ces nouvelles normes est impérative (1.3.). La définition de la bonne médecine qui était la prérogative des médecins va peu à peu devenir celle de la puissance publique. Dans un premier temps, ce qui change ce n'est donc pas tant la rationalité du médecin – pas encore – mais la rationalité des pouvoirs publics.

1.1. Du *New Public Management* à la reprise en main nécessaire de la médecine libérale

Les transformations en profondeur de la régulation de la pratique médicale en France à partir du début des années 1990 ne sont pas isolées de toutes influences extérieures et/ou rendues évidente par une organisation des soins primaires particulière. Bien au contraire, ce qui se passe pour la médecine libérale française, n'est qu'une traduction d'un mouvement plus général d'ampleur mondiale : l'éclosion et le développement de la Nouvelle Gestion Publique (*New Public Management*). Il convient alors de s'interroger sur ce qu'est le NPM et la façon dont il problématise les enjeux propres de la médecine.

1.1.1. Le New Public Management : il n'existe pas mais il est partout

Le NPM est un mouvement de réorganisation radicale de l'État et de ses administrations qui a fait son apparition à partir du début des années 1980. Si l'on parle le plus souvent de NPM au singulier, la littérature met cependant en évidence l'impossibilité à cerner le concept dans des termes consensuels tant il s'adapte à des contextes nationaux ou sectoriels – créant par la même une pluralité de situations interdisant une quelconque synthèse.

Par exemple, dans un article de 2001, Pollitt (2001) fait le constat de l'existence d'une grande variété de définition du NPM – presque autant que d'auteurs – sans toutefois parvenir à trouver des définitions identiques. Chacun souligne l'aspect du NPM qui lui semble le plus saillant sans qu'il n'y ait d'accord sur un trait unificateur. L'une des raisons principales expliquant l'absence de définition consensuelle tient au fait que le NPM a pour caractéristique de se fondre aux contextes locaux de telle sorte qu'il apparaît souvent insaisissable. Cette capacité à s'adapter engendre même certaines incohérences rendant l'analyse plus difficile encore. Par exemple, tandis que dans certains cas l'autonomisation des agents est présentée comme un atout pour améliorer leur l'efficacité, dans d'autres, ce sont les marges de liberté qu'il faut réduire afin de se prémunir de l'opportunisme naturel des fonctionnaires.

Selon Hood et Peters (2004), si les premiers écrits du NPM datent du début des années 1980, aucun livre ne peut prétendre au statut de traité théorique général de cette nouvelle approche de la gestion publique. Le pluralisme des approches est tel que, selon les auteurs, nous nous situerions encore au Moyen-Âge du *New Public Management*. C'est dans cette optique que nous proposons de reprendre pour définition du NPM la présentation extensive qu'en fait Philippe Bezes (2005) :

> *« Le statut du NPM est ambigu. Il s'agit d'un ensemble hétérogène d'axiomes tirés de théories économiques, de prescriptions issues de savoirs de management, de descriptions pratiques expérimentés dans les réformes (notamment dans les pays anglo-saxons) et de rationalisations doctrinales réalisées par des organisations transnationales (OCDE, Banque mondiale, etc.). Le New Public Management constitue donc largement un puzzle doctrinal à vocation générique, c'est-à-dire susceptible d'être appliqué à tous les services administratifs quels qu'ils soient, et alimentant des conceptions de réformes multiples et parfois contradictoires » (Bezes, 2005, p. 26).*

Pour illustrer la vocation générique du NPM, autrement dit son ambition d'universalité, il est possible de se référer à la cartographie réalisée par Hood (1995). L'objectif de l'auteur est, d'une part, de présenter la diffusion internationale des principes du NPM dès le début des années 1990, et, d'autre part, de montrer que ces principes sont partagés par les différentes tendances politiques au sein des pays de l'OCDE. Le NPM ne serait « ni de droite, ni de

gauche ». Le tableau 7.1., repris de Hood, présente l'importance des réformes inspirées du NPM dans certains pays de l'OCDE en rapport avec la « couleur politique » de la majorité alors en place. Il est remarquable de constater que, selon cette cartographie, le NPM n'a pas de frontières politique et géographique – ce qui explique sans doute la force avec laquelle il s'est imposé de telle sorte qu'aujourd'hui la quasi-intégralité des pays mentionnés par Hood en 1995 pourraient prétendre à figurer dans la classification « fort recours au NPM » [114].

Tableau 7.1. Adoption du NPM en fonction de la couleur politique dans les pays de l'OCDE

	Political incumbency emphasis		
NPM Emphasis	« Left »	« Centre »	« Right »
Hight	Sweden	Australia Canada New Zeeland	United Kingdom
Medium	France	Austria Denmark Finland Italy Netherlands Portugal USA	
Low	Greece Spain	Germany Switzerland	Japan Turkey

Source : Hood (1995, p. 100)

Le NPM serait en somme une dynamique réformatrice sans frontières cherchant à renouveler la rationalité propre de l'action publique. En ce sens, s'il y a nécessité de renouvèlement, c'est en rapport avec une tradition qui perdrait sa légitimité. Le nouveau suppose la présence d'un ancien qu'il faut combattre. Avant de présenter ce qu'il y a de nouveau, il faut donc décrire l'ancien : le modèle bureaucratique de l'action publique.

1.1.2. Rompre avec le modèle bureaucratique de l'action publique

La littérature attribue la première théorisation du modèle bureaucratique de l'action publique à Max Weber (1971) – c'est pourquoi l'on parle également de modèle weberien.

[114] La cartographie de Hood (1995) pourrait être légitimement critiquée sur de nombreux points importants (classification droite/gauche, degré d'adoption des principes du NPM, dilution du rôle de pays moteur comme le Royaume-Uni et les États-Unis, classification des États-Unis au « centre », etc.). Cependant, notre objectif ici n'est pas tant de discuter de la validité de cette cartographie que de la prendre comme un exemple d'une idée commune liée à notre thème de recherche : le NPM n'aurait pas de frontières parce qu'il serait dépouillé d'idéologie et fondé sur des critères neutres et objectifs.

Mais, il faut d'emblée préciser que cette influence de Weber ne se limite pas à l'action publique. Pour l'auteur allemand, si l'administration bureaucratique est la pratique de domination la plus rationnel du point de vue formel, elle n'est en rien l'apanage de l'État et de ses satellites : l'église, l'armée, le parti, l'entreprise, l'association, etc. sont également susceptibles d'adopter une administration bureaucratique. Pour Weber, cet idéal type n'a d'ailleurs pas la connotation péjorative qu'on lui donne aujourd'hui : au contraire, la bureaucratie est synonyme de la modernité capitaliste.

L'œuvre de Weber n'est pas la seule dans laquelle il est possible de puiser les fondements de l'organisation bureaucratique. D'autres auteurs classiques comme Woodorw Wilson, Friedrich Taylor ou Henri Fayol ont eux-aussi, en dépit de leurs fortes divergences, contribué à définir l'organisation bureaucratique. Par exemple, l'objectif de Wilson aux États-Unis est, de façon plus particulière que Weber, de poser les bases de l'administration de la chose publique. De la même façon, l'organisation scientifique du travail et le *one best way* de Talyor, comme l'administration par unité de commandement de Fayol, sont constitutifs d'une façon de penser l'organisation bureaucratique. Il est possible de synthétiser ce modèle, que l'on appelle alternativement classique ou ancien, en plusieurs principes d'organisation (Bartoli, 2005, Santo et Verrier, 2007).

Le premier principe du modèle bureaucratique est la séparation du pouvoir politique – qui adopte et vote les mesures - et du pouvoir administratif – qui les met en œuvre. L'objectif est alors de lutter contre l'abus de pouvoir, la corruption et le favoritisme consubstantiels à une organisation où celui qui décide de la règle est également celui qui l'applique. Les charges administratives ne doivent pas être personnalisées de façon à séparer les droits et les moyens personnels de l'activité professionnelle. Il faut garantir l'impersonnalité des relations et l'impartialité des agents. Pour y parvenir, l'État doit définir des statuts précis pour ses agents avec des obligations de services statutaires associés. Cette dissociation du politique et de l'administratif permet également de professionnaliser l'administratif, ce dernier n'étant plus sujet aux rotations issues des élections. Les agents désormais assurés de conserver leur poste développent des compétences spécialisées. Cette dépolitisation de l'action publique vise à favoriser une meilleure efficacité organisationnelle.

Un deuxième principe important de la bureaucratie porte sur le recrutement et la rémunération des agents. Le recrutement se réalise sur la base des qualifications professionnelles des agents. Afin d'éviter le favoritisme, des concours permettent de dépersonnaliser le recrutement. En outre, pour certains secteurs, des écoles spécialisées

peuvent prendre en charge la formation d'élite administrative. Concernant les rémunérations, celles-ci sont fonction du niveau de qualification initial et du degré de responsabilité professionnel. L'avancement est fonction de l'ancienneté et, le cas échéant, de nouvelles qualifications ou responsabilités – rendues possibles par une formation professionnelle continue. Encore une fois, l'objectif principal est de lutter contre le favoritisme et de cultiver un mérite d'ordre académique et professionnel – plutôt que l'arbitraire des chefferies. Cette organisation du recrutement permet également de disposer d'un personnel ayant une qualification et une culture homogène – dépersonnalisant à l'extrême les relations au sein de la bureaucratie. Par l'avancement à l'ancienneté, le modèle bureaucratique valorise l'expérience, la loyauté et la fidélité à l'administration.

Enfin, un dernier principe fondamental des bureaucraties classiques, est l'organisation centralisée, hiérarchique à structure pyramidale. La division du travail est maximale avec notamment la création de corps distinctifs. La structure pyramidale permet d'assurer la rapidité d'exécution des ordres et de maîtriser la chaîne des responsabilités. Les règles et procédures doivent être stables et explicites pour exploiter au mieux les économies de spécialisation. Dans le modèle bureaucratique, les avantages de la spécialisation sont en effet maximaux et ils assurent l'homogénéité du service rendu tel qu'il est défini par la loi.

En résumé, quelles que soit ses variantes, la spécificité de la bureaucratie réside dans le formalisme réglementaire strict, la structuration hiérarchique, la spécialisation des fonctions et la centralisation des pouvoirs de décision. Ce sont contre ses caractéristiques que vont s'élever les tenants d'un renouvellement de l'action publique au début des années 1980. L'organisation bureaucratique serait en effet la cause de lourdeurs, de rigidités, de démotivation et de désincitation des agents qu'il conviendrait de dépasser.

1.1.3. Les quatre formes du NPM

Les travaux autour du NPM semblent s'accorder pour dire qu'il faut reprendre le contrôle sur la bureaucratie. Les « dix commandements » d'Osborne et Gaebler (1992) sont à ce titre une illustration célèbre parmi d'autres (Encadré 7.1.). C'est sur la pluralité du sens de cette reprise en main qu'il est possible d'identifier les différentes influences – parfois contradictoires – du NPM. Philippe Bezes (2005) distingue ainsi quatre formes du NPM :

- L'amincissement de l'État :

Si aujourd'hui l'un des principaux slogans du NPM est « pas moins d'État mais mieux d'État », il est nécessaire de rappeler que la première vocation du NPM est l'amincissement de l'État. Avec la révolte fiscale et l'arrivée de gouvernements néo-conservateurs au pouvoir dès la fin des années 1970, le mot d'ordre était à la réduction de la dépense publique et à la privatisation. Ce que l'État fait, le privé pourrait le faire mieux et pour moins cher. Les réductions de coûts et des effectifs sont censés permettre de rapprocher l'État de ses « cœurs de métier » – la légitimité de la dépense publique étant constamment remise en cause même là où quelques semaines auparavant elle paraissait naturelle. Les externalisations et débureaucratisations par privatisation ou mise en concurrence des activités de l'État sont donc la première pierre du NPM.

- L'autonomisation de la gestion des services de l'État :

Une deuxième forme importante du NPM porte sur l'autonomisation de la gestion des services de l'État. Dans ce cadre, le modèle bureaucratique est jugé lourd et sclérosé. Le bon fonctionnement ne pouvant plus être assuré, il faudrait libérer les services de l'État de ses lourdeurs hiérarchiques. Pour cela, le NPM préconise le renforcement de l'autonomie de gestion des services et l'allègement des formes de contrôle. En contrepartie, l'État central doit mettre en place des procédures de « redevabilité » (*accountability*). Plutôt que de décider de tout, la hiérarchie ne doit que vérifier en aval le travail de ses services ainsi décentralisés. Inspirés des méthodes de gestion privée, les tenants du NPM n'en reconnaissent pas moins l'originalité du secteur public. En conséquence, plutôt que de rendre des comptes sur des ratios financiers, les administrations se voient de plus en plus évaluées sur des objectifs de qualité des services et de satisfaction de l'usager – désormais au cœur des réformes.

- Le contrôle des agents de l'État :

Une troisième forme du NPM se focalise quant à elle sur l'impérieuse nécessité du contrôle des agents. Sous l'influence des théories économiques *mainstream*[115] (école du *Public Choice*, économie des coûts de transaction, théorie de l'agence), la bureaucratie va peu à peu être vue comme une masse d'agents opportunistes agissant pour leur propre compte – donc ni pour celui des politiques élus, ni pour celui des citoyens usagers. Du fait des asymétries d'information, l'administration serait parvenue à imposer ses propres intérêts particuliers aux démocraties occidentales. L'autonomie des bureaucraties est alors un danger générant l'extension sans fin des budgets et des difficultés pour les politiques à appliquer les

[115] Cf. Première partie.

programmes pour lesquels ils ont été élus. Le contrôle de la bureaucratie devient un enjeu central. Pour résoudre ces problèmes il faudrait alors procéder à la contractualisation entre niveaux hiérarchiques afin d'expliciter et de contrôler les objectifs des fonctionnaires.

- La gestion des services et des agents de l'État par les indicateurs :

La dernière forme du NPM repose sur une vision dite rationnelle et gestionnaire du contrôle des organisations publiques. Le contrôle opère ici par la production et le traitement de l'information (Ogien, 1995). Dans cette optique, l'État devrait s'inspirer des pratiques de contrôle de gestion des grandes firmes multinationales. Celles-ci étant désormais constituées de nombreuses unités décentralisées (entreprises multi-divisionnelles), elles sont semblables à un État et ses diverses administrations. Or, depuis la fin du XXème siècle, les firmes multi-divisionnelles ont mis en place un système de contrôle à distance par les résultats financiers des filiales. Pour ce faire, il est nécessaire d'introduire des systèmes d'information permettant de recueillir et d'analyser les indicateurs. Cette perspective permet d'approfondir la logique de contrats d'objectifs déjà évoquée. L'État doit favoriser le développement de la comptabilité analytique, de systèmes de mesure de la performance des agents, de comparaison par *benchmarking*, de paiement à la performance, de chartes des clients ou d'usagers, de standards de qualité, etc.

Avant de voir comment ce mouvement a été importé en France, on peut déjà souligner l'importance du thème de la gestion de la main d'œuvre dans le NPM. Au fond, quelles que soient les sources théoriques, l'idée générale est de mettre les agents de l'État au travail – qu'ils y aient été éloignés par leur opportunisme naturel ou par la rigidité des procédures bureaucratiques. La question qui se pose est donc une question de gestion de ressources humaines assez banale : comment s'assurer de la discipline des salariés ? La réponse qui semble esquissée nous est familière : la discipline de marché ne doit pas connaître de limite – quitte à accommoder les institutions hors marché à l'effet prix.

Encadré 7.1. Les dix commandements de la nouvelle gestion publique

Osborne et Gaebler (1992)

Cette littérature s'apparentant plutôt à une succession sans fin de livres et essais sur l'impératif de renouvellement de l'action publique, il est impossible d'identifier un traité de NPM. On peut néanmoins citer à titre d'exemple le livre de David Osborne et Ted Gaebler (1992), Reinventing Governement, qui reste jusqu'à aujourd'hui un des plus grand succès d'édition de la littérature. Il est l'un des premiers à systématiser le nouveau paradigme gouvernemental (Bruno et Didier, 2013). Ce livre se présente sous la forme de 10 commandements à destination des gouvernants – 10 commandement pour dépasser les rigidités bureaucratiques qui enfoncent les États et les économies occidentales dans la crise :

(1) Piloter plutôt que ramer : « gouvernement catalyseur »

(2) Donner la capacité plutôt que servir : « Gouvernement communautaire »

(3) Introduire la concurrence dans les services publics : « Gouvernement compétitif »

(4) Transformer les organisations régies par des règles : « Gouvernement par mission »

(5) Financer les résultats et non les moyens : « Gouvernement à la performance »

(6) Satisfaire les besoins de la clientèle, et non pas de la bureaucratie : « Gouvernement orienté vers les clients »

(7) Gagner de l'argent plutôt que de le dépenser : « Gouvernement entrepreneurial »

(8) Prévenir plutôt que guérir : Gouvernement de précaution »

(9) Faire participer et travailler en équipe plutôt que de hiérarchiser : « Gouvernement décentralisé »

(10) Se servir du marché comme levier de changement : « Gouvernement orienté vers le marché »

1.1.4. Le NPM dans le contexte Français : l'exemple de la LOLF

Comme nous l'avons déjà remarqué le NMP se décline différemment selon les contextes locaux. En France, c'est avec la Loi organique relative aux lois de finance (LOLF), votée en 2001 et appliquée à partir de 2006, que l'on peut voir le plus clairement à l'œuvre la nouvelle rationalité de la gestion publique (Jany-Catrice, 2012b). En effet, la LOLF systématise dans l'administration française le recours à des techniques d'évaluation quantitatives et de pilotage par les indicateurs. D'une logique de moyens, l'État s'oblige à des résultats. Plus pratiquement, la LOLF modifie les modalités de présentation et de contrôle des lois de finances votées par le Parlement.

Les réformateurs ont mis en évidence deux objectifs principaux pour légitimer ce changement paradigmatique des finances publiques françaises. D'une part, la LOLF renforcerait le pourvoir des Parlementaires (députés et sénateurs) dans la mesure où, maintenant, toutes les lois de finances doivent être votées par eux. D'autre part, la LOLF aurait une légitimité démocratique puisqu'elle fait rentrer l'État dans l'ère de la performance – les nouvelles procédures de quantification ayant pour objectif d'évaluer constamment l'activité publique. Alors que l'attribution des budgets se faisait antérieurement ministère par ministère et type de dépense par type de dépense, depuis 2006, l'État français ne finance que

des missions – dans lesquelles on trouve un niveau plus détaillé des programmes, des objectifs et des indicateurs[116].

Le tableau 7.2., réalisé à partir du *Rapport sur l'évolution de l'économie nationale et sur les orientations des fiances publique. Tome 2* (Performance publique, 2014), présente ainsi le financement de la mission Santé. On observe que la mission santé se subdivise pour 2014 en 2 programmes, 4 objectifs et 13 indicateurs. Il est d'ailleurs notable que l'un des indicateurs soit proche des indicateurs de performance du ROSP : Programme 204, Objectif 1, Indicateur 1.2. Taux de participation au dépistage organisé du cancer du sein pour les femmes de 50 à 74 ans. Pour 2014, les médecins français participent ainsi – *via* le ROSP – à l'évaluation de la mission Santé de l'État.

En 2007, le chemin ouvert par la LOLF a été approfondi par la Révision générale des politiques publiques (RGPP). Les six axes de « modernisation » des services publics présentés dans le rapport d'étape publié en 2008 nous permettent de constater la continuité de la réforme menée par le NPM en France : adapter les missions de l'État aux défis du XXème siècle, (ii) Améliorer les services pour les citoyens et les entreprises, (iii) Moderniser et simplifier l'État dans son organisation et ses processus, (iv) Valoriser le travail et le parcours des agents, (v) Responsabiliser par la culture des résultats, (vi) Rétablir l'équilibre des comptes publics et garantir le bon usage de chaque euros. Comme le défendais déjà Hood en 1995, cette dynamique résiste bien à l'« alternance » politique puisque depuis 2012 et le changement formel de majorité, la RGPP a été remplacée par la Modernisation de l'action publique (MAP) – sans que ne change l'esprit devant animer l'action publique.

[116] Chaque indicateur peut par ailleurs adopter un point de vue potentiellement différent : celui du citoyen (efficacité de l'État qui fait ce qu'il annonce), celui de l'usager (qualité des services rendus) et celui du contribuable (efficience de l'administration).

Tableau 7.2. La mission santé de l'État français

Mission Santé		
Programme 204 – Prévention, sécurité sanitaire et offre de soins		
Objectif 1. Améliorer l'état de santé de la population et réduire les inégalités territoriales et sociales de santé		
	Indicateur 1.1.	Proportion de découvertes de séropositivité diagnostiquées à un stade de SIDA
	Indicateur 1.2.	Taux de participation au dépistage organisé du cancer du sein pour les femmes de 50 à 74 ans
	Indicateur 1.3.	Consommation annuelle d'alcool par habitant de plus de 15 ans
	Indicateur 1.4.	Taux de prévalence du tabagisme quotidien chez les jeunes
	Indicateur 1.5.	Nombre annuel de donneurs d'organes décédés et prélevés / nombre annuel de donneurs décédés à l'hôpital et recensés dans Cristal action
	Indicateur 1.6.	Pourcentage des étudiants en médecine s'orientant vers la médecine générale de premier recours
Objectif 2. Prévenir et maîtriser les risques sanitaires		
	Indicateur 2.1.	Pourcentage d'unités de distribution d'eau alimentant en eau potable plus de 5000 habitants présentant de manière récurrente des dépassements des limites de qualité
	Indicateur 2.2.	Nombre de signalements traités en 1h / Nombre total de signaux
	Indicateur 2.3.	Délai de traitement des autorisations de mise sur le marché
Programme 183 – Protection maladie		
Objectif 1. Assurer la délivrance de l'aide médicale de l'État dans des conditions appropriées de délais et de contrôles		
	Indicateur 1.1.	Délai moyen d'attribution de l'AME
	Indicateur 1.2.	Contrôle des ressources déclarées des bénéficiaires de l'AME
Objectif 2. Réduire les délais de présentation et de paiement des offres d'indemnisation du FIVA		
	Indicateur 2.1.	Pourcentage des offres présentées aux victimes de pathologies graves dans le délai légal de six mois
	Indicateur 2.2.	Pourcentage des offres payées dans le délai réglementaire de deux mois

1.1.5. NPM et médecine libérale : l'exigence du contrôle

À la lecture des paragraphes précédents, il peut naître un double paradoxe au moment de revenir à l'analyse de la médecine libérale française. D'une part, comment se fait-il que l'on cherche à interpréter les évolutions dans le champ de la santé à partir du thème du NPM

et de la critique de la bureaucratie alors que la médecine de ville est presque exclusivement libérale ? D'autre part, quand bien même cette association serait pertinente comment expliquer la confusion entre l'État – produit de la démocratie parlementaire – et la Caisse nationale d'assurance maladie (CNAM) – produit de la démocratie sociale donc en large partie hors du champ de la démocratie parlementaire ? Plus précisément, on comprend que l'État soumette l'hôpital aux principes du NPM car c'est lui qui contrôle la Fonction publique hospitalière. Mais, comment expliquer que la CNAM – qui a la charge de réguler la médecine de ville – choisisse, elle aussi, de se soumettre à ces principes ?

Concernant le premier paradoxe, si le médecin est formellement libéral, son revenu dépend de la solvabilisation de la patientèle et par conséquent de la puissance publique qui fixe les règles du jeu. Nous avons déjà montré comment historiquement, sans la solvabilisation de la demande par la puissance publique, l'activité médicale a eu les plus grandes difficultés à exercer. Le développement de la LOLF se conjugue avec la prise de pouvoir de l'État sur la santé au travers du vote par le parlement de l'Objectif quantifié national d'évolution des dépenses de santé. La prise de contrôle de l'État s'observe aussi dans la multiplication des agences (Agence Régionale de la santé, Haute Autorité de santé) qui portent les prérogatives de la puissance publique (nous y reviendrons). Enfin, le mode de financement de la santé a également évolué dans le sens d'une plus grande étatisation avec la création de la Contribution Sociale Généralisée, mise en place en 1991 et qui, en 1998, se substitue à la cotisation sociale maladie[117].

Concernant le second paradoxe – pourquoi la CNAM s'obligerait-elle à suivre des règles prévues pour l'État ? –, il suffit de se rappeler qu'historiquement la Sécurité sociale, puis ses différentes branches dont la CNAM, ont perdu peu à peu leurs prérogatives face au pouvoir de l'État. Rappelons par exemple que le directeur de la CNAM est nommé en Conseil des ministres – écornant ainsi le principe de démocratie sociale (Hassenteufel et Palier, 2005). De façon plus décisive, il semble que l'existence même de la CNAM dépende de ses capacités à agir de façon à ne pas rendre nécessaire sa reprise en main par l'État. On se rappelle, par exemple, qu'après la création de la Sécurité sociale, les syndicats majoritaires au sein de la FNOSS préféraient contenir les dépenses de santé avec l'appui de l'État plutôt que de

[117] La CSG est fixée à 1,1 % du revenu en 1991, 2,4 % en 1993 puis 3,4 % en 1997. En 1998, son taux passe à 7,5 %, soit une hausse de 4,1 %, tandis que la cotisation sociale maladie sur les revenus d'activité baisse de 4,75 %. Selon les chiffres du Haut conseil pour l'avenir de l'assurance maladie de 2013, l'assurance maladie est financée à hauteur de 48 % par les cotisations sociales et 36 % par la CSG, le reste relevant de taxes diverses (alcool, tabac, etc.).

négocier des tarifs médicaux élevés avec la CSMF – l'enjeu étant la crédibilité de la démocratie sociale à gérer à un budget comparable à celui de l'État[118]. Encore une fois, la médecine de ville semble être, elle-aussi, propice à l'adoption des préceptes du NPM – moyennant des adaptations au contexte sectoriel.

La médecine, à l'hôpital ou à la ville, pose un problème de gestion des ressources humaines à l'État « moderne ». Il faut contrôler l'activité médicale et maîtriser l'opportunisme naturel du médecin. D'après le rapport *Performance, incitation et gestion publique*, réalisé par Dominique Bureau et Michel Mougeot (2007), « la mesure de la performance dans la gestion publique répond à un impératif démocratique » qui n'est en rien étranger au système de santé.

La thèse en Sciences de gestion de Virginie Forest (2008) a déjà montré comment cet impératif démocratique – le NPM – a été difficilement mis en place à l'hôpital public. Contrairement à la médecine de ville, la reprise en main de l'hôpital s'est réalisée par un outil de contrôle des dépenses – la tarification à l'activité. Ici, le problème est assez différent dans la mesure où la régulation par les prix ayant échouée c'est la régulation par les pratiques qui cherche à s'imposer. En ce sens, la ROSP est un outil de contrôle parmi d'autres de la reprise en main de la médecine ambulatoire. Cependant, pour en arriver là – mettre en place un pilotage par les indicateurs –, il faut être capable de produire des indicateurs de performance médicale et être en mesure de les légitimer et de les diffuser.

1.2. Produire et certifier les normes médicales de contrôle : *Evidence Based Medicine* et agences indépendantes

Afin de mettre en place le contrôle de la pratique médicale par les chiffres, il nécessaire de se doter préalablement de nomenclatures permettant de quantifier l'activité du médecin. Si traditionnellement ces nomenclatures n'existent pas, l'émergence et le développement de la médecine fondée sur les preuves fournit aux adeptes du NPM de tels standards.

1.2.1. Produire les normes quantifiées : l'Evidence Based Medicine

Traditionnellement le champ médical était faiblement codifié et formalisé bien que le registre de légitimation soit celui de la « science ». L'apprentissage se faisait en situation, par l'expérience. Cette organisation a longtemps conféré aux médecins les plus prestigieux – les

[118] Cf. Chapitre 6.

dits mandarins – un pouvoir sur les autres membres de la profession, sur les patients et sur le pouvoir politique. La médecine n'en était pas pour autant du charlatanisme. Elle était essentiellement une science de laboratoire.

L'usage de groupe témoin pour évaluer la qualité d'un traitement est une pratique très ancienne. Dès le XVIIIème siècle, l'anglais David Lind faisait des expérimentations cliniques contrôlées. Il a montré ainsi qu'en mangeant des agrumes en abondance, les marins faisant de longs voyages pouvaient sensiblement réduire l'apparition de scorbut – maladie liée à une carence en vitamine C provoquant un déchaussement des dents, la purulence des gencives et, éventuellement, la mort. Afin de démontrer sa thèse, Lind a divisé un équipage de 12 marins en deux groupes de 6, l'un d'entre eux ajoutant à son régime de base des agrumes, l'autre ajoutant à son régime de base des traitements de l'époque (cidre, élixir de vitriol, etc.). Seuls les marins ayant consommé des agrumes ont guéri et la preuve a été faite d'un traitement efficace contre le scorbut.

Plus largement, dès la création de l'*American Medical Association* en 1847 aux États-Unis, la profession médicale s'est donné pour objectif la réforme thérapeutique (Marks, 1997) : promotion des spécialités pharmaceutiques, utilisation correcte des innovations pharmaceutiques, élimination des thérapeutiques jugées sectaires, etc. Peu à peu, la profession va même commencer à s'attaquer à l'industrie pharmaceutique qui, sans réglementation contraignante, se permet de vendre des substances sans effets – pourtant dotées par la force de la publicité de vertus quasi-magiques. Si à cette époque la lutte de la médecine libérale est une lutte pour le monopole de la dispense des soins, c'est aussi une lutte pour la promotion des soins les plus efficaces. C'est en effet de son efficacité que la médecine doit espérer tirer sa légitimité. Or, au début du XXème siècle, les preuves sont faites : insuline (diabète), antitoxine (diphtérie), antipyrétiques (fièvre), hormones purifiées, sérums et vaccins, pénicilline (syphilis), streptomycine (tuberculose), etc. Les médecins peuvent soigner.

Cependant, la connaissance n'est pas totale et la forme particulière de création de la connaissance crée des formes particulières de la dispute. Dans ce contexte, le débat sur la qualité des soins est empreint d'arguments d'autorité peu discutés même si très discutables. La variabilité des pratiques était alors une conséquence nécessaire de cette conception de la pratique médicale. Il y avait autant de pratiques légitimes que de médecins suffisamment charismatiques pour imposer leur vue.

C'est dans le contexte d'après-guerre et à partir du traitement de la tuberculose, que va naître au Royaume-Uni une nouvelle méthodologie révolutionnant la production de la preuve médicale : l'essai clinique randomisé[119] (encadré 7.2.).

Encadré 7.2. L'émergence du gold standard de l'essai clinique randomisé. Keel (2011)

Aux États-Unis, dans l'après-guerre, les recherches sur la tuberculose font l'objet d'une méthodologie innovante avec la réalisation d'essais multicentriques. Ce sont des essais réalisés dans un nombre important de centres de soins différents afin de brasser statistiquement un grand nombre de patients et de médecins. L'idée est de nettoyer l'effet observé du traitement d'une possible contingence locale. Le critère d'essai multicentrique est ainsi le premier des deux critères permettant de définir l'essai clinique randomisé, le second étant la randomisation. La randomisation est une procédure statistique prévoyant la répartition au hasard des malades sélectionnés pour participer à l'essai entre un groupe expérimental et un groupe témoin120. Cependant, les recherches réalisées aux États-Unis sur la tuberculose n'ont pas utilisé le critère de randomisation pour des raisons éthiques. En effet, à cette époque les médicaments étaient fournis gratuitement et en quantité suffisante par l'industrie pharmaceutique. Or, comme les médecins connaissaient déjà les effets bénéfiques de la streptomycine sur la guérison de la tuberculose121, il leur apparaissait éthiquement inacceptable de ne pas donner de traitement à des patients malades (le groupe témoin).

C'est au Royaume-Uni, dans un contexte de pénurie de médicaments et du fait de médecins peu convaincus de l'effet de la streptomycine, qu'a été défini le gold standard de l'essai clinique randomisé. En situation de pénurie et en raison de l'incertitude des médecins, la méthode l'essai clinique randomisé (multicentrique et avec assignation aléatoire) a paru être la plus performante pour évaluer un traitement.

L'expérimentation clinique de 1947/1948 du Medical Reaserch Council comportait les deux critères de l'essai clinique randomisé : essais multicentriques et randomisation. L'efficacité du traitement a été largement démontrée avec une mortalité de 7 % pour les patients traités contre 27% pour les patients non traités. Cet essai est depuis considéré comme le gold standard de la preuve en médecine122.

Suite à ces premiers essais cliniques randomisés, la médecine va peu à peu systématiser l'usage des statistiques et l'essai clinique randomisé sera l'épicentre de ce que l'on appelle désormais l'*Evidence Based Medicine* (EBM, médecine fondée sur les preuves). Dès les années 1950, une poignée de médecins – qu'Harry Marks (1999) appelle alternativement « réformateurs thérapeutiques » et « réformateur méthodologiques » – vont chercher à persuader leurs collègues chercheurs de la supériorité de l'argument statistique – et de la procédure de l'essai clinique randomisé – par rapport aux méthodes anciennes.

Le développement à grande échelle de ce l'EBM s'est notamment fait à partir de l'Université McMaster au Canada sous le contrôle de David Sackett. Les nouvelles

[119] Dans une perspective historique rigoureuse, il faudrait rappeler que les premières étapes menant à la définition précise du protocole d'essai clinique randomisé remontent aux années 1920. Mais, ce n'est qu'après-guerre que sera définit le *gold standard* de l'essai clinique randomisé (Marks, 1999, Keel, 2011).

[120] La séquence d'assignation doit provenir d'une série de nombres d'échantillonnage produits aléatoirement et le secret de l'assignation doit être préservé.

[121] Soulignons une dernière fois un point important : les médecins peuvent guérir avant l'apparition de l'essai clinique randomisé.

[122] On note ainsi que deux critères régulièrement discutés dans la littérature ne sont pas nécessaire pour réaliser un essai clinique randomisé : la pratique de double insu et l'usage d'un placebo.

technologies de l'information de la communication ont permis le traitement de masses gigantesques de données ainsi que la diffusion des résultats sur la planète entière. Chemin faisant, la valeur de l'essai clinique randomisé supplanta les autres éléments de preuve dans l'élaboration des règles de l'« art médical ». L'idée est de contester et/ou de compléter les connaissances venues de la pratique individuelle par des données externes ayant fait leur preuves statistiquement (Sackett et al., 1996). Ceci permettrait de réduire la variabilité des pratiques notamment en choisissant les traitements les plus efficaces.

La victoire de la statistique comme instrument de preuve est peu à peu assurée. Gordon Guyatt, l'un des plus grands promoteurs de l'épidémiologie clinique et inventeur de l'expression EBM résume parfaitement les termes du débat : « Au point où nous en sommes, je vais être incroyablement rigoureux et systématique dans mon accumulation de preuve, vous êtes dans le nouveau paradigme [l'EBM] et vous y faites de la science » (cité par Dumesnil, 2009, p.55). En dépit des avancées incontestables de l'EBM, dont il faut se féliciter, cette méthode signe la « victoire de la théorie sur la pratique et du standard sur la singularité » (Dumesnil, 2009, p.56).

À ce stade de l'analyse il convient de constater la convergence de l'EBM et des objectifs de politique publique. Cette méthode ambitionne de réduire les dépenses de santé et d'augmenter la qualité des soins par la baisse de la variabilité des pratiques et le choix des traitements les plus efficaces. Il nous faut également revenir sur le statut de la preuve. L'émergence de l'EBM est le résultat du retournement de la hiérarchie des preuves : l'autorité du *leader* charismatique et l'expérience individuelle du praticien sont aujourd'hui mises au second plan par les données issues du traitement statistique à grande échelle. Il s'agit, si l'on reprend les mots de Gordon Guyatt de « faire de la science ». Ce registre de légitimation est d'une importance centrale pour expliquer le développement hégémonique de l'EBM ainsi que son pouvoir de conviction permettant de construire la confiance des décideurs et des patients.

1.2.2. Certifier les normes quantifiées : les Agences

Le travail de création de normes n'est pas à lui seul suffisant pour construire le produit médical nécessaire au contrôle par les indicateurs. En effet, il est impératif de certifier ces normes pour que les médecins et les patients puissent s'y fier. La définition du produit ne suffit pas, il faut en certifier la qualité et en assurer la diffusion parmi les producteurs. Le problème de la diffusion des normes est d'autant plus central que dans le cas français,

l'identité libérale rend difficile l'application de guides généraux et que les médecins exercent souvent en solitaire et non en cabinet de groupe.

Ce sont les Agences sanitaires qui vont assurer ce travail de certification et de diffusion des normes. Ces agences, qui se sont multipliées dès les années 1980, sont le témoin d'une redéfinition de l'action publique (Benamouzig et Besançon, 2005). C'est suite aux différentes crises sanitaires qui ont eu lieu à cette époque (sang contaminé, vache folle, etc.) que la mise en place de ces agences s'est imposée. Les crises étant considérées comme provenant de la trop grande proximité de la sphère politique sur les décisions de santé, l'idée est née de créer des agences « indépendantes ». L'indépendance affichée et revendiquée de ces agences est essentielle à la crédibilité des normes pour des médecins libéraux refusant l'intrusion de l'État dans l'interaction médecin-patient. Les administrations centrales sont jugées inefficaces parce que par définition ankylosée dans les procédures bureaucratiques. Sous le contrôle direct du pouvoir politique – toujours suspect de visées électorales bien peu en phase avec les problématiques de santé publique –, l'administration souffrirait également d'un déficit de compétence dans un domaine particulièrement complexe et technique.

Sous l'influence des évolutions de l'action publique au Royaume-Uni et des préceptes du *New Public Management*, des institutions indépendantes des administrations centrales vont alors être créées en France (Benamouzig, 2009). Ces institutions appelées « Agences » vont se multiplier en quelques années[123]. Théoriquement ces agences sont censées découpler les fonctions de pilotage et de mise en œuvre des politiques publiques. La contractualisation de la direction générale de la santé avec ses agences est donc un pilier du système. Les contrats d'objectifs permettent de laisser une marge d'autonomie aux directeurs d'agences dans leur gestion du budget, et du personnel. Cette autonomie vis-à-vis de l'administration centrale est supposée être le gage de décisions neutres et efficaces.

Parmi cette pléthore d'agences certaines d'entre elles vont nous intéresser plus particulièrement. Ce sont les agences qui ont pour mission l'évaluation des pratiques médicales. Il faut encore une fois rappeler que l'évaluation des pratiques médicale n'est en rien dans la tradition de la profession médicale française. L'autorégulation était censée suffire à garantir la qualité des pratiques. Cependant, à partir des années 1960, parallèlement aux

[123] Agence nationale de recherche sur le sida (1992), Agence française de lutte contre le sida (1989), Agence du médicament (1993), Réseau national de santé publique (1992), Agence française du sang (1992), Etablissement français des greffes (1994), Institut de veille sanitaire (1998), Agence française de sécurité sanitaire des produits de santé (1999), Agence française de sécurité sanitaire des aliments (1999), Agence de biomédecine (2004), etc.

premiers pas de l'EBM c'est dans le corps médical lui-même que l'on va trouver des individus promouvant l'évaluation médicale.

Mais, c'est l'évolution des rapports de force entre les acteurs du système de soin (État, assurance maladie, experts de l'évaluation (médecin ou non), syndicats de médecins, Ordre des médecins, etc.) qui a déterminé les formes institutionnelles actuelles de l'évaluation médicale (Robelet, 1999). L'institutionnalisation de l'évaluation médicale s'est ainsi construite par une « double concurrence » (Robelet, 1999, p. 71). D'une part, des médecins experts en évaluation se sont confrontés aux représentations traditionnelles de la profession et seules titulaires du pouvoir d'autorégulation. Les premiers, détenteur d'un savoir spécifique particulièrement utile pour les politiques de maîtrise des dépenses, sont alors susceptibles de nouer des alliances avec le régulateur afin de remettre en cause l'autorégulation traditionnelle. D'autre part, une seconde concurrence se joue entre les médecins experts avec d'un côté ceux tournés vers une définition de l'évaluation en termes de santé publique et d'un autre côté des experts qui s'intéressent plus directement à la maîtrise des dépenses.

Avec cette clef de lecture, il est possible d'expliquer l'évolution du rôle des agences d'évaluation des pratiques médicales. En 1989, à la faveur d'un changement de majorité gouvernementale un rapport sur l'évaluation médicale est commandé à Jean-François Armogathe, président de l'Unaformec – Union d'associations professionnelles consacrées à la formation médicale. Ce rapport insiste sur une définition de l'évaluation en termes de formation. C'est dans cette optique qu'est créée, en 1990, l'Agence nationale de développement de l'évaluation médicale (ANDEM). Les missions – élaboration de recommandations de pratiques médicales via des « conférences de consensus » et des sociétés savantes – et le statut de cette agence – une association – illustrent bien la place et le rôle de l'évaluation. Il s'agit d'une institution non contraignante qui n'ébranle pas l'autorégulation de la profession médicale.

Le durcissement de la contrainte de budget fait de l'évaluation des pratiques médicales un outil de plus en plus pertinent. L'ANDEM se voit d'abord attribuer des compétences dans l'évaluation hospitalière (loi hospitalière de 1991) puis elle participe dès 1993 à l'élaboration des Références Médicales Opposables (RMO). Ces évolutions débouchent en 1995 sur le plan Juppé qui durcit les contraintes vis-à-vis du monde médical et cela notamment dans le champ de l'évaluation médicale. L'Agence nationale d'accréditation et d'évaluation en santé (ANAES) prend donc le relais de l'ANDEM. L'agence n'est plus une association et elle impose l'autodiscipline des médecins. En plus des missions d'élaboration, de validation et de

diffusion des méthodes d'évaluation et de recommandations de pratiques de l'ANDEM, l'ANAES obtient la mission d'accréditation des établissements de santé. Sur demande des établissements de santé les experts de l'ANAES viennent les évaluer et donnent leur rapport à l'agence régionale d'hospitalisation seule compétente en matière de planification. La contrainte est mesurée mais elle progresse.

Un nouveau pas est franchi en 2004, lorsqu'est créée la Haute autorité de santé (HAS) venant remplacer l'ANAES. En plus des missions de l'ANAES, la HAS se voit attribuer de nouvelles missions : définition de guides de bonnes pratiques en matière d'affection longue durée pour les médecins et les patients ; accréditation, sur base du volontariat, de certaines disciplines médicales ; élargissement des recommandations à des critères médicaux-économiques en matière de stratégie de soins et de prescriptions (notion de service médical rendu) ; obligation d'intégrer les recommandations de la HAS dans les logiciels d'aide à la prescription ; etc.

Sur le long terme, les contraintes sur la pratique des médecins se renforcent à mesure que l'objectif de maîtrise des dépenses s'affermit. De l'absence de recommandations de bonnes pratiques, nous sommes passés à des systèmes de recommandations et d'accréditation. Les contours du produit « soin médical » se précisent et le travail d'homogénéisation est en marche. Dans cette perspective, l'indépendance affichée des agences productrices de normes[124] est censée être garante de leur qualité et, en conséquence, réduire l'incertitude sur le bien à échanger.

1.3. L'évolution de la responsabilité médicale : la *Soft Law*

L'usage particulier fait par la puissance publique des normes quantifiées de la qualité des soins pourrait très bien n'avoir aucun impact sur la pratique médicale. Ces standards ne sont après tout que de l'encre sur du papier et les médecins – libéraux ou non – demeurent totalement libres dans leurs pratiques de prescription et thérapeutique. Pourtant, ces nouvelles normes produisent des effets – tant du côté des médecins que des patients. Ces effets sont d'ailleurs d'autant plus élevés que le droit vient peu à peu donner de l'autorité à cette *soft law* (droit mou).

[124]La HAS est, selon le code de la Sécurité sociale, une autorité publique indépendante à caractère scientifique.

1.3.1. La mobilisation des normes par les acteurs

Pour les patients-consommateurs, les normes sont une source d'autorité permettant de réduire l'asymétrie d'information dans le colloque singulier et de contester le pouvoir médical dans le cadre d'une démarché qualité (Setbon, 2000). L'effet recherché est de réduire l'autonomie du médecin en dotant les consommateurs de soin du pouvoir de contrôler l'exercice de l'« art médical ». Dès lors, l'introduction des normes quantitatives est réputée permettre de changer de paradigme de production de soin en entrant dans l'ère de la qualité des soins. En ce sens, les normes sont la source d'une redistribution des pouvoirs vers le consommateur de soin. Il est désormais possible pour le patient d'opposer à son médecin l'existence de normes de bonnes pratiques. Ce mouvement de reconquête du pouvoir du patient et des associations de patients n'est pas sans échos avec le programme de construction du consommateur de soin (Batifoulier et al. 2011). Il s'agit même d'un pilier essentiel : sans information abondante il ne peut pas y avoir de menace crédible sur le médecin (judiciaire ou marchande en faisant jouer la concurrence).

En outre, contrairement à ce que l'on pourrait croire, les consommateurs de soins ne sont pas les seuls à mobiliser les normes quantifiées dans la relation médicale. Evidemment les médecins peuvent les utiliser pour orienter leur décision mais, et c'est peut-être un effet inattendu, ils ont la possibilité d'utiliser la norme afin de s'opposer au patient lui-même et au pouvoir bureaucratique (Castel et Merle, 2002). D'abord, en sélectionnant pour le médecin les informations pertinentes dans la masse de données disponibles, les normes constituent des aides à la décision ayant une « dimension rassurante ». Ce ne serait pas tant l'autonomie qui serait remise en cause que l'isolement des médecins. Ensuite, et de façon décisive les normes permettraient aux praticiens de se légitimer face aux tiers. En respectant les normes, les médecins peuvent plus facilement faire attester la qualité des soins qu'ils prodiguent ceci facilitant la procédure d'accréditation. De plus, le respect de ces normes de qualité permettrait de désarmer la puissance publique lorsqu'elle cherche à appliquer la maîtrise médicalisée des dépenses. Puisque les pratiques normées sont les meilleures possibles, il est illégitime d'en dévier. Enfin, les normes sont un outil de résistance face aux patients mécontents de leur traitement. Traditionnellement plus sensibles aux pressions des malades et de leurs familles, les médecins peuvent désormais légitimer leurs pratiques à partir des normes quantifiées. Le médecin est inattaquable judiciairement s'il a respecté les normes.

Comme on le voit, les normes sont mobilisées et mobilisables dans le cadre de la relation de soin. Cependant, il reste à savoir quel pouvoir ont ces normes ? En effet, si ces

normes n'ont pas de pouvoir contraignant sur les comportements, alors elles ne servent à rien. Il faut donc s'intéresser à ce que dit le droit sur la place de ces standards.

1.3.2. Les principes de la responsabilité du médecin à partir des données « acquises » de la science

Traditionnellement, la responsabilité du médecin est définie par rapport à la conformité aux règles de l'art. Plus formellement, c'est l'arrêt Mercier (1936) qui définit les obligations du médecin :

> *« Attendu qu'il se forme entre le médecin et son client un véritable contrat comportant pour le praticien l'engagement sinon bien évidemment de guérir le malade, ce qui n'a d'ailleurs jamais été allégué, du moins de lui donner des soins, non pas quelconques […] mais consciencieux, attentifs, et réservé faite de circonstances exceptionnelles, conformes aux données acquises de la science » (Cour de cassation, chambre civile, 20 mai 1936).*

Dès cette date, la notion de données acquises de la science a constitué la « clef universelle de la responsabilité médicale » (Loiseau, 2005). Mais, aussitôt posée, cette définition a été sujette à interprétation : faut-il parler de données acquises ou de données actuelles comme le fait la Cour d'appel de Paris en 1946 ? Si ces questions terminologiques peuvent paraîtres anecdotiques elles ne le sont pas puisque la donnée actuelle semble plus exigeante pour le médecin que la donnée acquise. Alors que la première est une donnée dépendant au jour le jour de l'évolution des connaissances, la seconde suppose un temps de diffusion des informations à l'ensemble de la communauté. La jurisprudence a confirmé au début des années 2000 la primauté des données acquises de la science que le doyen Carbonnier définit ainsi :

> *« [une] donnée médicale n'est une donnée acquise de la science qu'autant qu'elle est reçue par la partie la plus considérable de l'opinion scientifique (question de fait). Plutôt que de compter des têtes, il s'agit d'apprécier la force du courant » (cité par Mascret, 2008, p. 167).*

Selon Pierre Sargos (1998) la donnée acquise de la science provient en toute logique des livres classiques, des traités médicaux, des articles de spécialistes, des travaux de congrès médicaux. L'engagement contractuel du médecin porte donc sur le fait d'apporter des soins – et non de guérir – conformes aux données acquises de la science.

En principe, la responsabilité du médecin peut être engagée sur faute. C'est par exemple le geste maladroit ou la maladresse du chirurgien qui occasionne un dommage non lié à l'aléa de la situation. Il est alors nécessaire de distinguer l'obligation de moyen – cas général – de l'obligation de résultat – cas particulier où il n'existe pas d'aléa inhérent à la situation en

cause. L'aléa en matière médicale étant la règle, l'obligation du médecin est une obligation de moyen – le patient ne pouvant pas attendre du praticien un résultat prédéterminé. L'obligation de résultat est beaucoup plus restreinte. Elle existe par exemple dans les cas d'infections nosocomiales dans la mesure où le code de santé publique prévoit que dans de tels cas, seule la responsabilité des professionnels de santé et des établissements de santé sont en cause.

Plus récemment s'est imposée une obligation nouvelle pour le médecin : le devoir d'information du patient. L'obligation d'information ne porte pas sur l'ensemble de l'activité médicale. En effet, concernant le diagnostic et le pronostic, le code de déontologie médicale prévoit que, « dans l'intérêt du malade et pour des raisons légitimes », le praticien peut décider d'entretenir l'ignorance du patient sur un diagnostic ou un pronostic grave. Ici, l'asymétrie d'information est assumée et défendue… dans l'intérêt du malade.

Néanmoins, le médecin est soumis à l'obligation d'information en ce qui concerne l'acte médical proposé. L'objectif de cette obligation est de donner les moyens au malade d'un consentement éclairé. Cette obligation d'information est un principe général qui ne connait d'exception que dans les cas d'urgence – où il est impossible d'informer le patient – et dans les cas où le patient demande à ne pas être informé. Cette nouvelle responsabilité du médecin peut se concevoir comme un contrepoint à l'obligation de moyen. Puisque il existe un aléa qui interdit d'exiger du médecin un résultat, il semble nécessaire d'informer le patient de la nature de cet aléa.

C'est sur ce cadre général que sont venus s'insérer peu à peu les standards chiffrés – produits par des agences indépendantes ou d'autres institutions – posant alors la question de l'opposabilité de ces normes au médecin.

1.3.3. L'émergence de la soft law à partir des normes de bonne pratique

La définition classique de la responsabilité du médecin autour de la notion de donnée acquise de la science s'est trouvée chamboulée par l'évolution récente du droit de la santé. Selon le constat d'Anne Laude (2009), la dernière décennie a en effet été marquée par un « développement tentaculaire » du droit de la santé, de la multiplication des domaines d'application de celui-ci ainsi que de l'apparition et la croissance d'une nouvelle source de droit : la *soft law*. Ce « droit mou » est constitué non seulement de toutes les normes produites par les institutions créées par le droit (Haut conseil de la santé publique, Commission transparence, Comité économique des produits de santé et Haute autorité de santé) mais aussi

par les normes moins « officielles » : *guidelines*, recommandations de bonnes pratiques, référentiels, etc. Cette source de droit est à ce point importante qu'elle peut être qualifiée de « droit parallèle » (*ibid.*, p. 55). Quelle est, dans ce contexte renouvelé, la responsabilité du médecin vis-à-vis de ces normes ?

Cette question s'est posée dès 1993 avec la loi du 4 janvier créant les Références médicales opposables (RMO) visant à maîtriser l'évolution des dépenses de santé (Encadré 7.3.). Il s'agit d'une forme négative de référence interdisant telle ou telle pratique (obligation de ne pas faire) et opposable par les caisses d'assurance maladie (et non par les patients). La pratique en dehors de la norme peut entrainer des sanctions financières. D'un statut privé (médecin-patient), la relation médicale devient un contrat à caractère réglementaire. Ces normes sont une nouvelle source d'appréciation pour le juge de ce qui relève de la responsabilité médicale (Laude, 1998). Les RMO peuvent en effet servir pour définir ce que sont les règles de l'art et, le cas échant, orienter la décision du juge. Les règles de l'art, qui correspondent aux données acquises de la science, sont bien plus larges et souples que les RMO. Elles expriment « la prudence, l'attention, et la conformité aux données scientifiques du moment (Penneau et Penneau, 1998, p. 4). La différence temporelle entre le moment de la fixation des normes (par définition figé) et l'état de l'art (par définition mouvante) implique pour ces normes une « autorité nulle » (*ibid.*, p. 6). La responsabilité du médecin étant fondée sur la faute commise, le respect des normes ne suffit pas à s'exonérer de sa responsabilité.

Encadré 7.3. Quelques exemples de références médicales opposables (1993)

Le Pape et Sermet (1998)

Dans une analyse sur les Références médicales opposables, Le Pape et Sermet (1998, p. 5) recensent une série de RMO qui, si elles ne sont pas encore des normes quantifiées comme on les trouve dans la ROSP, permettent de donner une idée de la dimension prescriptive de cette littérature :

Références prenant effet en 1994	Référence prenant effet en 1995
Prescription des antibiotiques en pratique courante (pathologie ORL) : « Il n'y a pas lieu d'utiliser une association antinopénivilline-inhibiteur des bêta-lactamases » « Il n'y a pas lieu d'utiliser les fuoroquinolones systémiques » Prescription des antiulcéreux : « Il n'y a pas lieu d'associer ou de prescrire simultanément deux antiulcéreux » « Il n'y a pas lieu de prescrire des antiulcéreux dans les gastrites chroniques » Diabète non insulino-dépendant : « Il n'y a pas lieu chez les sujets de plus de 70 ans, d'utiliser des sulfamides hypoglycémiants à ½ vie longue »	Traitement de l'adénome prostatique : « Il n'y a pas lieu d'utiliser l'association de 2 médicaments ou plus pour traiter les troubles mictionnels de l'hypertrophie prostatique bénigne » Prescription des hypnotiques et anxiolytiques : « Il n'y a pas lieu d'associer deux benzodiazépines pour un traitement anxiolytique » « Il n'y a pas lieu d'associer deux hypnotiques » Prescription des vaso-actifs : « Il n'y a pas lieu d'associer deux vaso-actifs ou plus, dans l'indication de l'artériopathie des membres inférieurs »

Cependant, le temps faisant son œuvre, la « normalisation galopante » (Mascret, 2008, p. 168) commence à produire des effets : le Conseil d'État s'est récemment appuyé sur les recommandations publiées par l'ANAES pour retenir la responsabilité d'un médecin. Le respect des normes n'est pas suffisant pour s'exonérer de sa responsabilité mais leur non-respect peut être retenu à charge. La jurisprudence tendrait ainsi à reconnaitre un caractère contraignant aux normes. Dans ce cadre, les RMO s'intègrent aussi aux données acquises de la science.

Olivier Smallwood (2006) synthétise les deux types de contraintes qui peuvent peser sur un médecin face aux normes de pratiques. La « contrainte indirecte », dont le principe a déjà été esquissé plus haut, porte sur le degré avec lequel il y a coïncidence entre les normes (figée) et les règles de l'art (mouvante) :

> *« La coïncidence évoquée n'est en réalité que ponctuelle au moment de l'élaboration des recommandations ou références médicales, lorsque la règle de l'art est figée pour une durée inconnue bien qu'obligatoirement éphémère. Il convient dès lors d'envisager la relation entre le contenant et le contenu, c'est-à-dire entre les règles de l'art et les recommandations, références, littérature médicale… sous l'angle d'une présomption simple ou « présomption d'identité de fait », l'affirmation d'une telle présomption n'étant pas sans incidence sur le terrain de la responsabilité des professionnels de santé » (Smallwood, 2006, p. 123).*

Dans ce cadre, il est possible d'établir une hiérarchie au sein même des recommandations médicales selon leur degré de présomption de coïncidence. Les recommandations non impératives non homologuées seraient en se sens les moins contraignantes alors que les recommandations impératives et homologuées décrirait parfaitement les règles de l'art. Ainsi, le respect des normes ne peut exonérer le médecin de sa responsabilité dans la mesure où il reste nécessaire de prouver leur conformité aux règles de l'art. En cela la portée contraignante des normes est indirecte : elle dépend de leur adéquation aux règles de l'art. Mais, d'après l'auteur, en plus de cette contrainte indirecte, la jurisprudence tend à reconnaitre une force contraignante directe aux normes :

> *« La jurisprudence administrative semble s'accorder sur le fait que le respect des recommandations de bonnes pratiques et des références médicales opposables constitue une obligation déontologique à la charge de tout praticien » (ibid., p. 124).*

Les décisions du conseil d'État du 12 janvier 2005 et de la Cours administrative d'appel de Nancy du 10 janvier 2005 se sont fondées sur la *soft law* (nomenclature générale des actes professionnels et RMO) pour reconnaitre la responsabilité déontologique du médecin. Il ne s'agit pour l'instant que de responsabilité déontologique mais il est légalement possible de se fonder sur ce type de faute pour attribuer une responsabilité civile ou administrative. On peut alors se demander, si comme le suggère Grossieux (2003), les RMO ferons « bientôt office de règles de l'art ». La force contraignante directe des normes semble gagner de l'importance. La normalisation des règles de l'art est une première étape qui peut déboucher logiquement sur le changement des sources de responsabilité déontologique, civil et administrative du médecin.

2. Critique de la normalisation du travail médical

Le second grand tournant de la médecine libérale est lancé. L'activité médicale est régulée de façon croissante à partir de standards chiffrés – constituant potentiellement autant d'indicateurs de la performance publique. Mais, cette façon de produire la qualité des soins à partir d'un jeu sur les quantités ne va pas sans difficultés. Elle développe un gouvernement de la qualité par le chiffre qui n'est pas neutre et oriente vers une définition particulière de la performance (2.1). En systématisant le recours aux résultats de l'*Evidence Based Medicine* (EBM), la politique publique opère également une double réduction épistémologique problématique : la maladie n'est pas un objet où peut être accompli l'idéal de scientificité et

de neutralité de l'EBM (2.2.) et le service de soins n'est pas assimilable à un produit – ce qui interdit d'appliquer aux soins le paradigme productiviste (2.3).

2.1. Le gouvernement par le chiffre

La normalisation des soins n'est que la division médicale d'un mouvement plus général visant la « qualité totale ». Cet objectif modifie les comportements par l'injonction à la « performance totale » qui, en naturalisant la donnée chiffrée, évince le débat démocratique sur ce qu'il convient de faire (2.1.1.). Dans une perspective critique cherchant à dénaturaliser les termes du débat, il faut alors s'interroger sur l'épistémologie des objets statistiques qui est à l'œuvre (2.1.2.).

2.1.1. Qualité totale et performance totale

Le concept de performance totale forgé par Florence Jany-Catrice permet d'éclairer les ressorts du gouvernement par le chiffre. La performance totale est conçue sur le modèle de la « qualité totale » : c'est son négatif photographique. Pour que les acteurs puissent assurer la qualité totale – entendue comme une mode de gestion de la qualité qui ne connait pas de limites –, ils doivent – au double sens d'être redevable (devoir moral) et d'être contraint (devoir contractuel) – un engagement illimité au travail. L'appel à l'autonomie des acteurs et l'usage intensif d'outils statistiques forment le socle de ce gouvernement par la qualité. Cette double obligation est cependant paradoxale dans la mesure où les indicateurs quantitatifs définissant le travail prescrit sont de plus en plus éloignés du travail réel et du sens que les personnes donnent à leurs actions. La performance totale, s'entend alors comme un ensemble vaste de dispositifs régulant le comportement des acteurs auquel personne ne croit mais auquel tout le monde se plie :

> *« […] il existe une tendance structurelle à ce que les organismes, les États comme les individus, se conforment à la performance totale définie comme régime, qui, tout en visant à une évaluation du produit et ses finalités, finit par dicter, en les y soumettant, les actes des individus et leurs finalités. La performance totale est ainsi le dispositif qui coordonne les agents à partir de l'exhibition d'une grandeur particulière, celle énoncée par le chiffre. Celui-ci est à la fois parfaitement décontextualisé et totalement instituée par une matérialité qui s'avère durable (du fait des propriétés particulières du chiffre), et auquel chacun est sommé de se conformer. Cette norme chiffrée ainsi créée sert alors de lien organique et symbolique entre les individus » (Jany-Catrice, 2012, p. 13).*

Le recours à un grand nombre de statistiques a pour fonction principale de légitimer un mode d'organisation de la santé qui n'est plus questionné. Orienter le comportement des acteurs vers une finalité conçue comme bonne en soi, car méthodologiquement neutre et impartiale, permet de ne parler que de l'évolution des indicateurs : sont-ils bien mesurés (critique du codage) ? ; y-en-a-t-il suffisamment (critique du multitâche) ? ; peut-on écarter les cas difficiles (critique de l'homogénéité de la patientèle) ? ; le niveau des objectifs est-il suffisamment exigeant (critique du taux cible) ? ; etc.[125]

Si ce niveau de la critique est important, il occulte la discussion sur la pertinence de l'indicateur lui-même. Alors que dans le premier cas la question porte sur la façon (technique) dont on mesure, dans le second c'est la pertinence de la chosée mesurée elle-même qui est remise en cause. Plus largement encore, l'utilité du pilotage des comportements par les indicateurs est naturalisée alors qu'elle relève d'une pratique sociale historiquement située. Les outils de gestion ne sont pas neutres et révèlent un conception normative de les organisations et de leurs finalités (Palpacuer et al., 2010). Les débats sur les indicateurs de richesse montrent également que ceux-ci ne sont pas que des outils techniques mais, malgré la naturalisation dont ils font l'objet, des institutions relevant du politique (Gadrey et Jany-Catrice, 2009, Jany-Catrice et Méda, 2013).

Il est possible de repérer l'éviction de la délibération démocratique sur les indicateurs de performance en distinguant plusieurs régimes de performance, selon le nombre d'indicateur de performance et le mode de décision retenus (Jany-Catrice, 2012). Dans ce cadre, soit la performance est unidimensionnelle et tous les comportements sont orientés vers un seul indicateur de performance, soit la performance est multidimensionnelle et c'est une batterie d'indicateurs qui conduit l'action. Du point de vue du mode de sélection du ou des indicateurs, soit le processus de décision est planifié par un régulateur central, soit il est négocié par les parties prenantes. Au total, on peut repérer quatre régimes de performance (Tableau 7.3.).

[125] Cf. Chapitre 5 pour une critique des indicateurs de la ROSP.

Tableau 7.3. Les régimes de performance (à partir de Jany-Catrice, 2012)

	L'évaluation est…	
Le processus est…	Unidimensionnel	Multidimensionnel
Planifié	Régime autocratique (A)	Régime managérial (B)
Négocié	Régime fordiste (C)	Régime délibératif (D)

Le régime de performance autocratique (A) est assimilable au capitalisme financier contemporain avec le taux de profit comme seul indicateur de performance des entreprises. Le régime managérial (B) se distingue du premier dans la mesure où, s'il y a toujours une planification centralisée, l'évaluation de la performance se fait sur une multitude de critères. Ce cas de figure est particulièrement pertinent pour analyser l'action publique qui, par nature, ne peut pas être évaluée par une norme de rentabilité. C'est l'esprit qui anime la nouvelle régulation de la médecine libérale lorsque la puissance publique essaie de gouverner à distance par l'usage d'indicateurs de performance plus ou moins imposés. Le régime fordiste (C) se caractérise par la négociation collective de l'objectif unique à atteindre. Le compromis fordiste est un exemple paradigmatique. Enfin, le régime délibératif (D) ajoute à la négociation collective de l'indicateur la pluralité d'objectifs à atteindre. Ce régime de performance est au fondement des démarches associatives ou coopératives.

Cette typologie permet d'éclairer nettement le déficit démocratique du système de santé. Bien qu'il existe une pluralité d'indicateurs justifiée par la complexité du travail médical, le fait que ces indicateurs soient décidés par l'assurance maladie (Contrat d'Amélioration des Pratiques Individuelles) ou dans le cadre des rapports conventionnels (Rémunération sur Objectif de Santé Publique) évince de fait les usagers du système de santé. Si le déficit démocratique est suffisant en soi pour fonder la critique, il se trouve que les indicateurs choisit posent également des problèmes, en dépit de la tentative de naturalisation dont ils sont l'objet.

2.1.2. Réalisme et constructivisme des objets statistiques

Le dispositif de paiement à la performance médicale s'inscrit dans ce régime de la performance totale. Il repose sur l'idée qu'il est possible de mesurer objectivement le travail du médecin. Équipé de cette information, la mise en place d'incitations adéquates permettrait d'améliorer la qualité des soins. Dans ce cadre théorique, la question de la mesure n'est censée être qu'une question purement technique à laquelle les méthodes statistiques et des

investissements technologiques peuvent aisément répondre. Tout se passe comme s'il y avait un consensus sur la définition de l'objet mesuré.

Selon Alain Desrosières (2010), cette appréhension du travail statistique peut être qualifiée de réaliste. La conception réaliste postule que les objets étudiés sont des choses qui existent indépendamment de la mesure. Ils préexistent à la réalité sociale. Dès lors, il est possible de mesurer objectivement le travail du médecin parce que l'on sait parfaitement définir de quoi l'on parle. La maladie (M) se définit par telle liste exhaustive de symptômes mesurables et elle se soigne par le protocole (P_M) dont l'application permet de mesurer le travail du médecin (L_M). Le médecin performant est celui qui diagnostique la bonne maladie et qui applique le traitement adapté. Les indicateurs efficaces (concernant la maladie, le protocole ou la performance du médecin) existent et sont simplement en attente de leur propre découverte. C'est de la « recherche-scientifique-objective » qu'il faut attendre la révélation.

À l'inverse, la position constructiviste postule que les objets étudiés sont le fruit d'un travail de définition qui s'inscrit dans un contexte sociohistorique. Ils sont des conventions statistiques dont on ne peut pas taire le processus de construction. La question de la mesure du travail médical suppose alors que puisse être définis les contours de ce qui est à mesurer. Or, nous ne disposons pas d'une définition de l'activité médicale. À défaut de définition, ce sont les normes quantifiées de l'EBM qui sont utilisées pour ajuster les mécanismes de régulation du travail médical (notamment pour le P4P). En conséquence, c'est la méthode de l'EBM qu'il convient de questionner. Si elle peut apparaître comme une méthode neutre et purement scientifique, c'est-à-dire une méthode débarrassée de toute l'incertitude relative aux questions politiques, nous allons désormais étudier sa part d'ombre. À *minima*, deux problématiques peuvent être soulevées : peut-on mesurer la maladie ? Jusqu'à quel point la méthode de l'EBM est neutre à son contexte d'élaboration ?

2.2. Une épistémologie réductrice de la maladie

2.2.1. Deux conceptions de la maladie

La quantification de la qualité des soins repose sur l'hypothèse selon laquelle l'EBM serait une méthode de connaissance médicale permettant d'épuiser l'ensemble de la réalité. Elle permettrait de doter les médecins (puis les économistes) d'une mesure objective de la maladie. À partir de ses résultats, il serait possible de produire des protocoles universels à l'usage des praticiens. Mais, il existe un débat épistémologique important au sein même de la

communauté médicale concernant la portée de cette méthode. La question posée est de savoir ce qu'est la maladie et s'il est possible d'en faire une métrique universelle.

La controverse épistémologique évoquée ici oppose deux conceptions de la maladie que nous pouvons présenter brièvement. La première, la conception positiviste défendue initialement par Auguste Comte, définit la maladie comme une variation quantitative de l'état normal. En ce sens, la maladie est une situation de déséquilibre par rapport à la situation normale d'équilibre. Dès lors, le savoir scientifique est censé définir ce qu'est le normal et en déduire secondairement la définition de la maladie. Par exemple, c'est après avoir défini les bornes « normales » de la température du corps humain qu'il est possible d'en déduire les situations de maladies. La maladie serait « axiologiquement neutre » (Dumesnil, 2009) puisque totalement indépendante du malade. L'homogénéité entre la maladie et le normal implique nécessairement une sorte de dépossession de la maladie. Cette épistémologie est celle à l'origine de l'*Evidence Based Medicine*. Dans cette perspective, il est possible de créer des normes quantifiées à partir de l'analyse statistique (les essais cliniques randomisés) parce que le pathologique est un phénomène objectivement et scientifiquement mesurable.

C'est face à cette conception que Canguilhem écrit *Le normal et le pathologique* en 1966 (Canguilhem, 1966). Il s'oppose frontalement à la définition positiviste de la maladie faisant de celle-ci une simple variation quantitative du normal. Pour lui, il y a une hétérogénéité entre l'état normal et l'état pathologique. La maladie doit être considérée comme une « autre allure de la vie » ou encore comme une rupture avec l'état de bien-être. Ici, la rupture n'est pas le fruit d'une variation quantitative et mesurable entre deux états puisque pour cela il faudrait une homogénéité entre normal et pathologique. La maladie est l'expression de la subjectivité du patient ce qui en fait un jugement de valeur et non plus un fait indiscutable comme dans la conception positiviste. Ce n'est plus le scientifique qui est à l'origine du savoir mais le patient.

Si cette discussion peut sembler distante de notre propos initial, nous pensons que le choix entre ces deux épistémologies n'est pas neutre vis-à-vis de l'organisation de la médecine. En effet, comme le montre Julien Dumesnil (2009), il existe une forte convergence entre l'épistémologie positiviste de la maladie et les théories de sciences économiques visant à quantifier l'« art médical ». Si le pathologique n'est qu'un écart quantitatif par rapport à une norme axiologiquement neutre, alors la maladie est un concept mesurable indépendant du patient et du médecin. Il est possible de détacher la maladie du malade (et du soignant). La définition de l'objet maladie étant en ce cas parfaitement claire et objective, il devient

possible d'appliquer les méthodes de gestion issues d'autres sphères de la production économique : standardisation, normalisation, production à grande échelle, etc. Il n'y a plus de différence entre la production d'une voiture et la production de soin.

À l'inverse, suivre Canguilhem lorsqu'il énonce que la maladie est « une autre allure de la vie » mettant en jeu la singularité et les valeurs du patient implique l'impossibilité d'établir une métrique universelle du pathologique. La différence qualitative entre le normal et le pathologique impose une individualisation du soin bien peu compatible avec les méthodes éprouvées dans l'industrie. Cette épistémologie fonde plutôt une régulation de la médecine par ce qu'il est de coutume d'appeler l'« art médical ». Le médecin particulier fonde son action non seulement sur ses connaissances, mais aussi sur la singularité du malade.

2.2.2. Quelle neutralité méthodologique ?

La légitimité de l'EBM provient de sa prétention à fournir des preuves exemptes de contestations. La méthodologie de l'essai clinique randomisé empreinte de rigueur et d'impartialité est censée être un « juge de paix » permettant de clore les débats (au moins momentanément). Cependant, cette neutralité supposée peut être sérieusement questionnée dans la mesure où les résultats des essais cliniques randomisés ne sont pas indépendants du contexte sociohistorique qui les ont vus naître. Dans cette optique, on peut reprendre l'argument d'Agnès Labrousse (2010) selon lequel il faut dénaturaliser l'EBM en envisageant le processus de construction sociale des essais cliniques randomisés sous au moins trois angles : leur durée, leurs objectifs et leurs contraintes pratiques.

Du point de vue de la durée des expériences, les résultats de l'EBM doivent porter à caution puisque près d'un quart des médicaments mis sur le marché engendrent des effets indésirables qui ne sont pas détectés pas les essais cliniques. De plus, l'industrie pharmaceutique fait pression sur les régulateurs pour réduire la durée légale des essais cliniques ce qui conduit à se reposer la question de l'impartialité de cette méthode. Du point de vue des objectifs, l'auteure fait remarquer que pour accélérer la mise sur le marché des médicaments l'évaluation de nombreux essais portent non sur les résultats finaux (réduction de la mortalité ou morbidité) mais sur des résultats intermédiaires. Enfin, du point de vue des contraintes pratiques de leur mise en place, les essais cliniques randomisés ne sont pas plus à l'abri que d'autres méthodes de « petits arrangements caractéristiques de toute méthodologie en action ».

L'ouvrage d'Harry Marks, *La médecine des preuves. Histoire et anthropologie des essais cliniques [1900-1990]*, permet de dénaturaliser l'*Evidence based medicine*. Tout en se félicitant des réussites nombreuses de la méthodologie de l'essai clinique randomisé et du renversement de la hiérarchie des preuves, l'auteur averti le lecteur contre une vision idéalisée de l'usage des statistiques en médecine :

> *« […], un écart demeure entre le monde des dictats méthodologiques et les réalités sociales de la recherche clinique, écart qu'il est possible d'explorer en s'intéressant aux remarques et aux actions de ceux qui doivent conduire, interpréter et appliquer les études cliniques. Je ne nie pas que de nombreux essais contrôlés aient été organisés plus facilement, ou reçus avec plus d'enthousiasme, que ceux que j'ai choisis ici. Néanmoins, même l'essai contrôlé le plus simple est le produit d'un ordre social négocié, foisonnant de décisions – contesté ou non – et de présupposés restés sans examens. D'ordinaire, ces discussions sont le pain quotidien des chercheurs cliniques, et ne méritent pas de commentaires. Seules les études les plus célébrées, les plus élaborées, laissent une trace écrite permettant au profane d'avoir un aperçu du monde social de la recherche clinique » (Marks, 1999, p. 194-195).*

Pour illustrer cet « ordre social négocié » – cette construction sociale –, il est possible, comme le fait l'auteur, de braquer les projecteurs sur certaines études cliniques pour montrer comment d'autres critères que les critères scientifiques entrent en jeux pour la production des standards. Par exemple, pour Marks, (1999, p. 235-278), il est intéressant de voir comment dans les années 1960-1970 une étude sur les maladies cardiovasculaires – *Diet-Heart*, lancée par le *National Heart and Lung Institute* – n'a pas été réalisée du fait d'un environnement social peu favorable (Encadré 7.4.). La rareté des financements et la préférence pour des recherches sur des médicaments plutôt que sur les habitudes alimentaires ont ainsi bloqué la possibilité de réaliser un essai clinique randomisé. Même lorsque l'on accepte sans restrictions le principe de l'essai clinique randomisé, il apparaît que les conditions sociohistoriques expliquent pourquoi certaines études se font et d'autres ne se font pas. Or, ces conditions prêtent souvent à la critique – notamment lorsque l'on s'aperçoit que la recherche est presque intégralement financée (et contrôlée) par l'industrie pharmaceutique.

Encadré 7.4. La dimension politique des essais cliniques : le cas de l'étude Diet Heart. Marks (1999)

En 1950, les maladies cardiovasculaires et maladies du système circulatoire causaient 40 % des morts enregistrées aux États-Unis. Pour de nombreux praticiens c'était une bonne nouvelle : les patients meurent de problèmes cardiaques parce qu'ils ne meurent plus de maladies infectieuses. Mais, pour une minorité d'entre eux, dont le chef statisticien de la Life Insurance Company, Louis Dublin, ces maladies proviendraient d'un régime alimentaire trop riche – lui-même résultant de la disparition de la paysannerie active au profit de citadins inactifs. Afin de tester cette hypothèse, un groupe de chercheurs issu de 6 institutions médicales décide de mettre en place une étude dans laquelle ils escomptent réduire le taux de mortalité due aux maladies cardiovasculaires en proposant aux « hommes d'âge moyen » un régime alimentaire adapté à leur activité physique.

Dès le départ, cette étude a posé des problèmes de conception dépassant les problématiques habituelles comme le nombre de participants, le temps du programme, etc. Premièrement, cette étude porte en elle une contradiction entre la tradition de recherche académique – cherchant une base rigoureusement scientifique aux pratiques thérapeutiques –, et la tradition pragmatique de santé publique – qui cherche avant tout à modifier la politique publique et le comportement social général. Comment alors concilier exigence méthodologique de l'essai clinique tout en menant un essai de santé publique de masse ? Plus l'échantillon est élevé plus les résultats sont représentatifs mais plus il est difficile de maîtriser de bout en bout la méthodologie utilisée. Deuxièmement, cet essai nécessite la coopération active des entreprises alimentaires devant produire et fournir des mets spéciaux totalement homogènes. Troisièmement, plus qu'ailleurs, cette étude exige une forte coopération des patients qui ne sont pas, à vrai dire, malades et qui doivent accepter d'être enrôlés dans l'expérimentation, d'y rester jusqu'au bout – mais aussi de respecter le traitement prescrit (le régime alimentaire).

Etant données ces limitations inhérentes à la dimension sociale du sujet de l'expérience, si les résultats ne montrent pas de différences entre le groupe suivant le régime alimentaire et les autres, cela prouvera-t-il qu'un régime alimentaire pauvre n'a aucun rôle dans la prévention des maladies cardiovasculaire ? Ou cela prouvera-t-il que les médecins n'ont pas la capacité et/ou la volonté de convaincre leurs patients ?

Face à ces difficultés, les médecins ont décidé de procéder en deux temps. Dans un premier temps, ils ont procédé à une étude à faible échelle testant la capacité de l'industrie alimentaire à fournir les plateaux repas exigés et celle des patients à respecter leur régime alimentaire. Portant sur seulement 6 000 individus, les résultats sont très bons : chacun suit son protocole et le groupe test voit son taux de cholestérol baisser de 11 % par rapport au groupe témoin. Dans un second temps, il s'agit de généraliser l'expérimentation pour étudier l'impact du régime alimentaire sur les maladies cardiovasculaires. Cependant, cette étude n'a jamais vu le jour pour trois catégories de réponses principales.

D'une part, l'abandon du projet est lié au manque de financements de l'institution publique censée financer l'étude : la *National health institute* (NHI). Celle-ci était effrayée par l'ampleur de l'étude : les maladies cardiovasculaires étant relativement rares, il fallait un échantillon conséquent pour mesurer des effets significatifs – les estimations tournaient autour de 100 000 personnes pendant près de 10 ans. À cette époque, le NHI subissait par ailleurs une réduction de ses financements de la part du gouvernement fédéral et l'obligation de le distribuer entre un plus grand nombre d'équipes de recherche.

D'autre part, outre la dimension financière – première dans tout essai clinique randomisé –, l'abandon du projet s'explique également par les conflits internes au champ médical dans lequel les opposants et septiques vis-à-vis de l'utilité d'un régime alimentaire pauvre étaient nombreux et coriaces. Avant même de passée l'épreuve de l'essai clinique randomisé, une hypothèse doit passer l'épreuve du financement et de l'approbation – au moins relative – par les pairs.

Enfin, il est intéressant de noter qu'au moment même où les fonds était refusés pour l'étude *Diet-Heart*, d'autres fonds exceptionnels étaient débloqués par le congrès (650 millions de dollars) et attribués par le NHI à une étude sur le même sujet mais impliquant l'usage de médicaments – *Coronary Drug Project*.

2.2.3 Evidence based medicine et industrie pharmaceutique

Désormais conscient de la dimension nécessairement sociohistorique de l'EBM, il se pose la question de l'indépendance de la recherche à l'égard de ses sources de financement.

Des études commandées par l'industrie pharmaceutique risquent d'introduire des biais notamment en ne publiant pas systématiquement des résultats négatifs pour la firme. Cet argument est développé par Othmar Keel (2011). Selon lui, s'il y a un nombre croissant d'études peu fiables se réclamant de l'EBM, c'est à cause de leur mode de financement. Celui-ci étant presque entièrement privé et sous le contrôle des firmes pharmaceutiques, l'exigence de rentabilité prime sur la rigueur scientifique. Deux effets de la mainmise du profit sur la recherche médicale peuvent être détaillés. D'une part, l'industrie pharmaceutique multiplie les essais de nouvelles molécules contre placebo plutôt que contre un médicament déjà éprouvé. Cela conduit à multiplier le nombre de médicament de type « moi aussi », c'est-à-dire ayant une meilleure efficacité que rien (le placebo), que l'on est incapable de hiérarchiser. Les firmes peuvent alors vendre ces « moi aussi » plus chers (parce que neufs et brevetés) sans que l'efficacité des traitements soit meilleure que celle des médicaments existants : il n'ont été testés que comparativement à des placebos ! D'autre part, les recherches portent majoritairement sur des médicaments où existent un potentiel de vente important, soit plutôt en direction des populations aisées plutôt que des plus modestes, plutôt pour les pays du nord que ceux du Sud et pas pour les pour les maladies rares ou orphelines.

La neutralité et l'objectivité des résultats de l'EBM ne sont plus totalement assurées. La mise en ordre et la hiérarchisation des connaissances n'est pas aussi facile à réaliser qu'à théoriser. Logiquement, c'est le rôle des agences sanitaires indépendantes de mettre de l'ordre dans la mise en ordre. Indépendante du pouvoir politique et du pouvoir économique, la Haute Autorité de Santé devrait être la clé de voute de l'architecture du système de santé. Elle devrait être le « juge de paix des juges de paix ». Pourtant une récente décision de justice vient écorner cette aspiration. En décembre 2009, l'Association pour une formation médicale indépendante (Formindep) a déposé une requête auprès du Conseil d'État afin d'annuler une recommandation professionnelle concernant la prise en charge du diabète de type 2. Le 27 avril 2011 le Conseil d'État a annulé cette recommandation car les déclarations de conflit d'intérêt n'étaient pas respectées (seules 6 déclarations étaient disponibles sur 27) (Duguet, 2012). Plus généralement, on observe que ce type d'agences, étant faiblement doté en moyens financiers et humains pour produire les normes, a tendance à solliciter des personnes extérieures potentiellement en conflit d'intérêt (Benamouzig et Besançon, 2005).

Au total, l'accumulation de critiques peut conduire à affirmer : « D'une *evidence based medicine*, on passe alors à une *evidence biased medicine* » (Labrousse, 2010).

2.2.4. L'incertitude sur la qualité persiste et le P4P pose question

L'EBM censée fournir des normes quantifiées et objectives peine à achever son projet. Loin de réduire l'incertitude sur la qualité comme elle était censée le faire, elle ne fait que la dissimuler derrière une caution scientifique perméable à la critique. Le problème de l'incertitude n'étant pas résolu, la question de la qualité des soins demeure pertinente – même dans le cadre de l'EBM.

Un exemple éclatant de la force de l'incertitude sur la qualité des soins est relaté par Castel et Dalagarrondo (2005). La cancérologie pédiatrique française est traversée par une opposition entre deux pratiques différentes : l'une approuvant les méthodes issues de l'EBM et une autre, très minoritaire et personnalisée par le Dr Nicole Delépine, valorisant l'exercice individualisé. Cette opposition révèle des enjeux de pouvoir au sein du segment professionnel concerné avec un désaccord sur le régime de la preuve. Pour les services de cancérologie pédiatrique cherchant à normaliser les pratiques, les preuves issues de méthodes statistiques sont supérieures aux preuves issues de l'expérience individuelle du médecin. L'analyse des auteurs s'attardent longuement sur la médiatisation de l'affrontement permettant « d'apprécier la mise en public de ce débat » (Castel et Dalagarrondo, 2005, p. 18). Selon eux, la médiatisation s'est appuyée sur la critique de la standardisation et de l'expérimentation médicale pour mobiliser l'opinion publique en faveur du docteur Delépine. En jouant sur la peur et l'incertitude existant en cancérologie pédiatrique, les méthodes issues de l'EBM ont été discréditées.

Il ne s'agit pas ici de trancher un débat médical. Le point important est que dans le secteur de la santé, mettant en jeu la souffrance des patients et parfois leur vie, toute incertitude prend des dimensions extraordinaire. Quel que soit la vérité médicale, cet exemple met en évidence l'enjeu central de normalisation pour la puissance publique. L'incertitude sur la qualité a conduit à polariser les croyances sur telle méthode plutôt que telle autre par un effet mimétique. Dans l'optique de la quantification de la qualité des soins, il ne doit pas rester de place pour l'incertitude sous peine de remettre en cause l'intégralité du système de production des soins.

Or, la question de l'incertitude sur la qualité n'est pas résolue par le développement de l'EBM. L'une des raisons qui n'a pas été encore évoquée tient au fait que, même avec cette méthode, l'étendue de l'ignorance médicale reste vaste. Sur de nombreux sujets, les médecins

n'ont pas de réponses – même provenant d'essais cliniques randomisés. L'incertitude sur la qualité reste un problème central pour l'organisation du système de santé.

Dans cette perspective, nous pouvons légitimement questionner l'usage qui est fait des normes quantifiées (Salais, 2010). La mesure en elle-même n'est pas un problème : elle invite au débat et, le cas échéant, à la production d'autres mesures. Mais, connaissant les limites de des normes quantifiées dans le domaine médical, est-il raisonnable de les utiliser à travers le paiement à la performance pour orienter le comportement des médecins ? Deux risques sont encourus si la norme quantifiée n'est pas adaptée. D'une part, la santé des patients est en jeu si les médecins décident d'accomplir leurs objectifs et sont performants. La performance ainsi définie peut nuire à la santé. D'autre part, si les médecins considèrent trop importante la santé des patients et décident de ne pas accomplir les objectifs, c'est leur performance qui est en jeu. Voilà deux hypothèses permettant d'expliquer que, d'une part, la performance n'est pas forcément gage de qualité des soins et que, d'autre part, les médecins soient peu performants.

2.3. Une épistémologie réductrice des services de soins

La critique de la quantification des soins peut également s'appuyer sur l'économie des services qui met en évidence l'impossibilité d'appliquer aux services le paradigme du produit (2.3.1.). Il n'est ni possible d'établir une nomenclature des services de soin (2.3.2.) ni d'en occulter la dimension collective (2.3.3.).

2.3.1. Le soin n'est pas un produit

Les questionnements épistémologiques sur ce qu'est la maladie invitent à se protéger de l'idée selon laquelle elle serait un état indépendant du malade. Si, comme le considère Canguilhem on ne peut pas détacher la maladie du malade, alors il est impossible d'organiser le soin à partir d'une application systématique des méthodes de la production industrielle. Ce cadre d'analyse nous semble d'autant plus pertinent que l'on s'interroge ave Jean Gadrey (1991) sur l'originalité des activités de services comparativement aux activités de production :

> *« […] penser les services en utilisant le paradigme – au sens linguistique – associé aux termes de produit et de productivité est aujourd'hui dans de nombreux cas un 'obstacle épistémologique' à leur analyse économique et à leur gestion efficace. […] Cette thèse s'oppose donc clairement aux idées et aux pratiques de la transposition, selon lesquelles les services seraient seulement en attente de leur propre industrialisation, par un double mécanisme d'introduction de technologie (notamment informatique) et de normalisation voir d''objectivation' du service, visant soit à lui donner une forme de produit tangible, susceptible*

d'être acheté et consommé comme un bien industriel, soit à produire le service en série selon des protocoles standardisés » (Gadrey, 1991 : 1).

Voilà résumé le mouvement d'industrialisation de la « production » de soin en marche depuis plusieurs années[126]. Alors que ce modèle était balbutiant dans le système de soin à la date où écrivait l'auteur, l'introduction de la tarification à l'activité à l'hôpital en 2004 et du paiement à la performance en médecine de ville dès 2009 confirme la stratégie d'industrialisation du soin. Opposé à une vision productiviste des services, Jean Gadrey les définit comme : « une opération de transformation d'un état ou d'un traitement portant sur un support C détenu, utilisé ou contrôlé par un agent économique A, effectué par un autre agent B (le prestataire) pour le compte de A, mais n'aboutissant pas à la production par B d'une nouvelle marchandise susceptible de circuler indépendamment de C (on serait alors dans le cas de la production industrielle ou artisanale) » (ibid. : 8). Parmi les différents types de supports C possibles, dans le cas des soins médicaux C est « l'individu A lui-même, dans ses dimensions physiques et corporelles ». De cette définition on peut retenir que le service n'est pas un objet et qu'il ne peut être considéré indépendamment de son support, ici le malade. À *priori*, triviales ces observations portent en elle une critique radicale de l'application du modèle industriel aux services.

2.3.2 .L'impossible nomenclature des maladies et des actes

Parce que le service (de soin) n'est pas un objet indépendant du patient, il est impossible d'en établir une nomenclature. Sinon, il devrait y avoir autant de maladies listées que de malades. Or, l'organisation industrielle, pour qui la définition du produit est acquise, repose sur l'existence de protocoles permettant la standardisation de la production. La quantification permet alors aux gestionnaires de s'appuyer sur des normes reproductibles pour optimiser l'allocation des ressources. Mais, pour ce faire, il faut être capable de définir précisément le produit : dans le cas de la santé, il faut définir toutes les maladies ainsi que tous les protocoles et actes de soins associés. C'est l'hypothèse de nomenclature qui stipule que chaque bien soumis à un échange doit avoir une qualité homogène, parfaitement définie et connue de tous les agents. C'est une condition nécessaire pour que les échanges aient lieu sur le marché. Il faut non seulement définir l'ensemble des services qui peuvent être fournis, mais aussi que la qualité de ces services soit homogène afin de rendre possible l'échange marchand. Tout

[126] Le même mouvement a été repéré et discuté dans le secteur des services à la personne (Devetter et Jany-Catrice, 2010, Jany-Catrice, 2010, Gallois, 2012, Vatan, 2014). Nous nous inscrivons dans une démarche similaire, appliquée au cas des soins médicaux.

devrait se passer comme si les maladies existaient indépendamment des patients et qu'elles n'étaient simplement que dans l'attente de leur propre découverte par la recherche scientifique neutre et impartiale.

Dans le cas des soins médicaux, il est traditionnellement considéré que la production porte sur les maladies traitées. C'est ainsi qu'ont été élaborés les Groupes Homogènes de Malades (GHM) à l'hôpital. Il s'agit de regrouper des maladies, selon des critères médicaux et économiques, dans une même catégorie afin de leur associer un tarif (les groupes homogènes de séjour). Cette opération a pour but, dans le cadre de la mise en place de la tarification à l'activité, d'établir une correspondance entre les « produits » et les recettes. À partir de ce mécanisme s'organise la mise en concurrence des hôpitaux (Belorgey, 2010 ; Mougeot, 2000).

Mais ces GHM ne sont pas une nomenclature dans la mesure où, contrairement aux activités industrielles produisant un ensemble discret de biens, les activités de services engendrent un ensemble continu (potentiellement infini) de résultats. L'homogénéité revendiquée par les GHM n'est qu'une façade. Depuis la première liste de GHM publiée en 1983, leur nombre a explosé passant de 471 pour la version 0 à 2291 pour la version 11. Par ailleurs, les codages des diagnostiques ne vont pas sans poser de problèmes. Ils sont effectués à partir de la Classification internationale des maladies (CIM) et de la Classification commune des actes médicaux (CCAM). La CCAM est un classement qui ambitionne de rassembler tous les actes médicaux (dans le but de faciliter la rémunération des médecins avec le paiement à l'acte, …). 31 versions ont été publiées depuis sa création en 2002 avec plus de 7000 actes. La CIM en est, elle, à sa 10ème version (CIM-10) avec l'élaboration actuellement de la 11ème. Plus de 150 000 codes sont disponibles. Là encore la question de la construction de ces listes qui se veulent être des nomenclatures est à poser. Si la CIM est peu connue, son 5ème chapitre consacré à la psychiatrie est lui à l'origine de nombreuses controverses. Reprenant le très contesté DSM (*Diagnostic and statistical manual of mental*), ce chapitre du CIM-10 connaît le même sort (Gonon, 2013, Di Vittorio et al., 2013).

Au final, il apparaît que le travail du médecin est imperméable à toute tentative de nomenclature parce qu'il y a une absence de définition claire du produit. Les GHM ne sont pas une nomenclature au sens où, par définition, ils ne peuvent pas être exhaustifs et ne sont pas indépendants des conditions de leur construction. Il s'agit plutôt d'une « échelle de complexité, à un moment donné, des cas à traiter » (Gadrey, 1991).

Puisqu'il n'y a pas d'homogénéité et une définition claire des maladies et des services associés, le soin ne peut pas être dispensé comme un produit. L'incomplétude contractuelle ne peut être résolue par l'évaluation quantifiée du produit. Il parait évident que dans ces conditions le paiement à la performance est voué à l'échec. Toutes les tâches non rémunérées sont susceptibles de ne plus être effectuées et l'accroissement du nombre d'indicateurs est synonyme de bureaucratisation (Hibou, 2012) et de perte de sens du travail (de Gaulejac, 2009). L'objet de l'indicateur est incertain (incertitude sur la « bonne » épistémologie) et la mesure est sujette à caution (sélection des tâches, *gaming*, etc.). Les limites de la « dictature du chiffre » (Fouquet, 2010, p. 316) invitent à repenser l'évaluation de la politique publique. Voilà peut-être une explication concernant le peu d'efficacité de ce type de mécanisme. On peut à nouveau se demander dans quelle mesure la mise en place du paiement à la performance demeure un outil de régulation raisonnable de la médecine ?

2.3.3. Le patient est au cœur du service : la « coproduction »

Une autre conséquence importante de la définition du service par Jean Gadrey porte sur la place du client. Il n'en a aucune dans l'organisation industrielle : il est en dehors du processus de production. Le produit existe indépendamment du client, c'est pourquoi il est possible d'organiser la fabrication sans se soucier du client. Il ne s'agit pas de dire que le client est un étranger au royaume de la production industrielle. Au contraire, le produit est destiné à être vendu ce qui implique que le produit soit pensé pour plaire au client. Mais, cette présence du client est fictive au sens où, matériellement, il n'est pas là au moment de la production – dans l'usine.

Dans les services (de soin), le client fait partie de la production. Sans lui pas de production : sans malade, pas de maladie à soigner et pas de guérison possible. Le patient apporte sa maladie avec lui. On peut repérer trois dimensions de la matérialité du service (Gadrey, 2008) dont il est aisé de comprendre les conséquences dans le cas de la santé :

- Le service implique un déplacement physique des personnes – le plus souvent du patient, bien que, plus rarement, le médecin est également amené à se déplacer. Or, rien n'oblige le patient à se déplacer pour aller consulter son médecin. Il existe d'ailleurs une différence importante entre les patients inscrits auprès d'un médecin traitant et les patients consommant des services de soins au cours l'année[127]. Ces déplacements génèrent par ailleurs

[127] Cf. Chapitre 5.

des coûts (titre de transport, véhicule, carburant, etc.) qui peuvent potentiellement expliquer une partie du renoncement aux soins pour raisons financières.

- Le service implique la matérialité des espaces de relation. Autrement dit, le service à lieu « quelque part ». Pour la relation de soin en médecine ambulatoire, ce lieu est le cabinet du médical. Or, les cabinets médicaux ne sont pas répartis de façon régulière sur le territoire. La thématique des « désert médicaux » illustre bien le fait que dans de nombreux cas, l'accès aux soins est limité ce qui a des conséquences sur l'état de santé des patients et sur la charge de travail des médecins restants. Dans les zones à densité médicale plus élevée, les médecins sont conduits à adopter des stratégies de sélection des patients qui ont également une conséquence sur le *case mix* auquel ils font face. Par exemple, certains médecins généralistes développent des formes de médecine alternative (homéopathie, acuponcture, kinésithérapie, etc.) ou se spécialisent sur certains types de patients (patients CMUC, patients jeunes, etc.).

- La matérialité des outils techniques d'appui à la relation est également un élément important distinguant la relation de service de la production. Si en bout de chaine, l'observation profane affiche la relation nue patient/médecin, il est aisé de dévoiler l'ensemble des supports qui concourent à la qualité du service rendu. La question du suivi du patient à travers le temps est importante de ce point de vue. L'outil informatique est aujourd'hui assez unanimement considéré comme un préalable nécessaire pour le suivi médicalisé du patient – les modalités concrètes faisant néanmoins l'objet de conflits important comme le montre la difficile mise en preuve du Dossier médical personnalisé. Comme nous l'avons souligné au chapitre 5, la ROSP a pris en considération ce problème dans la mesure où il est possible d'interpréter les indicateurs d'organisation du cabinet comme une « subvention cachée ». Cependant, le questionnement sur les outils techniques d'appui à la relation ne s'arrête pas là si l'on étend la notion d'appui à la relation aux personnels hors-médecin. En effet, le service est-il le même lorsque le cabinet médical emploi du personnel administratif, des infirmiers ou des paramédicaux ?

Dans ces conditions, il peut paraitre aberrant d'attribuer au médecin la responsabilité pleine et entière de l'état de santé de ses patients. Et si le patient ne vient pas ? Et si la zone géographique d'exercice réduit le *case mix* ? Et si le cabinet souffre de sous-investissement ? Quelles que soient les réponses, peut-on être assuré que la responsabilité du médecin est seule en jeu ? Que la variété des situations ne résulte que de son mérite ? Si la réponse à ces questions est non, la ROSP ne peut pas avoir pour vocation de récompenser l'implication du médecin.

En plus de ces problèmes liminaires, la coproduction du soin oblige l'évaluateur rigoureux à se questionner sur les deux coproducteurs : le médecin – ce qui est fait par la littérature – et le patient – ce qu'il reste à faire. Evaluer le médecin passe nécessairement par la prise en compte des caractéristiques du patient : âge, éducation, sexe, catégorie socioprofessionnelle, état de santé, etc. Le patient participe à l'élaboration du service. En quelque sorte, il travaille aussi puisque de son comportement dépend la qualité du service reçu. Il faut dès lors relativiser la question de la performance selon ces critères.

Pensons par exemple à la question du respect par le patient des consignes du médecin. C'est le paradigme de l'observance (ou *compliance* en anglais). Une littérature croissante insiste désormais sur le rôle du patient dans sa propre guérison. L'Organisation Mondiale de la Santé a publié un rapport à ce sujet : *Adherence to long-term therapies : evidence for action* (2003). Ce rapport s'intéresse à l'observance des prescriptions médicales dans le cas de maladies chroniques. Par exemple, il explique dans le cas du diabète de type 2 que la faible observance engendre deux effets principaux : un surcoût pour les systèmes de santé et l'augmentation de la souffrance des patients. Le contrôle du diabète repose sur autre chose que la simple prise de médicaments : autocontrôle du taux de sucre dans le sang, régime équilibré, examens ophtalmologiques, etc. Le phénomène est si important qu'on estime aux États-Unis que seuls 2 % des malades adoptent le comportement requis par leur état de santé (comportements définit par l'*American Diabetes Association*). Des conclusions similaires sont tirées par ce rapport concernant d'autres maladies chroniques : asthme, cancer, dépression, épilepsie, VIH, hypertension, tabagisme et tuberculose.

La coproduction du soin est en conséquence un élément décisif lorsqu'il s'agit d'évaluer le service. Contrairement au produit dont l'évaluation est directe, le service donne lieu à deux types d'évaluation (Gadrey, 1991). D'une part, le service immédiat est l'obligation de moyen (sur laquelle est fondée la plupart des normes de bonnes pratiques). Traditionnellement, le médecin a une obligation de moyens étant données « les règles de l'art» ces dernières excluant la plupart des normes de bonnes pratiques. Quelles sont alors les obligations de moyen du médecin si le « droit mou » (le *corpus* des normes de bonnes pratiques) tend à être reconnu et privilégié pour fonder la responsabilité du médecin ? Le patient n'encoure-t-il pas les risques évoqués précédemment liés à la quantification de l'« art médical » ?

D'autre part, le résultat du service est une seconde source d'évaluation. Dans le cas de la médecine, il s'agit de l'atteinte d'un état de santé particulier. Le résultat attendu n'est pas toujours la guérison, il peut être la stabilisation d'un état de santé déjà mauvais ou le

ralentissement de la progression de la maladie. Evidement le résultat est encore plus incertain que les effets immédiats du service. Cette incertitude est évidement liée au patient (santé, morphologie, observance, etc.), mais aussi aux conditions sociohistoriques dans lesquelles vivent le médecin et le patient.

Le cas du traitement de l'obésité est utile pour illustrer cette idée. Analysant la perception par les pédiatres des contraintes liées à la prise en charge de l'obésité, Franc et al. (2009) rappellent quelques-uns des obstacles à un tel travail : contexte socio-économique défavorisé, environnement alimentaire peu favorable, accès difficile au système de santé, lobbying de l'industrie agroalimentaire, etc. Certains auteurs ont pu récemment démontrer la relation entre la proximité de *fast-foods* des écoles californiennes et la prévalence de l'obésité chez élèves du secondaire (Davis et al. 2009). L'étude montre également que ces élèves consomment moins de fruits et légumes et plus de sodas que ceux qui sont relativement plus loin des *fast-foods*. Puisque les médecins reconnaissent la relation entre obésité et diabète, ne peut-on pas s'interroger sur le sens des nouveaux indicateurs de performance de la convention médicale de 2011 dans la mesure où ils portent sur des critères de résultat concernant les dosages d'HbA1c et de LDL cholestérol ? Il parait raisonnable de se soucier de l'amélioration de l'état de santé des patients diabétiques, mais peut-on en imputer l'entière responsabilité au médecin ? Voilà en partie comment il est possible d'expliquer la réticence des médecins face aux normes de bonne pratique dans le cas de diabète de type 2 (Bachimont et al., 2006).

Une dernière caractéristique des services de soin porte sur l'état de l'autre qui fait l'objet de la relation : il souffre. Or, comme l'a montré le chapitre 2, cette souffrance est à l'origine d'une éthique en santé, cette dernière influençant l'éthique des médecins. Les services de soins mettent donc en jeux quelque chose d'anti utilitaire qui, par construction, ne peut pas être mis en comptes. L'attention, le sourire, l'écoute, etc. ne se mettent pas en chiffres. Ces comportements étant mis en concurrence avec le respect des normes quantifiées, ils sont vouées à se marginaliser alors qu'ils constituent un pilier essentiel, parfois de la guérison du malade, souvent de la qualité des soins (Caillé, 2008). Selon Philippe Chanial (2010), le *New Public Management* en santé organise dans ces conditions l'éviction du « souci de l'autre » par le « souci de l'acte ». L'inestimable de la relation de soins fait alors place au geste technique ce qui, paradoxalement, se fait au détriment même de l'efficacité des soins.

Au final, ce qui est en jeu avec la coproduction des soins, c'est la politique de santé en général. Toutes les questions discutées précédemment sont hiérarchiquement dominées non

seulement par la problématique de l'organisation de la dispense de soin mais aussi par la question plus générale des déterminants de la santé.

Du point de vue de l'organisation des soins, la France ayant un système essentiellement tourné vers le soin curatif, il semble logique qu'il y ait des défaillances sur le plan de la prévention. Celles-ci ne sont pas entièrement imputables au médecin, celui-ci n'étant qu'un maillon du système de santé. Le résultat des soins est évidemment dépendant de l'action du médecin. Cependant, il reste le résultat d'une organisation collective elle-même dépendante du contexte sociohistorique (Palier et Bonoli, 1999, Grignon, 2009). La comparaison des systèmes britannique et français est exemplaire à cet égard. Indépendamment de toute évaluation, il semble clair qu'une organisation libérale comme en France produise des effets dissemblables à l'organisation en cabinet de soins primaires au Royaume-Uni. La mise en place de politiques publiques telles que la ROSP et plus largement la quantification sont nécessairement incapables de remédier à un problème qui les dépasse : l'absence d'organisation de la médecine ambulatoire en France. L'attribution d'une performance individuelle (bonne ou mauvaise) à un individu particulier placé dans une organisation collective mérite d'être questionnée. Rémunérer un médecin à la performance met de côté la contribution décisive du patient et du système de santé à l'activité de soin.

De façon encore plus générale, les limites de la quantification de la qualité des soins invitent à questionner la place de la politique de santé en France. En effet, la ROSP est un instrument de « bout de chaîne » qui cherche à améliorer la qualité de soins et l'état de santé par les professionnels de santé. Si cette ambition est légitime, elle masque le rôle de facteurs autrement plus importants dans l'état de santé de la population. Il est utile d'organiser le suivi du diabète, mais n'est-il pas aussi important de prévenir cette maladie par une régulation de l'industrie agroalimentaire ? Il est utile de prévenir les différents cancers, mais n'est-il pas aussi important mettre en place une régulation des activités polluantes ? Il est utile d'orienter les prescriptions médicales vers les médicaments génériques, mais n'est-il pas aussi important de réguler l'industrie pharmaceutique ?

Conclusion

Les essais cliniques randomisés et, plus largement, l'*Evidence Based Medicine* sont de toute évidence une avancée fondamentale de la médecine contemporaine. L'usage des

statistiques pour repérer les régularités des réactions des corps aux traitements ne doit pas être balayé d'un revers de la main. En plus de diversifier les sources de preuves, l'usage de l'EBM permet de faire le tri dans la vaste connaissance médicale. Avec le développement rapide des technologies de communication et d'information depuis les années 1990, les données médicales disponibles sont pléthoriques et doivent faire l'objet d'évaluations qui peuvent se nourrir de cette méthodologie.

La problématique liée à l'EBM qui importe dans cette thèse porte sur la rencontre entre EBM et *New Public Management*. Lorsque la puissance publique s'accapare certains résultats issus des essais cliniques randomisés et systématise le recours à des données quantitatives – parfois même sans qu'elles soient issues de cette méthode, entretenant une confusion –, alors les dangers apparaissent. C'est l'usage social qui est fait de la donnée chiffrée qui est en cause. Le recours systématique à ces chiffres, *via* notamment des outils comme le paiement à la performance, pose en effet problème parce qu'il renforce l'idée fausse selon laquelle la qualité des soins serait solvable dans les quantités. On peut repérer 4 grands types de critique à cette assertion :

(i) Il n'est pas certain que le pathologique soit une variation quantitative de la maladie. La norme quantifiée du normal et du pathologique n'existe peut-être pas et il faut penser la délivrance du soin de santé comme l'adaptation permanente au cas particulier.

(ii) L'essai clinique randomisé n'est pas une méthode produisant des résultats purs et objectifs, dénués de biais. C'est une construction sociohistorique produisant des résultats nécessairement orientés – notamment vers les sources de profit pour l'industrie pharmaceutique qui finance ces essais.

(iii) Il n'est pas possible d'établir une nomenclature des produits dans le secteur des soins. Il y a autant de service de santé que de patients différents. Il est alors trop « coûteux » - pour ne pas dire impossible – de dresser à l'avance la liste des standards à appliquer pour chaque cas.

(iv) Il n'est pas possible d'évacuer la participation d'autres acteurs que le médecin à la production des soins. Le service de soins est une co-production engageant au moins le patient et plus largement le reste du monde social (organisation du système de santé, industrie alimentaire, industries polluantes, normes sociales portant sur la pratique sportive, etc.).

Il apparaît que le soin et la santé – fondement de la Rémunération sur objectif de santé publique en France – ne sont pas tant une production individuelle du médecin qu'une « servuction » collective. Il ne s'agit pas évidemment de dire que le médecin n'a aucune responsabilité dans le système de santé. De façon plus mesurée, l'argument est que le médecin n'est qu'un maillon d'une chaine plus vaste sur lequel il est douteux de faire peser tout le poids du système de soin. Comme à l'accoutumée dans le système de santé français, même lorsqu'il s'agit de penser à la prévention et au suivi des maladies chroniques le paradigme reste le curatif – c'est-à-dire que le soin et la santé proviennent nécessairement et en priorité de l'action du médecin. Or, par exemple, si les diabétiques français sont en moyenne en meilleure forme que les diabétiques états-uniens, faut-il en déduire que c'est grâce ou à cause des médecins ? Il y a comme un paradoxe à vouloir régler un problème de santé publique avec des moyens « d'urgence ».

L'émergence de l'EBM reste néanmoins une avancée de la médecine moderne et pose un problème légitime à la puissance publique chargée d'organiser le système de santé : comment faire tenir ensemble le général provenant de l'usage des statistiques impersonnelles et le particulier de la rencontre personnelle entre le médecin et son patient ?

Imaginons le problème suivant. Au niveau général, si un traitement A est efficace pour 80 % de la population, mais qu'il n'est utilisé que dans 30 % des cas, il est clair qu'il existe une sous-utilisation du traitement entrainant une faible qualité des soins. C'est à cette situation que fait face le médecin de santé publique. Au niveau particulier, si un patient fait partie des queues de distribution, c'est-à-dire qu'il fait partie des 20 % pour qui le traitement A ne marche pas, il a tout intérêt à ce qu'une règle uniforme imposant le traitement A n'existe pas. En effet, pour lui, appliquer le traitement A, c'est 100 % de chance de rester malade – ou pire. Le fait que le traitement A soit efficace pour 80 % des malades n'a aucun intérêt pour lui. C'est à cette situation, à cette angoisse permanente que fait face le médecin dans son cabinet.

L'impensé de la politique de santé est là : comment doivent s'articuler dans la pratique médicale le général et le particulier ? Comment tirer le meilleur de ces deux sources de légitimité ? Historiquement en France, il semble que jusqu'au début des années 1990 la politique de prix assumait implicitement la prévalence du particulier alors que depuis le second grand tournant de la médecine libérale, c'est le général qui semble prévaloir. Dans un cas comme dans l'autre, l'articulation du général et du particulier ne sont pas pensés par la

politique publique. Soit l'on pense particulier, soit l'on pense général – mais jamais les deux simultanément.

Chapitre 8

Une analyse conventionnaliste de la régulation industrielle des pratiques médicales

Introduction

Le chapitre 6 de cette thèse a analysé l'évolution historique de la régulation de la médecine libérale. Si cette histoire a été marquée par une lutte incessante entre la puissance publique et la profession médicale autour de la définition d'un tarif commun, le début des années 1990 est le témoin d'une rupture stratégique de la part du régulateur. Avec l'introduction de références médicales opposables en 1993, la puissance publique change radicalement de stratégie : plutôt que de réguler l'offre de soin par les prix, elle préfère la réguler par les pratiques. C'est le second grand tournant de la médecine libérale – dont l'introduction de mécanismes de paiement à la performance médicale[128] est un outil.

Le chapitre 7 a permis de définir cette rupture : la réforme cherche à imposer la quantification de la qualité des soins. Ce projet, qui se présente sous les aspects de la neutralité et de l'impartialité, rencontre pourtant de fortes résistances, qui apparaissent légitimes. Dès lors, nous sommes en position de relire la rupture qui s'opère depuis les années 1990 comme l'opposition entre deux conceptions concurrentes de la santé sans qu'il ne soit possible de dire que l'une des deux est plus « rationnelle » que l'autre.

L'objectif de ce chapitre est de proposer un cadre théorique – celui de l'Économie des conventions – permettant de comprendre le second grand tournant de la médecine libérale en définissant et en étudiant les deux conceptions de la santé qui s'opposent.

L'opposition porte sur la conception de la qualité des soins. Comme l'a montré le chapitre précédent, le second grand tournant de la médecine libérale s'appuie sur une conception standardisée et normalisée de la qualité de soins. Plutôt que de laisser la médecine aux médecins, il s'agit de définir à un niveau centralisé les protocoles pertinents pour la

[128] En France, le Contrat d'amélioration des pratiques individuelles en 2009 et la Rémunération sur objectif de santé publique en 2011 – ce dernier dispositif faisant l'objet du chapitre 5.

pratique médicale. Le rôle de l'*Evidence based medicine* et des agences de santé indépendantes est alors central pour certifier la qualité des soins. Cette conception s'oppose à une définition plus traditionnelle de la qualité des soins selon laquelle la relation de soin est si singulière qu'il est impossible de prévoir une nomenclature complète des traitements. L'autonomie du médecin face à l'État et au marché est alors valorisée parce que c'est la seule garantie d'un jugement indépendant. Cette liberté n'est pourtant pas gratuite puisqu'elle implique le respect de valeurs professionnelles éloignant le comportement du médecin d'un comportement purement opportuniste. À l'autonomie de la pratique doit répondre le désintéressement de la profession médicale. La qualité du travail médical repose alors de façon fondamentale sur les engagements éthiques de la profession[129].

Ces deux représentions du travail médical peuvent être soutenues par des arguments légitimes : d'un côté on met l'accent sur la médecine comme un art, valorisant l'autonomie professionnelle et l'expérience clinique au bénéfice du patient. La confiance du patient envers un médecin est centrale et doit reposer sur la crédibilité de l'engagement éthique du professionnel. De l'autre, on valorise l'excellence scientifique et l'homogénéité du traitement pour une même pathologie. Chaque patient doit pouvoir bénéficier de la même médecine. Le standard doit être labélisé pour être digne de confiance. Le patient s'en remet alors davantage aux arènes scientifiques qui ont homologué la prestation médicale qu'à l'engagement éthique du médecin particulier.

Nous nous appuyons sur la notion de convention de qualité pour saisir ce conflit de représentation. La première représentation de la qualité des soins repose sur une convention que l'on peut appeler « inspirée/domestique ». Elle valorise les relations de proximité entre un patient et un médecin et mobilise la créativité du médecin pour répondre à l'incertitude face à la maladie. La seconde est portée par une convention « industrielle ». Elle s'appuie sur la standardisation des soins et les données de la science pour mettre en protocole le travail médical (1.).

Ce conflit portant sur la définition de la qualité des soins n'est pas sans impact sur la conception que ce font les acteurs (médecins comme patients) du système de santé (2.). En

[129] Comme l'a suggéré Arrow (1963), l'autonomie accordée au professionnel repose sur les institutions de la confiance. L'existence de ces institutions dépend de l'histoire propre à chaque pays. En France, le chapitre 6 a montré comment la profession médicale s'est construite contre l'État au nom de la qualité des soins. Dans cette perspective, le libéralisme médical est avant tout une manière de s'assurer de l'autonomie de décision du professionnel. Les codes de déontologies et autres règles éthiques viennent garantir que l'autonomie octroyée à la profession ne sera pas utilisée de façon instrumentale. Comme l'a montré Karpik (1989) au sujet des avocats, l'autonomie concédée à la profession médicale relève de la logique du don/contre don.

premier lieu, en apportant une nouvelle définition de la qualité des soins, la convention industrielle modifie l'éthique médicale. En effet, le comportement attendu du médecin change selon la convention de qualité de soins. Alors que dans le cadre de la convention inspirée/domestique les valeurs hippocratiques sont essentielles pour établir le lien de confiance entre médecin et patient, avec la nouvelle convention, il n'est attendu du médecin aucune vertu morale dans la mesure où la confiance provient des agences indépendantes. Autrefois dépositaire de la santé, le médecin devient un maillon parmi d'autres de la division du travail sanitaire.

En second lieu, c'est la perception du patient qui se modifie. L'industrialisation de la médecine constitue un terreau de la marchandisation du soin. La standardisation est en effet en phase avec la comparaison et la mise en concurrence. Elle invite le patient à faire usage de son pouvoir de consommateur. Pour que la concurrence soit possible, les biens doivent être homogènes. Or, le rôle de la nouvelle convention est justement d'homogénéiser les biens. Ce processus permet par ailleurs de désingulariser la relation de soin de façon à ce qu'il soit plus facile pour les patients/consommateurs d'arbitrer entre les médecins/producteurs – sur la base d'un produit homogène. À travers la régulation de la médecine libérale c'est donc la conception des soins de tous les acteurs du système de santé qui est en jeu.

1. La convention de qualité du « soin » : convention inspirée/domestique versus convention industrielle

Comme nous l'avons vu dans les précédents chapitres, le second grand tournant de la médecine libérale est une preuve éclatante du fait que la définition du « produit de soin » ne va pas de soi. Alors que dans la théorie *mainstream* les biens sont homogènes par hypothèse (c'est « l'hypothèse de nomenclature »), dans le monde réel, la définition du bien et de sa qualité est le résultat d'un processus sociohistorique. La production de normes de bonnes pratiques, de *guidelines*, de systèmes d'information, de références médicales opposables, l'appui sur les agences indépendantes, la démarche qualité et les processus d'accréditation, etc. visent à canaliser la définition d'une bonne façon de soigner. C'est parce que les biens et services de santé sont foncièrement hétérogènes que les dispositifs de réduction de l'hétérogénéité ont été mis en œuvre. Cette industrie de l'industrialisation n'a en effet pas de sens dans un monde où les produits seraient homogènes.

Elle ne cherche pas uniquement à combattre l'hétérogénéité des biens et des services de santé mais à justifier cette entreprise au nom d'arguments légitimes. Il existe en effet tout un registre d'argumentation qui permet de valoriser le monde industriel qui se construit en matière de santé mais aussi de critiquer l'ancien monde fondé sur l'autonomie professionnelle. La standardisation des soins ne manque pas d'arguments pour la mettre en œuvre et pour imposer la quantification contre les « règles de l'art ». Elle fait ainsi valoir la capacité de manipulation des « règles de l'art » par les médecins et le poids du corporatisme

D'une façon symétrique, l'ancien monde professionnel dominé par la singularité de la relation des soins est aussi tout à fait légitime. Il s'appuie sur une vision du soin qui peut emporter l'adhésion quand elle met en avant le patient particulier dont le médecin doit soulager la souffrance et non le patient statistique moyen. La rigidité des normes et des standards peut être dénoncée quand elle s'oriente vers la routinisation des actes et la dé-professionnalisation de la médecine.

Il existe ainsi une pluralité des modes d'évaluation de la qualité du soin de santé. Il n'y a donc pas qu'une seule façon de définir la qualité dans le secteur des soins et il serait inexact de présenter l'opposition entre ces différentes façons de faire comme des résistances illégitimes à une sorte de « rationalisation » consubstantielle à la « modernité ».

C'est dans cette perspective que nous mobilisons la notion de convention de qualité qui désigne à la fois des ressources de justification et des ressources pour la critique (1.1). Fort de l'appareillage théorique de l'Économie des conventions, nous opposons la convention industrielle qui conçoit le service de soin comme un produit et met en avant les standards de qualité (1.2.) à la convention inspirée/domestique qui se nourrit de l'autonomie professionnelle au sens de la sociologie des professions (1.3.). Nous récapitulons l'opposition entre ces deux types de convention dans une dernière section (1.4).

1.1. Le paradigme de l'Économie des conventions et les conventions de qualité

La reconnaissance de la pluralité des formes de l'accord dans la définition de la qualité des produits est l'un des fondements majeurs de l'Économie des conventions (EC). Ainsi, les travaux de François Eymard-Duvernay (1989, 1993, 2002) ont mis en avant le fait que la qualité des biens ne peut se réduire à des critères exclusivement marchands. Il s'agit, en particulier, de reconnaitre que « l'opposition traditionnelle entre l'efficacité économique du marché et le rôle protecteur des règles est rendue plus complexe par le développement des

modes d'action où la frontière entre ces deux ordres de référence est brouillée » (Eymard-Duvernay, 1989, p. 329). Au contraire, il existe une pluralité des formes d'accord que les agents peuvent prendre en compte. En ce sens, l'histoire de la régulation de la médecine libérale révèle en creux la pluralité des modes d'évaluation du soin. Cette pluralité s'inscrit aussi bien au niveau de la règle qu'au niveau des représentations des acteurs.

En effet, l'EC définit la notion de convention à un double niveau (Favereau, 1986, 1999, Batifoulier, 2001). Au premier niveau, la convention est un type particulier de règles, entre la règle-contrainte et la règle-contrat, qui se caractérise par quatre critères : formulation vague, origine obscure, caractère arbitraire et absence de sanction juridique. Le code vestimentaire est l'exemple typique de règle conventionnelle. Lors d'un entretien d'embauche dans une banque, le costume et la cravate sont de rigueur. De formulation vague, ces règles ne présentent pas la dimension coercitive des règles-contrats ou des règles-contraintes. Leur origine est obscure et elles sont arbitraires au sens où une autre solution pourrait faire l'affaire : peu importe que la cravate soit remplacée par un nœud papillon ou un panama tant que la coordination est réalisée.

Mais à ce premier niveau, qui est loin d'être le plus important, s'ajoute un second niveau qui constitue véritablement l'originalité de l'EC : pour toute règle de comportement et pour toute discussion sur la qualité des produits, il existe une marge d'interprétation. Les individus jugent les règles pour les appliquer. Ce jugement fait appel à des principes normatifs qui ont les mêmes caractéristiques qu'une convention (arbitraire, formulation vague, histoire floue, etc.) : « on verra dans cette forme de conventions un trait constitutif de l'ensemble des règles et non plus seulement une catégorie de règles parmi d'autres. Le message devient : pas de règles, ou plus généralement pas d'institutions, sans convention sous-jacente » (Favereau, 1986, p. 167).

La convention est donc aussi et surtout un modèle d'évaluation d'un produit, d'une règle ou d'une action. En reprenant l'exemple précédent, si on peut avoir du mal en théorie (et en théorie économique particulièrement) à expliquer pourquoi on opte pour une cravate, un nœud papillon ou un panama, ce n'est pas un réel problème pour les individus s'ils sont capables d'évaluer la situation et le contexte dans lequel ils se situent (un entretien d'embauche, une réunion de famille, un mariage, etc.).

Cette représentation de la situation est une représentation du « collectif » dans lequel l'individu s'insère, à condition d'accorder au mot « collectif » une acception large. Le

« collectif » peut aussi bien désigner une interaction interindividuelle (comme l'interaction entre un recruteur ou un recruté ou celle entre le patient et le médecin) que l'ensemble de l'entité collective (l'entreprise ou le système de santé dans leur globalité).

Les représentations ne sont conventionnelles que si elles permettent la coordination. Elles doivent donc être partagées ou collectives. Aussi, la convention se définit par une représentation collective du collectif. Les règles n'étant jamais complètes et les produits n'étant jamais homogènes, la coordination entre les acteurs passe par l'accord sur des « modèles d'évaluation » (Biencourt et al. 2001). Le modèle des économies de la grandeur de Boltanski et Thévenot (1991) a fourni une première approche permettant de distinguer les différentes conceptions du « bien ». Les « cités » sont des conventions générales qui organisent la justification et la critique des actions individuelles au nom du bien commun – en rapport avec un nombre limité de philosophies politiques (Encadré 8.1.). Il n'y a pas, en ce sens, une cité plus « rationnelle » qu'une autre dans la mesure où elles renvoient toutes à une conception légitime et incommensurable du collectif. La décision n'est pas qu'une question de calcul mais un choix politique entre plusieurs communs possibles.

Sur la base de ce modèle des économies de la grandeur, il est possible de repérer différentes conceptions de la qualité. Eymard Duvernay (1989) oppose ainsi une conception industrielle de la qualité qui relève de standards chiffrés, où les procédures font l'objet d'une objectivation poussée, à une convention domestique de la qualité, fondée sur la marque et l'apprentissage *via* des liens de proximité. Cette opposition a donné lieu à de nombreux développement de travaux conventionnalistes particulièrement féconds qu'ils s'appliquent à un secteur d'activité particulier comme l'industrie laitière française (Boisard et Letablier, 1987) ou plus généralement au monde de l'entreprise (Thévenot, 1989, Eymard Duvernay, 2004).

Encadré 8.1. Les Économies de la grandeur : une pluralité de valeur

Le modèle des Économies de la grandeur, développé par Boltanski et Thévenot dans « De la justification » (1991), propose une théorie de la coordination des comportements comme des représentations. Il s'appuie alors sur les justifications observées lors des disputes ordinaires. En situation de conflit, les acteurs vont mobiliser des arguments pour défendre leur point de vue. À partir d'enquêtes de terrain et de la littérature de philosophie politique occidentale, Boltanski et Thévenot effectuent une taxinomie des différents registres du « bien » mobilisés par les agents. Ces registres de justification définissent ainsi différentes cités.

Boltanski et Thévenot identifient six registres de justification qui correspondent aux six « cités » : civique, industrielle, marchande, domestique, inspirée et du renom. Boltanski et Chiapello (1999) en ajoutent une septième, la cité par projet. Le tableau, ci-dessous, permet de qualifier chacune de ces cités en fonction des principes de justifications qui leur(s) sont associés et d'un appui théorique et empirique de référence.

Cités	Principes de justification	Référence
Marchande	Concurrence, intérêt particulier	Smith - guide pour réussir dans les affaires
Civique	Volonté générale, action collective	Rousseau - guide syndical
Industrielle	Efficacité, performance	Saint-Simon - guide de productivité
Domestique	Proximité, tradition	Bossuet - guide de savoir-vivre
Du renom	Prestige, reconnaissance de l'opinion	Hobbes - guide de relations publiques
Inspirée	Jaillissement de l'inspiration	Saint Augustin - guide de créativité

Il y a donc une pluralité de mondes possibles. Ceci ne suppose pas que chaque individu s'inscrive dans un monde particulier, ni que les individus agissent sciemment en suivant les préceptes de l'une de ces cités. Au contraire, les agents en fonction des interactions avec les autres, peuvent sans difficultés passer « d'une cité à l'autre ». De même, les justifications apportées à chacune de ces activités sont implicites ; les individus ne les expliciteront qu'en cas de conflit. La coordination suppose alors un accord sur un principe commun ou sur la réalisation d'un compris, qui peuvent émerger entre différents registres de justification.

Pour distinguer ces différentes formes d'actions, Thévenot a ainsi été amené plus récemment à doter l'EC d'un « second pluralisme » (Thévenot, 2006). Les modalités d'engagement sont alors hiérarchisées selon leur plus ou moins grande capacité à être étendues au-delà de la situation particulière. Si toutes les situations ne mettent pas en cause le bien commun, elles s'appuient toujours sur une forme ou une autre de normativité plus ou moins générale et généralisable.

Nous tenterons ici, en s'inspirant de ces travaux, de montrer que le champ de la santé n'échappe pas à cette conceptualisation. Le second grand tournant de la médecine libérale est exemplaire de la confrontation entre deux conventions de la qualité des soins : la convention (de qualité des soins) industrielle cherche ainsi à remplacer la convention (de qualité des soins) inspirée /domestique.

1.2. La convention industrielle de qualité des soins

Dans la mesure où elle fait l'objet du Chapitre 7 de cette thèse, il n'est pas nécessaire de s'attarder longuement sur la convention industrielle de qualité des soins. Rappelons néanmoins ses caractéristiques principales.

La convention industrielle se caractérise par la conception qu'elle a de la maladie : c'est une variation quantitative de l'état normal. L'organisation légitime de la médecine découle de ce point de départ. Puisqu'il est possible d'isoler la maladie du malade, il faut soigner la maladie et non le malade. Le problème médical majeur n'est donc pas l'adaptation à la singularité du cas ou la résolution d'un problème complexe. Il réside dans la réduction de la variabilité des pratiques par le recours à des normes chiffrées. Dans un contexte où l'information est pléthorique et où la formation est famélique, il faut que des agences indépendantes produisent des synthèses scientifiques incontestables. En France, c'est le rôle de la Haute Autorité de Santé. L'*Evidence based medicine* fournit les données médicales qui prétendre être neutres et objectives et qui servent de fondement à la rédaction de normes de bonnes pratiques. Le médecin doit alors utiliser ces standards pour soigner le « patient moyen ». Une fois le diagnostic réalisé, sa tâche principale est d'attribuer le cas spécifique à l'une des catégories générales de sa nomenclature – et de fournir le traitement associé.

Or, « la soumission à des standards de qualité, constitue un élément de base de la logique industrielle » (Eymard Duvernay, 1986, p. 346). Cette convention est industrielle parce qu'elle conçoit le service de soin comme un produit et parce que son objectif prioritaire est de gagner en productivité et en efficacité. La rationalisation dont il est souvent question renvoie à la réduction de la complexité par standardisation. Il s'agit là d'une étape nécessaire pour contrôler le travail médical et faire des économies d'échelle. Cette convention renvoie également à un usage intensif des statistiques. L'indicateur chiffré à une place primordiale dans la nouvelle convention de qualité des soins : la norme chiffrée est le socle privilégié du contrôle de qualité et de l'incitation à la qualité (entre autre, par l'intermédiaire du paiement à la performance médicale).

Notons enfin que l'on peut déceler deux justifications principales à cette convention de qualité. D'une part, la convention industrielle peut relever d'une démarche de santé publique. Face à la variabilité des pratiques, l'usage de normes chiffrées permet d'homogénéiser les comportements pour le bien des patients. Si la norme statistique n'est pas optimale pour les queues de distribution, elle l'est en moyenne. D'autre part, cette convention est totalement en lien avec la théorie économique *mainstream* et le mouvement du *New Management Public* quand ils veulent lutter contre les rentes abusives. L'autonomie professionnelle est alors conçue comme un vecteur de l'opportunisme et la standardisation se présente alors comme une tentative de combattre le pouvoir médical.

1.3. De la sociologie des professions à la convention inspirée/domestique de qualité des soins

En dépit de sa prétention à l'objectivité et à l'universalité, la convention industrielle ne vient pas sur un terrain vierge de considérations normatives sur la qualité de soins. Historiquement, la médecine s'est organisée autour de codes de déontologie, d'ordres professionnels, et autres valeurs hippocratiques qui codifient la pratique. La sociologie des professions offre des ressources précieuses pour analyser cette forme particulière d'organisation.

Deux grandes traditions théoriques traversent l'histoire de la sociologie des professions (Dubar et Tripier, 1998). *La sociologie fonctionnaliste* des professions fait un usage limité de la notion de profession en tant que type particulier de métiers. Dans ce courant, la profession se distingue du métier sur au moins trois critères. D'une part, la profession suppose une compétence propre ou l'usage de savoirs d'un type particulier (impliquant par exemple une formation de haut niveau ou des savoirs scientifiques). D'autre part, le professionnel doit détenir une compétence exclusive ou du moins un statut particulier. Enfin, le professionnel dispose d'une forte autonomie dans la conduite de son travail – cette autonomie étant justifiée par les deux conditions précédentes. L'activité archétype du modèle professionnel est la médecine.

Face à la sociologie fonctionnaliste est né *le courant interactionniste* pour lequel l'usage du concept de profession peut s'étendre à tous les métiers. L'interactionnisme analyse la profession au sens du fonctionnalisme comme une croyance qui confère des avantages illégitimes aux professionnels. Dans le même esprit que celui de la théorie économique *mainstream*, il n'y a pas de différence entre le médecin et le garagiste – si ce n'est le potentiel illimité de l'opportunisme médical (Batifoulier et Gadreau, 2007). Pour la sociologie interactionniste des professions, il n'est pas nécessaire de distinguer la profession du métier. Si pour les fonctionnalistes il y a une différence de nature entre professions et métiers, pour les interactionnistes, il s'agit avant tout d'une différence de représentations sociales.

Récemment Florent Champy (2009, 2011) a proposé un dépassement de cette opposition classique en sociologie. L'une des justifications à cette tentative de synthèse est l'impossibilité qu'a la théorie interactionniste des professions d'expliquer la légitimité des professions (de santé ou autres) lorsqu'elles se défendent face aux attaques faites à leur statut. Plus clairement, le travail de cet auteur cherche à comprendre en quoi la défense des

professions face à la normalisation est légitime. Pour ce faire, contre l'interactionnisme il affirme la spécificité des professions et, contre le fonctionnalisme il abandonne l'idée de définir toutes les professions. Au lieu de définir la profession par le statut, il s'agit de la définir par l'activité. En ce sens, ce sont les caractéristiques du travail qui font d'une activité un métier ou une profession. D'après Champy (2009), la spécificité des professions tient au type de problèmes qu'elles ont à résoudre qui renvoie à la singularité à la complexité des cas (Encadré 8.2.).

Encadré 8.2. Singularité et complexité : deux caractéristiques majeures de l'activité permettent de distinguer les métiers des professions

(i) Singularité des cas auxquels les professionnels sont confrontés. Il est intéressant de noter que cette singularité des cas tient pour l'essentiel de la singularité du « matériau humain » constituant le fondement de l'activité médicale. Ce postulat est concordant avec celui, déjà étudié, de l'économie des services : le soin est singulier parce que le service médical ne peut pas se penser en dehors du patient qui le demande.

(ii) Complexité des situations à partir desquelles travaillent les professionnels : cette complexité renvoie au fait que les problèmes posés au médecin ne peuvent pas être explicités *a priori*. Il est impossible d'établir une nomenclature exhaustive du travail médical. La complexité implique par exemple que, contrairement au métier, le professionnel doit, à un moment ou à un autre, décider d'arrêter de rechercher la solution à un problème. Dans le cas du métier, l'investigation aboutie toujours rapidement à une analyse complète du problème. Le professionnel fait face à une situation ouverte qui ne peut pas avoir de solution *a priori*.

En raison de la singularité et de la complexité des cas, l'incertitude de la décision du professionnel est majeure. Ce sont ces caractéristiques de l'activité qui engendrent la création de modalités spécifiques de gestion de l'incertitude par le professionnel. Plutôt que des capacités de calcul, il est attendu du professionnel des capacités de jugement « Ce travail ne consiste pas – ou pas principalement – à appliquer *mécaniquement* des savoirs scientifiques. Face à des problèmes singuliers et complexes, les professionnels prennent des décisions qui comportent une dimension de *délibération* et même de pari : le travail professionnel est conjecturel, parce qu'il porte sur une réalité qui échappe inévitablement à toute maîtrise systématique. On retrouve ici les idées de réflexivité ou d'inférence : c'est bien parce que la réponse ne coule pas de source qu'une délibération plus ou moins longue est nécessaire » (Champy, 2009, p.84-85).

Le savoir formalisé n'est donc pas suffisant pour le professionnel. Il lui faut également une capacité de délibération qui s'acquiert notamment par l'expérience et la pratique. Selon l'auteur, la médecine est un cas emblématique de cette tension entre savoir et délibération : si le savoir médical est évidemment d'un haut degré d'abstraction et de formalisation, il n'est pas suffisant pour la pratique – qui laisse une grande place à la délibération étant donné la singularité du cas.

Le professionnel fait ainsi face à une double incertitude. D'une part, une forte incertitude porte sur la singularité du cas étudié et, d'autre part, l'incertitude provient de la tension entre les différents principes pris en compte dans la délibération sur les fins de l'action. Ce choix entre plusieurs conceptions possibles du bien, entre plusieurs conventions, est le fondement de la dimension éthique des professions. La délibération sur les fins de l'action implique des dilemmes éthiques constitutifs du travail du professionnel.

Nous nous proposons de qualifier cette convention de la qualité du travail d'inspirée/domestique.

En effet, l'une des conditions essentielles pour que le professionnel puisse gérer l'incertitude convenablement est l'octroi d'une forte autonomie dans son travail. Cette autonomie ne porte ni sur les fins de l'activité (le médecin ne peut pas faire n'importe quoi) ni sur les moyens de l'activité (l'autonomie sur les moyens n'est pas propre aux professions savantes). L'autonomie des professionnels porte sur deux dimensions :

(i) La marge d'interprétation des cas étudiés. Puisque l'activité est marquée par la singularité des cas et la complexité, les médecins ont une marge d'interprétation importante.

(ii) La capacité à délibérer sur la hiérarchie des fins et sur le contenu précis de ces fins dans le cadre des fins générales de l'activité. Etant donné les impératifs généraux de l'activité médicale, les médecins peuvent, à l'intérieur de ce cadre, définir et hiérarchiser les fins de l'activité : ne pas nuire plutôt que soigner, soulager la douleur plutôt que ne rien faire, etc.

En ce sens, la pratique médicale prend les attributs propres à la cité inspirée. La délibération peut ainsi être vue comme un moyen pour trouver l'inspiration afin de répondre à la singularité et la complexité des cas. L'écart entre l'universel et la pratique étant irréductible, le professionnel doit faire preuve de prudence[130] – entendue au sens de sagesse pratique – lors de ses délibérations sur ce qu'il convient de faire. Dans la cité inspirée, « les personnes sont créatives quand elles sont séparées des autres, retirées en quelque sorte en elles-mêmes, dans leur intériorité » (Boltanski et Chiapello, 1999, p. 192).

Mais le caractère inspiré de la convention de qualité de soin du professionnel de santé, ne permet pas, selon nous, de saisir pleinement la complexité de « l'art médical ». L'espace de délibération qui s'ouvre pour le médecin implique de ne pas déconnecter la maladie du malade. Son détachement est donc avant tout métaphysique, sa pratique, quant à elle, est bien attachée au patient, à l'individu dans toute sa singularité avec qui le soignant doit pouvoir rentrer dans une relation de confiance, de proximité. Cette pratique rentre alors en résonnance avec la « personnification des ressources » propre à la cité domestique. « La qualification

[130] Afin d'affiner la description de l'activité du professionnel, Champy (2009) se réfère au concept de prudence chez Aristote – c'est pourquoi l'auteur parle de profession à pratique prudentielle. Le philosophe grec oppose l'universel – relevant de la science – et le particulier – pour lequel la science est inappropriée. Dans les situations particulières, la contingence met le savoir scientifique en défaut d'où l'importance chez Aristote de la délibération. La délibération du professionnel porte tant sur les moyens – dimension technique de la délibération – que sur les fins – dimension politique de la délibération. Dans la mesure où il existe toujours une dissonance entre les fins idéales et les moyens disponibles, la délibération sur les fins n'est pas donnée *a priori* – c'est en cela qu'elle est politique : elle dépend des représentations que l'on se fait du collectif.

domestique des êtres repose, en effet, sur des rapprochements des relations temporelles et en appellent à l'*ancestral*, ou par des correspondances topologiques se référant au *proche*, ou encore par des comparaisons hiérarchiques recourant à l'*autorité* » (Thévenot, 1989). On rejoint là encore les caractéristiques du professionnel de santé lorsqu'il fonde sa pratique sur sa connaissance traditionnelle de la médecine et celle du patient avec qui il est suffisamment proche pour rentrer dans la sphère de l'intime en devenant « le médecin de famille » (bien avant le médecin traitant). La familiarité de cette relation se trouve au fondement de la confiance sur laquelle le médecin peut baser son autorité.

Parce qu'elle revêt à la fois les attributs de la cité inspirée et de la cité domestique, nous nous proposons de qualifier cette convention de qualité des soins d'inspirée/domestique en suivant en cela la logique du « compromis » qui peut exister entre deux ordres de grandeur. En effet, « aucun ordre de justification ne suffit à lui seul à régir les coordinations complexes que réclame une vie organisée en société et la confrontation entre plusieurs formes de justification nécessite la confection de compromis » (Thévenot 1996).

1.4. Convention inspirée/domestique versus convention industrielle

Le passage de la régulation de la médecine libérale par les prix à une régulation par les pratiques peut se lire comme la tentative de la puissance publique de substituer une convention de qualité des soins par une autre. La réforme ainsi que ses résistances expriment le conflit entre la convention inspirée/domestique et la convention industrielle que nous proposons de résumer ici (Tableau 8.1.) :

Tableau 8.1. Convention inspirée/domestique *versus* convention industrielle

	Convention inspirée/domestique	**Convention industrielle**
Définition du pathologique	Rupture qualitative avec le normal	Variation quantitative avec le normal
Objet de l'activité	Couple patient-malade	Maladie
Problème médical résoudre	Singularité et complexité	Hétérogénéité des pratiques
Objet du soin	Malade singulier	Malade moyen
Particularité du savoir médical	Fondé sur l'expérience	Fondé sur les standards
Tâche du médecin	Adaptation au cas	Attribution du cas
Définition de la qualité	Professionnel	EBM
Éthique médicale	Primordiale	Secondaire
Institution de la confiance	Éthique médicale	Agence de certification

L'opposition entre ces deux formes de définition de la qualité des soins peut être développée en revenant sur les attributs fondamentaux des deux conventions.

- **La convention inspirée/domestique :**

L'activité médicale est par excellence un lieu où s'exprime avec force l'incertitude radicale : l'occurrence de la maladie est aléatoire, la connaissance même de la maladie est limitée et dépendante du progrès médical, la réaction du malade au traitement (lorsqu'il existe) est incertaine, etc. Le médecin, aussi doué et aussi bien intentionné soit-il, ne connait donc ni l'ensemble des événements de la nature ni les probabilités associées.

Dans ces conditions, il est impossible d'écrire un contrat de soin complet liant le médecin au patient et à l'assureur. Puisque l'on ne peut pas écrire des contrats complets, comment faire en sorte que les parties prenantes acceptent de prendre part au processus de soin ? Comment s'assurer de la qualité des soins en environnement incertain ?

(a) Le médecin est détenteur d'un savoir spécialisé. Une première justification de l'asymétrie est purement économique : l'investissement dans ce savoir complexe et en mouvement est coûteux. Si le patient était aussi bien informé que le médecin, il n'irait jamais consulter. En fait, si l'on ne s'intéresse qu'à la partie diagnostic, le marché des soins est essentiellement un marché de l'information. Il s'agit pour le médecin, à partir de ses connaissances de traduire des symptômes en diagnostic. Ce qui est échangé étant des informations – par hypothèse inégalement réparties –, pourquoi déduire toutes les recommandations de politiques publiques du « problème » d'asymétrie d'information ? L'existence du médecin est la solution du problème d'asymétrie d'information. Plus

généralement, dans le cas où ce qui est échangé est une information, s'intéresser à l'écart entre les observations et la situation de marché à information parfaite et complète devient absurde. Cela revient en effet à prendre pour étalon une situation où le marché… disparaitrait. Ici, l'asymétrie d'information ne détruit pas les conditions de l'échange, elle les crée : sans asymétrie d'information, pas besoin de médecin. La question pour l'économiste n'est alors pas tant de discuter des asymétries d'information à l'avantage du médecin que de discuter de la prise de décision en situation d'incertitude radicale.

Quel processus de décision adopter ?

(b) Le savoir médical est singularisé. L'existence d'une incertitude radicale dans le domaine médical invite à se poser la question de ce qu'est la maladie et, en creux, de se demander ce qu'est le service de soin. Traditionnellement, l'épistémologie médicale pensait la maladie comme un état de rupture avec la situation normale. Cette rupture étant qualitative seul le malade est capable de définir sa maladie. Dès lors, il existe autant de maladie que de malades et le travail du médecin est nécessairement ancré dans la singularité du patient. L'hypothèse de nomenclature est impossible à réaliser : il est impossible d'attribuer à l'avance à chaque maladie un processus de soin. L'« art médical » est alors l'exercice d'adaptation du savoir médical (par définition imparfait) à la singularité du malade. Une deuxième justification de l'asymétrie d'information est épistémologique : la maladie étant un phénomène singulier, le médecin s'engage dans un processus de décision personnalisée (l'« art médical ») qui a des chances de remporté l'adhésion. Notons que ce qui est en cause ce n'est pas le fait de prendre la meilleur décision mais d'adopter le meilleur processus de décision.

Comment enfin s'assurer de l'engagement du médecin ? Comment se prémunir de son opportunisme ?

(c) La pratique de la médecine est encadrée par des institutions de la confiance. Il est impossible d'expliquer que les patients se remettent à une catégorie spéciale de la population appelée médecin si l'on ne prend pas en compte la construction sociale de la confiance (en la médecine). Les études longues, les barrières à l'entrée, les codes de déontologie, les codes éthiques, le serment d'Hippocrate, etc. sont des éléments qui assurent une certaine légitimité à la profession médicale. Celle-ci n'affiche pas son orientation pro-sociale pour rien. L'éthique médicale a pour but d'encadrer le comportement médical de façon à ce que les patients aient confiance dans la personne du médecin. Une troisième justification de l'asymétrie

d'information est morale : puisque les médecins agissent de façon éthique, il est possible de s'en remettre à eux.

La proposition (a), (b) et (c) forment le cœur de la convention inspirée/domestique. Dans ce cadre, l'incertitude radicale qui règne sur le domaine médical est atténuée par une pratique singulière et « pro-sociale ». Il devient possible de s'en remettre au médecin. La qualité des soins ne pose pas de problème particulier puisqu'elle est assurée par le professionnel. Le patient va alors pouvoir se fier à la fois à l'inspiration du médecin mais également à la proximité qui est née de la relation qui a vocation à tirer sa force du passé pour se poursuivre dans l'avenir. L'incomplétude contractuelle n'est plus un problème mais une solution. En effet, l'écriture d'un contrat complet est non seulement impossible (incertitude radicale et hypothèse de nomenclature) mais aussi très coûteuse. Or, puisque le savoir médical est évolutif, il est préférable de se fier au professionnel plus tôt que de le contraindre par des règles qui, en réduisant sa marge d'autonomie, l'empêche d'adapter sa décision à la singularité de la situation.

- **La convention industrielle**

La convention industrielle est l'antithèse de la convention professionnelle. Elle est fondée sur le refus, en bloc et en détail, des propositions (b) et (c) substituée par les propositions (b') et (c').

(b') Le savoir médical peut être standardisé. Rompant avec l'épistémologie traditionnelle, l'épistémologie positiviste de la maladie analyse la maladie comme une variation quantitative de la situation normale. En ce sens, plutôt qu'une rupture liée à la subjectivité du patient, la maladie est un phénomène purement objectif et quantifiable. S'il est possible de définir et de borner à l'aide des nombres la maladie alors, il est possible de construire une nomenclature des actes et des traitements. En d'autres termes, l'épistémologie positiviste proclame que le soin est un produit. En tant que produit objectif, il est alors souhaitable d'appliquer à la maladie le paradigme productiviste : standardisation, normalisation, démarche qualité, économie d'échelle, contrôle, etc. Un outil caractéristique de cette épistémologie est l'*evidence based medicine*. La preuve médicale ne relève plus de l'analyse singulière des conditions du malade mais du traitement statistique. Soigner la maladie, c'est soigner le malade moyen. Pour soigner les malades, il faut donc réduire l'écart à la norme. L'information pertinente existe et il est nécessaire de produire des systèmes d'information visant à développer les bonnes pratiques.

Comment s'assurer de l'engagement du médecin ? Comment se prémunir de son opportunisme ?

(c') Si certains fondements de l'épistémologie positiviste proviennent de la nécessité de traiter une information abondante et en permanente évolution, d'autres sources s'appuient sur la crise de confiance du système médical. Il ne serait plus pertinent de se fier au médecin. Dans cet esprit, les différentes crises sanitaires (comme celle du sang contaminé) ainsi que le constat d'une variabilité considérable de la pratique médicale poussent à plaider pour la mise en place de dispositifs de contrôle. Il n'est plus alors question de s'appuyer sur les institutions de la confiance traditionnelles. La confiance au médecin s'est mue en défiance. L'asymétrie d'information redevient problématique et les systèmes d'information sont le moyen du contrôle et de l'incitation. Les agences indépendantes et le nouveau régime de la preuve sont le socle de la nouvelle confiance.

La convention industrielle repose sur l'idée que la pratique médicale est susceptible d'être standardisée et qu'elle nécessite d'être contrôlée et incitée afin de s'assurer d'un niveau de qualité minimal. Peu importe le professionnel de santé tant qu'il applique la norme. Le second grand tournant de la médecine libérale implique alors que le médecin se dé-professionnalise en s'industrialisant.

2. Bouleversement de l'éthique médicale et marchandisation de la santé

La section précédente s'intéressait à la convention de qualité des soins – c'est-à-dire à la façon dont il convient de juger l'activité médicale. Désormais, il est nécessaire de replacer la discussion au niveau du système de santé dans sa globalité. Le passage d'une convention inspirée/domestique à une convention industrielle redistribue les cartes du légitime et de l'illégitime. Elle transforme la vision que se font les acteurs du système de santé et bouleverse ainsi leur représentation du collectif. Cette hypothèse est déclinée à un double niveau.

Nous considérons tout d'abord que la nouvelle représentation de l'« action qui convient » (Thévenot, 1990) redessine les contours de l'éthique médicale. Sans toutefois faire table rase du passé, avec la transformation de la convention de qualité des soins, c'est l'éthique médicale elle-même – définie en tant que guide pour l'action – qui évolue. Le développement des standards de qualité modifie en effet la rationalité médicale. Ce qui

pouvait être considéré comme éthiquement correct dans le monde inspiré / domestique l'est moins dans le monde industriel et inversement (2.1).

Par ailleurs, le changement de convention de qualité se traduit aussi par un bouleversement de la vision que peut avoir le patient du système de santé. Le monde industriel rend plus aisé la mise en concurrence. Il est plus facile de comparer avec des standards de qualité qu'avec des prestations uniques et dédiées. Le patient est invité à faire valoir de nouvelles compétences et à se conduire en consommateur. L'industrialisation de la médecine participe alors à la construction marchande du secteur de la santé (2.2)

2.1. L'éthique médicale sous influence de la convention industrielle

Après avoir présenté les « valeurs propres » de la médecine libérale (2.1.1.), nous montrons ce que fait l'introduction de la convention de qualité industrielle à ces valeurs traditionnelles (2.1.2.) : elles préparent un réagencement des valeurs médicales au profit des libertés tarifaire et d'installation (2.1.3.).

2.1.1. Les « valeurs propres » et l'éthique médicale

L'approche interprétative des conventions est extrêmement fertile pour comprendre les enjeux liés au second grand tournant de la médecine libérale. En effet, les travaux conventionnalistes ont démontré, depuis plusieurs années, que l'éthique médicale peut s'analyser comme une convention assurant la coordination dans le système de soin (Batifoulier, 2009). Les règles déontologiques sont par nature incomplètes et évasives. Elles ne peuvent à elles seules prescrire le comportement à adopter. Les codes éthiques n'ont d'intérêt que s'ils sont interprétés par les médecins. Doter les médecins d'une rationalité interprétative dans le cadre de l'Économie des conventions permet de combler l'écart entre la règle et le contenu de la règle. L'interprétation des situations se fonde alors sur les représentations que se font les médecins sur le collectif – sur ce qu'il convient de faire.

L'éthique médicale peut ainsi se comprendre comme l'ensemble pluriel des ressources normatives spécifiques à la profession médicale. Dans cette optique, il est possible d'isoler trois « valeurs propres » qui bornent les contours de l'éthique médicale sans en donner une définition rigide (Batifoulier, 2011). La normativité des médecins est faite d'une diversité de valeurs qui permettent d'interpréter les règles. Ces valeurs sont mobilisées différemment selon les segments professionnels et les lieux d'exercice de l'activité médicale. Le médecin en

hôpital public est sans doute plus sensible à la qualité de service public que le médecin en exercice libéral intégral ou en clinique privée à but lucratif. Les médecins, comme les autres corporations, ne forment pas un tout uniforme. Mais, à un certain niveau d'identité, celui de la profession, une forme d'unité (qui ne suppose pas une identité des représentations) peut s'organiser autour de valeurs distinctives, dans lesquelles peut se reconnaître chaque médecin, même si le poids donné à chacune de ces valeurs peut varier d'un médecin à l'autre. On peut, dans cette perspective, recenser trois types de valeurs propres qui soutiennent la capacité de jugement éthique des médecins : les valeurs hippocratiques, les valeurs de la Santé et les valeurs libérales (Batifoulier, 2009).

Nous présentons ces trois valeurs propres avant d'étudier l'impact de la convention industrielle sur celles-ci.

- Les valeurs hippocratiques :

Ces valeurs expriment le souci du médecin – en tant que membre d'un collectif professionnel – pour le patient. L'hippocratisme est au fondement du paternalisme médical et du *care* exercé par le médecin. Ces valeurs relèvent des quatre principes fondateurs de la médecine énumérés par Gillon (1994) : respect de l'autonomie du patient, non malveillance, bienveillance, justice[131]. Ces valeurs ont pour objectif d'organiser la socialisation des membres de la profession (Gadreau, 2009). La profession doit être tournée vers l'intérêt du patient singulier. Les valeurs hippocratiques, associées aux valeurs humanistes, participent des institutions de la confiance qui assurent la société que le médecin donne de son temps et de sa personne. De là naissent les avantages sociaux de la profession : autorégulation et autocontrôle.

- Les valeurs de la santé :

Le médecin porte la responsabilité de ce bien précieux qu'est la santé. Cette fois-ci, au-delà du cas singulier, la santé est une valeur en soi qui doit être protégé par l'organisation sociale. Plusieurs traditions théoriques s'inscrivent dans cette approche : la santé est un bien premier naturel (Rawls, 1997 [1971]), une modalité propre du bien vivre (Ricoeur, 1996), ou, comme nous l'avons défendu au chapitre 2, la santé est un besoin et non une préférence. Ces différentes approches justifient que le médecin ait pour rôle de promouvoir la santé en toute généralité comme les avocats, en tant que profession, défendent l'idée de Justice (Bessis, 2008). Le médecin n'est pas donc uniquement responsable de l'amélioration du bien-être

[131] Nous avons développé ce point au chapitre 2.

particulier mais il est, par ailleurs, dépositaire de l'idée de Santé. La défense de la valeur Santé octroie alors au médecin une légitimité particulière d'où provient son pouvoir politique.

- Les valeurs libérales :

D'après ces valeurs, la seule façon de défendre la mission hippocratique et l'objectif de Santé est l'organisation libérale. Ces valeurs sont plus récentes et correspondent, en France, à la construction de la médecine libérale face à l'État Social. La Charte de la médecine libérale de 1927 peut être considérée comme le cœur du projet libéral de la médecine française. Les principes d'entente directe, de paiement direct et de liberté d'installation sont conçus comme les conditions garantissant des soins de qualité et une morale exigeante.

Ce libéralisme n'est pas un libéralisme économique : la profession doit se trouver à égale distance de la règle bureaucratique (État) et de la règle marchande (marché). La concurrence est prohibée et la question des revenus est largement relativisée. D'ailleurs, le libéralisme médical se construit autour du devoir de générosité du médecin. Les actes gratuits et la variabilité des prix sont ainsi le symbole de la charité libérale selon laquelle le seul juge du « don de soin » doit être le médecin lui-même et non l'administration. Or, la socialisation prive le médecin de cette capacité de faire justice[132].

Au total, l'éthique médicale se nourrit d'une pluralité de représentations sur ce qu'il convient de faire. Ces principes normatifs ne sont pas parfaitement définis et fournis clef en main : ils sont susceptibles d'évolution dans le temps et l'espace. La politique publique est l'une des sources les plus importantes permettant d'expliquer l'évolution de l'éthique médicale. De nombreux travaux ont déjà montré comment à force de réformes à dominante marchande, l'éthique médicale s'est peu à peu déplacée vers la légitimation de comportements marchands dans la relation médicale (Encadré 8.3.).

[132] Notons une dernière fois que cette liberté n'est pas gratuite. Contrairement à ce que défendent les valeurs libérales de la médecine, il est possible de soigner en dehors du libéralisme. La légitimité de ces valeurs provient du fait qu'elles sont le contre don du respect des valeurs hippocratiques et de santé. Tel est le prix de la liberté politique.

Encadré 8.3. La coloration marchande de l'éthique médicale

La politique publique n'est pas sans effet sur les comportements et les représentations assurant la coordination dans le système de santé. Au moins depuis 1980, l'orientation marchande de la politique publique de santé modifie la conception de ce qu'il convient de faire pour la profession médicale (Batifoulier et al. 2007) : ouverture du secteur tarifaire à dépassements d'honoraires, multiplication des incitations financières à destination du médecin, contribution directe grandissante du patient au financement des soins, etc. Toutes ces évolutions contribuent à modifier la conception du soin. Les soins tendent à devenir un bien comme un autre sur un marché comme un autre.

Cependant, parler de marchandisation des soins dans le secteur des soins serait abusif. En effet, l'influence de la politique publique sur les représentations n'est jamais totale. Les principes marchands diffusés par la politique publique ne viennent pas sur un terrain vierge de valeurs. Ils doivent se conjuguer avec les valeurs médicales traditionnelles. La modification des règles ne transforment pas le médecin en marchand pur. La pluralité des valeurs demeure, c'est pourquoi la littérature conventionnaliste en santé parle de « coloration marchande » de l'éthique médicale (Batifoulier et Biencourt, 2005, Batifoulier et Gadreau, 2007). Mais, à l'intérieur de cette pluralité, la politique publique a rendu saillant le cadre marchand. Le service de soin devient alors davantage marchand à deux niveaux : (i) au niveau des règles et des instruments de gouvernance et (ii) au niveau des représentations et des comportements des acteurs – qui appliquent ou résistent à la politique publique.

Au niveau des règles et des instruments de gouvernance, certains travaux ont montré que le droit tend à être colonisé par le vocabulaire économique. Abecassis et Domin (2008) ont ainsi analysé l'évolution du vocabulaire utilisé dans la rédaction des conventions médicales nationales et montré la présence croissante des préoccupations financières. La convention de 2005 instituant le statut de médecin traitant est celle qui offre le terrain le plus favorable aux notions économiques et comptables. La maîtrise médicalisée des dépenses au centre de l'analyse écarte par ailleurs toutes notions de sanctions ou de contraintes pour les médecins. Plus généralement, ce formatage économiste du droit place au cœur de l'analyse l'incitation financière comme l'instrument privilégié de régulation (Allouache et Vacarie, 2008).

La diffusion des principes marchands par les textes transite également par les représentations. Contrairement à l'individu Pavlovien de la théorie économique *mainstream*, le stimulus n'engendre pas de réponse apprise. Avant de modifier le comportement, l'évolution des règles modifie les représentations. L'incitation modifie les conceptions de ce qu'il convient de faire… dans un sens marchand. Le libéralisme économique de la politique publique trouve néanmoins un terrain favorable avec le libéralisme politique traditionnel de la médecine libérale. La marchandisation de la relation médicale est alors légitime et il est possible d'observer empiriquement l'envolée de ce type de comportement.

Deux grandes évolutions sur le plan des comportements peuvent être distinguées. D'une part, l'orientation marchande de la politique publique implique une banalisation des actes chers. Les dépassements d'honoraires deviennent de plus en plus importants et de plus en plus courants – même chez les médecins de secteur 1 (Aballea et al. 2008). Le tarif opposable est en recul depuis 1980 et rien ne semble arrêter la croissance de l'écart entre ce tarif et le prix effectivement payé par le patient. D'autre part, la coloration marchande de l'éthique médicale se repère par le mouvement de dégratuitisation (Batifoulier et al. 2009). La littérature a ainsi mis en évidence l'érosion de la « part gratuite » des soins en médecine libérale (Batifoulier et Ventelou, 2003). Les actes gratuits sont de moins en moins courants (Lievaut, 2010) et cette tendance est complétée par le phénomène nouveau de refus de soin aux titulaires du dispositif CMUC (Desprès et al., 2009).

2.1.2. Ce que fait la convention industrielle à l'éthique médicale

À la suite de ces travaux, nous souhaitons renouveler l'analyse des politiques publiques en questionnant l'impact du second grand tournant de la médecine libérale sur les valeurs propres de la médecine que nous venons d'identifier. En substituant la convention industrielle de qualité des soins à la convention inspirée/domestique, la nouvelle stratégie du régulateur participe au renouvellement de l'éthique médicale.

- Les valeurs hippocratiques : un supplément d'âme dangereux

La transformation la plus radicale qu'apporte la convention industrielle à l'éthique médicale porte sur les valeurs hippocratiques. L'automatisme de la décision consécutif à l'application des standards anémie la relation personnelle entre le médecin et son patient. Dans la nouvelle convention de qualité des soins, l'hippocratisme n'est plus primordial : le patient ne doit pas attendre du médecin une qualité morale dans le cadre du colloque singulier car l'application de la norme suffit à obtenir un soin de qualité. Le changement de convention appelle à un renouvellement de l'éthique médicale. Il n'est plus attendu que le médecin ait une compétence morale attachée à sa personne : évaluer le bien et le mal. La compétence doit avant tout être technique (Jaunait, 2004).

L'hippocratisme peut même être néfaste : le médecin qui décide de ne pas respecter le standard dans le souci du patient peut être considéré comme un « danger de santé publique ». En voulant être bienveillant, le médecin s'écarte de la norme, ce qui cause au niveau agrégé une baisse de la qualité des soins et une hausse des coûts[133].

Avec la convention industrielle, la bienveillance ne disparaît pas pour autant : plutôt que dans la personne du médecin, celle-ci est dans le protocole. La confiance ne siège plus dans la personnalité du médecin mais dans l'« im-personnalité » de l'agence indépendante. La médecine de protocole remet en cause le paternalisme médical pour reléguer au second plan l'autonomie au médecin. La confiance du patient ne doit plus résider dans le seul professionnel. Elle doit être rabattue vers la norme chiffrée et les processus de certification. C'est le paternalisme de l'expert qui est valorisé[134].

- Les valeurs de la Santé : le déclassement du médecin

La convention industrielle de qualité des soins ne remet pas en cause l'idée de Santé. En ce sens, l'atteinte aux valeurs propres est moins radicale dans ce cas-là que dans celui des valeurs hippocratiques. Il est néanmoins possible de repérer deux changements importants sur la valeur Santé ayant pour conséquence un déclassement de la profession médicale.

[133] En forçant le trait, on pourrait dire que le second grand tournant de la médecine libérale organise alors un détachement du médecin vis-à-vis de la plainte de son patient contraire à la logique antérieure de *care*. S'intéresser au patient c'est prendre le risque de voir son jugement brouillé par des considérations affectives qui ne font pas le poids face aux résultats de la médecine fondée sur les preuves. La règle uniforme permet ainsi de ne pas faire reposer la décision médicale sur un médecin qui, au mieux, est mal informé des dernières avancées scientifiques ou qui, au pire, est susceptible d'opportunisme malveillant.

[134] Or, les experts ne sont pas organisés en profession et les règles déontologiques sont très souples.

D'une part, la nouvelle convention incite à relativiser grandement la figure du médecin comme dépositaire de la santé. Le médecin n'est plus le gardien de ce bien précieux qu'est la santé. La définition et l'application de standards chiffrés de production est une source de désingularisation du soin tant du côté du patient que de celui du médecin. Du point de vue du patient, la maladie devient un objet détachable, désencastré du malade. Du point de vue du médecin, l'expérience et le jugement ne sont d'aucune utilité dans la mesure où le standard dit quoi faire en toutes circonstances. Ce n'est donc plus un médecin singulier qui soigne un malade singulier mais un standard qui soigne une maladie. Historiquement dépositaire de la santé, le médecin n'est plus qu'un rouage dans la division du travail. Dans le cadre de la convention industrielle, le médecin n'est plus le centre de décision. La santé est toujours une valeur en soi mais elle est promue avant tout par les agences indépendantes productrices de normes sur la base de la médecine fondée sur les preuves. Les patients sont invités à faire confiance aux règles et procédures organisant la standardisation des soins – l'agence indépendante étant au sommet de la hiérarchie – et non au médecin. Hier savant charismatique, le médecin est dans la nouvelle convention victime d'un déclassement.

D'autre part, la nouvelle convention met en évidence le conflit entre plusieurs conceptions de la santé. Alors que l'on pourrait penser qu'il n'y a pas de conflit entre la santé au niveau particulier (micro) et la santé au niveau général (macro), l'émergence de la convention industrielle vient contredire cette hypothèse. En effet, comme les avocats peuvent avoir à informer leurs clients d'une contradiction possible entre l'idée (générale) de Justice et leurs revendications (particulières) de justiciables (Bessis et Favereau, 2010), les médecins font face au même type de problème lorsque l'intérêt ou les désirs des patients s'opposent à l'idée de Santé. Ainsi, si la convention industrielle enjoint à soigner le patient moyen pour minimiser la variabilité des pratiques et l'efficacité des traitements, dans certains cas (les queues de distribution), l'intérêt du patient particulier est d'être soigné différemment du patient moyen. Chaque cas porte en lui un dilemme entre général et particulier[135].

L'application systématique des protocoles peut alors produire des situations où le médecin a l'impression de mal faire son travail parce qu'il y a un conflit entre les deux niveaux de généralité de la Santé. La conception du métier change. L'éthique médicale tend à

[135] Sans même recourir aux problèmes posés par la médecine fondée sur les preuves, la question de la vaccination illustre également le dilemme entre général et particulier. Au niveau particulier, il est possible que le vaccin ne soit pas dans l'intérêt médical du patient mais, il est dans l'intérêt général que le maximum de patients soient vaccinés pour que le traitement soit efficace. Notons néanmoins que le continuum entre intérêt général et intérêt particulier semble bien plus restreint dans le cas de la Santé que dans celui de la Justice.

devenir une éthique bureaucratique ou administrative : peu importe les cas, il faut un détachement vis à vis du patient afin de « bien cocher les cases ».

- Les valeurs libérales : d'une catégorie de liberté à l'autre.

À première vue, la convention industrielle a le mérite de lutter contre le libéralisme médical et constitue une attaque à la toute puissance de la profession qui entend exclure l'État social de l'organisation du « bien soigner ». La médecine à identité libérale la vit comme une intrusion insupportable dans le « colloque singulier » et comme un « flicage » du médecin. Cependant, si les libertés liées à la pratique médicale sont effectivement assez sévèrement remises en cause, les libertés tarifaires et d'installations sont quant à elles renforcées.

Comme nous l'avons rappelé à plusieurs reprises, la nouvelle convention de qualité repose sur la possibilité de dissoudre la qualité des soins dans les quantités. Si les valeurs traditionnelles de la profession médicale pouvaient assez aisément s'accommoder du libéralisme marchand développé par la politique publique, il semble que la convention de qualité industrielle des soins, en remettant plus fondamentalement en cause l'autonomie du médecin, menace son statut même de professionnel. La prolifération des standards de production, dénie au médecin sa capacité à exercer avec prudence – au sens de la convention de la pratique prudentielle propre à un professionnel (Cf. plus haut). Il n'est plus un professionnel mais un rouage interchangeable dans la division du travail propre au système de soin. L'identité libérale politique s'effrite au profit des règles bureaucratiques. Le médecin perd sa liberté thérapeutique et sa liberté de prescription.

C'est dans cette perspective que la Rémunération sur objectif de santé publique incite le médecin à prescrire le dépistage du cancer du sein à ses patientes alors que l'efficacité des mammographies fait l'objet de controverses importantes dans la profession[136]. L'autonomie autrefois pierre angulaire de l'organisation professionnelle est aujourd'hui une menace. La convention industrielle orchestre la dépossession des libertés thérapeutique et de prescription. C'est du point de vue du législateur son principal intérêt. Mais ce n'est plus au médecin de dire s'il faut faire une mammographie mais à la norme produite par l'agence indépendante. Seule la règle bureaucratique est censée produire la qualité des soins – et non le jugement autonome du médecin.

[136] Cf. Chapitre 5.

2.1.3. Un réagencement des valeurs

Cette reprise en main du processus thérapeutique par le régulateur engendre un résultat original qu'il convient d'analyser au regard de la situation antérieure. Dans la convention inspirée / domestique, le don de liberté politique peut s'interpréter comme le contre don du respect des valeurs hippocratiques. De ce processus social doivent découler des soins de qualité. Dès lors, toute remise en cause du libéralisme médical se concevait comme une menace pour la qualité de soins – la socialisation ou la maîtrise des dépenses se faisant alors nécessairement contre le patient. La convention industrielle renverse la logique de la critique : la critique des médecins face à l'immixtion de l'État dans le « colloque singulier » devient plus difficile dans la mesure où ces deux institutions se prévalent de la qualité des soins. Il est plus facile de défendre la qualité des soins contre la maîtrise des dépenses que contre… la qualité des soins elle-même (convention industrielle). Cette évolution de la régulation de la médecine libérale désarme la critique de la profession face à l'État[137].

Cependant, il ne faut pas croire que la profession médicale s'est trouvée démunie face au second grand tournant de la médecine libérale. Bien au contraire, il faut y voir un arbitrage entre des libertés qui ne sont pas équivalentes. En cédant du terrain sur la liberté thérapeutique et la liberté de prescription, la profession médicale a renforcé ses libertés tarifaires et d'installation.

Du point de vue de la liberté d'installation, il est possible de voir dans le développement de la télémédecine une illustration du second grand tournant de la médecine libérale (Da Silva et Rauly, 2014, Rauly, 2013). La standardisation est un moyen de satisfaire et de pérenniser la liberté d'installation. Plutôt que de légiférer sur la liberté d'installation chère à la profession médicale depuis la Charte de 1927, la puissance publique encourage la télémédecine[138] dans la mesure où elle constitue un levier de la normalisation des pratiques. En effet, si la télémédecine permet au médecin de consulter à distance – donc de s'installer où il le désire, indépendamment de la démographie médicale –, cet outil technologique est aussi un outil de

[137] Au moins dans un premier temps, car, dans un second temps, la critique peut et doit porter sur la possibilité de dissoudre la qualité dans les quantités.

[138] La télémédecine utilise les technologies de l'information et de la communication pour mettre en rapport un médecin et un patient. Ainsi, si le médecin ne vient plus au patient, c'est le patient qui ira virtuellement au médecin grâce à la téléconsultation. Cette stratégie revient à missionner les industriels privés pour construire les équipements nécessaires à l'accès au médecin (cabines de téléconsultation ou mallettes de télémédecine), afin que l'ordinateur médical à la maison remplace le médecin de proximité. Le patient, sa famille ou un aidant sont intégrés dans la filière de soin. Ils collaborent avec le médecin qui établit et transmet les données, et participent au développement d'un marché lucratif pour les plateformes privées de télésurveillance et télétransmission.

contrôle puisqu'il permet de constituer des bases de données sur l'activité médicale. Il ne reste alors plus qu'à prescrire le travail à partir de la médecine fondée sur les preuves et à le vérifier informatiquement pour réguler les pratiques.

Du point de vue de la liberté tarifaire, il n'est pas nécessaire d'insister : les années 1980 ont vu la revanche de l'entente directe et les années 2000 l'abandon de toute politique de prix[139]. L'acceptation (relative) par la profession médicale de la convention industrielle révèle ainsi une sorte de troc institutionnel entre les libertés tarifaires et de prescription d'un côté et les libertés d'installation et tarifaire de l'autre.

2.2. L'industrialisation de la médecine, levier à la marchandisation des soins

Dans cette section nous repartons de l'analyse du paiement à la performance médicale pour montrer en quoi il s'inscrit dans le cadre du néolibéralisme sanitaire. Si la double logique industrielle et marchande du paiement à la performance médical (2.2.1.) peut contrarier les motivations médicales (2.2.2.), l'usage intensif des statistiques comme outil de preuve et outil de gouvernement (2.2.3.) permet de dissimuler la propagation d'une gouvernementalité néolibérale (2.2.4). C'est sous ces conditions que la convention de qualité de soins industrielle peut propager le néolibéralisme médical (2.2.5.) et poursuivre le programme de privatisation du secteur sanitaire (2.2.6.).

2.2.1. Le paiement à la performance médicale : entre logique industrielle et logique marchande

Le dispositif de paiement à la performance (P4P) médicale est emblématique des 30 dernières années de politique publique. Il a des airs de famille avec le salaire à la performance de la période fordiste. Le paiement au résultat est proche du paiement à la tâche ou à la pièce. En ce sens, il est emblématique d'une approche industrielle de la production, en l'occurrence ici, la production de soins. Mais, il intègre aussi une logique marchande dans la mesure où le travail doit être motivé par des incitations financières.

Ce type de mécanisme incitatif développe à la fois une logique industrielle (« la qualité est résumée par des quantités mesurables ») et une logique marchande (« le médecin est motivé par son intérêt lucratif, qu'il faut susciter »). Ainsi, le médecin devient un salarié « ordinaire » face à la logique de la marchandisation. Contrairement aux cadres dans les

[139] Cf. Chapitre 6 et Encadré 8.1.

entreprises privées, il n'a pas à craindre un licenciement pour motif personnel (Seignour et al., 2007). Néanmoins, ses pratiques sont pilotées grâce au couple rémunération/évaluation : la rémunération vient encourager ou sanctionner l'activité professionnelle dont l'évaluation est nécessairement standardisée.

Le P4P suppose en effet que les médecins ne sont pas motivés au travail et que la motivation à l'effort doit être financière (Wynia, 2009). Il fait ainsi du médecin un *homo economicus* comme un autre et, à ce titre, il participe du libéralisme marchand. D'ailleurs, il est intéressant de noter que le P4P est une confirmation implicite du paiement à l'acte car il repose sur une logique similaire. Dans le cadre du paiement à l'acte, il est justifié de faire ce pourquoi l'on est payé (l'acte) et il est justifié de ne pas faire ce pour quoi l'on n'est pas payé (l'absence d'acte). Ce principe d'action devient dans le cas du P4P : il est justifié de faire ce pour quoi on est surpayé (l'indicateur) et il est justifié de ne pas faire ce pour quoi on n'est pas surpayé (l'absence d'indicateur). Autrement dit, en dehors de la rémunération du travail prescrit, il n'y a pas de travail. Or, l'incomplétude contractuelle, que l'on pourrait qualifier de radicale dans le cas des services de soins, implique qu'il est impossible de rémunérer toutes les tâches – parce que l'on ne les connait pas à l'avance. Le P4P approfondit alors la logique marchande du paiement à l'acte selon laquelle ce qui n'est pas explicitement rémunéré n'est pas à réaliser.

La logique marchande portée par le P4P n'est néanmoins pas tout à fait la même qu'auparavant. Désormais, le versement de la rémunération (la prime) se fait sous condition de la modification des pratiques. Le paiement à l'acte diffère donc du P4P sur la conditionnalité de la rémunération : la rémunération de l'acte est inconditionnelle et laisse au médecin le soin de décider ce qu'il convient de faire – même le pire – alors que la rémunération à la performance est conditionnelle à l'indicateur et contraint le médecin réceptif à l'opportunité de la prime.

C'est ici qu'opère la stratégie du Nirvana analysée dans le premier chapitre. La concurrence étant proscrite dans le champ médical, c'est l'incitation qui doit conduire le médecin à adopter un comportement « comme si »[140] : comme si l'on était sur un marché de concurrence pure et parfaite. La perte d'autonomie du médecin consécutive à la conditionnalité de la prime peut alors se voir sous son jour positif : l'amélioration du service rendu au patient. De ce point de vue, il y a un approfondissement de la logique marchande car

[140] En référence au fameux « as if » de Milton Friedman.

désormais il n'est plus question de financer le médecin « pour rien ». Contrairement, par exemple, à l'ouverture du secteur à dépassement d'honoraires qui ne s'est accompagnée d'aucune contrepartie, le P4P relève bien plus de la logique marchande puisqu'à un prix correspond un service.

Avec l'introduction du P4P il y a donc continuité et rupture avec les politiques publiques antérieures. Il y a continuité parce que le P4P s'engouffre dans la brèche des valeurs marchandes, à la suite de l'abandon de la politique de prix. Il y a rupture parce que désormais la rémunération est (a) conditionnelle à une pratique (b) fondée sur une convention de qualité des soins industrielle. Le P4P est un instrument de régulation par le prix et par les pratiques. C'est un pont nécessaire dans l'évolution de la régulation de la médecine libérale.

En dehors de la politique publique, le P4P peut s'appuyer sur la légitimité de l'évolution historique de la médecine : la médecine fondée sur les preuves. C'est ainsi que l'on peut comprendre le « méli-mélo » de motivations (Bloy et Rigal, 2012) à l'origine de l'acception et du rejet du CAPI et du ROSP en France. Etant fondée sur une logique plurielle (marchande et industrielle), le P4P multiplie les justifications et les critiques dans une profession médicale aujourd'hui très segmentée.

2.2.2. Un interprétation du crowding ou effect médical

À partir de cette théorie conventionnaliste de l'évolution de la politique publique, il est possible d'incorporer à l'analyse la question de la motivation posée par les psychologues sociaux et de donner une interprétation du *crowding out effect* (COE)[141]. En effet, l'Économie des conventions comme la Théorie de l'autodétermination (définie au chapitre 3) partagent le souci de mettre au centre de leurs théories la dimension interprétative du comportement. Rappelons à ce titre que, d'après Théorie de l'autodétermination, le COE repose sur deux mécanismes d'interprétation : le *perceived competence* – l'incitation est une preuve du manque de confiance du régulateur dans les capacités de l'incité – ou *perceived locus of causality* – l'incitation remet en cause l'autonomie de l'incité puisqu'il n'est plus l'origine de l'action et passe pour quelqu'un d'intéressé. Dans les deux cas, un besoin psychologique fondamental (compétence ou autonomie) est remis en cause par l'incitation. Ce processus n'a pourtant rien de mécanique. Les individus interprètent la motivation extrinsèque (ME) à partir de la situation particulière à laquelle ils font face.

[141] Cf. Chapitre 3. Le COE se définit comme l'éviction des motivations intrinsèques lorsqu'un régulateur introduit des motivations extrinsèques.

Le problème le plus important de cette théorie est que l'on ne sait pas en fonction de quoi se fait l'interprétation. La ME est interprétée en situation mais rien n'est dit sur ces situations et sur les ressources pour l'interprétation. Or, c'est justement ce que propose l'Économie des conventions : les individus interprètent les situations en mobilisant une pluralité de conventions possibles[142]. Dans le cas de l'introduction du P4P dans la médecine libérale, deux conflits de représentations majeurs entrainant potentiellement un COE sont possibles selon que l'on s'intéresse à l'une au l'autre des valeurs portées par ce type de dispositif :

- Contre les valeurs marchandes :

Du point de vue des valeurs libérales marchandes, si l'incitation en appelle à l'intérêt lucratif, elle est néanmoins conditionnelle à une pratique particulière. La subordination de la rémunération à une forme prescrite d'activité n'est rien d'autre qu'une subordination du médecin aux recommandations de l'assureur public. Il est évident que de ce point de vue le libéralisme médical n'est pas totalement en phase avec le libéralisme marchand. Plus classiquement, le libéralisme marchand n'a aucune raison de s'accommoder des valeurs hippocratiques. Le P4P, à travers l'approfondissement de la logique marchande, impose une réduction de l'autonomie du médecin. La mise à mal du besoin psychologique d'autonomie peut alors être à l'origine d'un *crowding out effect*.

- Contre la convention de qualité des soins industrielle :

Le P4P porte en lui une atteinte plus fondamentale du besoin d'autonomie par la substitution de la convention inspirée/domestique de qualité du travail médical par la convention industrielle. En plus de la conditionnalité de la rémunération, le P4P prescrit une règle de comportement hétéronome sur laquelle le médecin n'a aucune prise – sous peine de non rémunération. À aucun moment le médecin ne participe de la construction du standard. Non seulement l'autonomie du médecin est déniée mais cet état de fait repose sur un fort scepticisme quant à sa compétence. Le médecin n'est plus le dépositaire de la santé. Si la puissance publique prescrit ce qu'il convient de faire, c'est bien que le médecin (i) ne veut pas le faire et/ou (ii) ne sait pas le faire. La critique opposée à la convention inspirée / domestique censée justifier la convention industrielle repose sur le déficit de compétences et la trop forte

[142] Le COE peut alors se comprendre comme un conflit de représentations. Si dans une situation donnée un individu adopte la convention A et que la ME véhicule la convention B, alors le résultat de la ME va dépendre de la complémentarité ou de la substituabilité entre A et B. Si A et B sont complémentaires, il y a effet d'entrainement des motivations. Si A et B sont substituables, il y a éviction des motivations intrinsèques (MI) par les ME (c'est le COE).

autonomie du médecin. La convention industrielle implique donc logiquement la corrosion des besoins psychologiques fondamentaux (d'autonomie et de compétences) et, probablement, un effet d'éviction des motivations intrinsèques médicales.

Les règles de bonnes pratiques comme la rémunération à la performance sont des informations qui sont interprétées par leurs destinataires. L'affichage de l'industrialisation de la médecine donne à penser au médecin que son flair, son intuition, son expérience, son savoir faire sont devenus secondaires et peut induire une dé-motivation à mobiliser son art face à un patient, d'autant plus s'il s'agit d'un patient complexe. C'est une première source d'éviction.

La transformation du médecin en chasseur de primes par le développement des récompenses financières peut induire un paradoxal désintéressement pour la prime et surtout pour l'acte qui lui est subordonné. Donner un prix à l'acte médical et au fait de faire simplement son travail lui enlève de la valeur et rend rebutant le respect de la « bonne pratique ». En gratifiant la bonne pratique, on peut aussi la dé-gratifier. C'est une seconde source d'éviction.

Ces effets d'éviction potentiels sont imputables à la façon dont les professionnels interprètent les dispositifs et aux ressources conventionnelles qu'ils mobilisent pour cette interprétation. Le flair du médecin n'est opposé aux bonnes pratiques industrielles que si elles sont vécues comme une attaque à ce qui fait la singularité du métier de soignant. La récompense financière produit des effets contreproductifs que si le médecin lui associe le développement d'une vision péjorative de son métier. Or, le climat marchand qui anime les réformes de la santé oriente les représentations et modifie la rationalité médicale. L'industrialisation et le développement de la logique de la performance sont des éléments de la diffusion de ce climat marchand.

2.2.3. La statistique : l'instrument de gouvernement derrière l'outil de preuve

Pour comprendre la diffusion des valeurs marchandes dans les représentations, on peut s'intéresser au rôle particulier que joue la statistique dans la convention de qualité des soins industrielle. La réforme de la santé, quand elle s'inscrit dans le registre marchand, veut faire de la santé un bien comme un autre et légitimer les comportements marchands dans un secteur qui en était dépourvu. Afin de contourner les résistances à cette nouvelle forme de légitimité dans le système de santé, la nouvelle convention s'appuie sur l'une des propriétés importantes du chiffre et des statistiques – la statistique est un instrument de preuve

Dans la plupart des disciplines de sciences sociales, la littérature fait état d'une propension particulièrement développée de la période contemporaine à utiliser les nombres afin de décrire la réalité et d'orienter les comportements. Cette inclinaison à mettre le monde en chiffres plutôt qu'en mots serait d'ailleurs surdéveloppée tant est si bien que de nombreux auteurs s'accordent à y voir une pathologie sociale parfois qualifiée de « quantophrénie » (de Gaulejac, 2009).

Les problèmes liés à la surutilisation des nombres sont alors à rechercher dans la double propriété de l'argument statistique (Desrosières, 2008) : c'est un outil de preuve et un outil de gouvernement. Il est assez clair que l'argument statistique est un outil de preuve. L'usage des chiffres permet de développer les connaissances sur le monde. Compter, mesurer, comparer, etc. permet de se doter d'une représentation du monde[143].

L'intérêt de l'argument statistique en tant qu'instrument de preuve est qu'il objective le débat. Les français vivent en moyenne plus longtemps que les américains. Dans une certaine mesure, le chiffre clos la discussion : 3 est plus grand que 2 qui est plus petit que 4. Evidemment, le débat ne peut être clos qui si le problème d'identification est résolu. Comme l'explique Alain Desrosières (2008), « quantifier, c'est convenir puis mesurer ». Ce n'est que parce qu'il y a accord sur la chose mesurée que le résultat statistique détient le pouvoir de mettre tout le monde d'accord. On peut alors se demander si vivre longtemps c'est bien vivre ou si le calcul d'une moyenne ne cache pas de fortes inégalités sociales. L'usage éclairé de la statistique ne peut pas se passer d'une discussion permanente sur la définition de la chose mesurée. Quantifier, ce n'est pas que mesurer.

Or, dans la perspective épistémologique réaliste qui domine aujourd'hui, la réalité préexiste à la chose mesurée et en effet, 1 est plus petit que 2 sans qu'il ne soit possible d'apporter une critique sur l'existence et/ou la pertinence de la chose mesurée. Dans le cas du ROSP, si l'on considère que le taux de réalisation global n'est pas discutable, alors les médecins peuvent être classés du meilleur au plus mauvais sans autre forme de procès. Le chiffre est « sacralisé » (Ogien, 2013) et n'appelle pas à la critique. Tout au mieux les questionnements portent sur la précision de l'indicateur mais il n'est jamais question de

[143] Par exemple, le calcul de l'espérance de vie permet de donner une idée du temps séparant la naissance du décès des individus dans un territoire donné. Comparer dans le temps et dans l'espace les espérances de vie permet de trouver où les individus vivent le plus longtemps. Dire que sur le territoire « A » la population a en moyenne une espérance de vie supérieur de 8 ans à celle du territoire « B » permet de décrire une différence objective de temps de vie à travers l'espace. L'argument statistique permet de comparer et de hiérarchiser des états du monde.

remettre en cause la réalité de la chose mesurée. Plutôt que de questionner l'idée de performance médicale, la seule critique possible est d'améliorer les outils de mesure de cette performance.

Cette posture réaliste est d'autant plus dommageable qu'elle accélère la mutation du chiffre comme outil de preuve en instrument de gouvernement. En effet, puisque le chiffre n'est pas discutable, il porte en lui une légitimité imposant l'orientation les comportements. Dans cet esprit, les données chiffrées sont couramment utilisées par diverses institutions ou organisations afin de modifier les comportements. Dans l'entreprise par exemple, les tableaux de bord permettent de fournir une vision de l'activité qui oriente l'action. Pour la médecine libérale, un « bon » médecin est nécessairement un médecin ayant de bons indicateurs ROSP. Il s'agit de gouverner les comportements désormais pilotés par le nombre.

L'usage des statistiques dans un cadre réaliste cache alors une certaine conception de l'action juste qui la sous-tend. Alain Desrosières (2008) a ainsi montré que les méthodes statistiques sont loin d'avoir connues une évolution croissante linéaire propre à un hypothétique progrès empreint de neutralité scientifique. Au contraire, derrière les discours d'objectivité se cachent l'évolution de la conception de l'État et de ses rapports au marché. La statistique est alors un instrument de pouvoir qui permet au prince de voir ce qu'il veut voir.

L'auteur a distingué cinq formes d'État correspondant à cinq types de statistiques et, donc, à cinq conceptions différentes des relations entre État et marché. Ces différentes catégories ne sont pas exclusives les unes des autres et sont présentées dans la chronologie de leur apparition (Tableau 8.2.).

Tableau 8.2. L'État, le marché et les statistiques (Desrosières, 2009, p. 12)

	Façon de penser la société et l'économie	Mode d'action	Formes de statistiques
État ingénieur La production et les hommes (depuis le XVIIème siècle)	Institutions hiérarchisée et organisée rationnellement. France, de Colbert à De Gaulle. URSS.	Optimisation sous contraintes. Réduction des coûts. Planification centralisée. Technocratie. Grands travaux. Vision à long terme.	Démographie. Production en quantité physique. Tableaux d'échanges interindustriels. Balances matières (URSS).
État libéral Le commerce et les prix (depuis le XVIIIème siècle)	Physiocratie. Un grand marché. Concurrence libre.	Lutte contre les corporatismes. Libre échangisme. Loi anti-trust protégeant la concurrence.	Statistiques favorisant la transparence des marchés (cas de l'agriculture américaine). Mesures éventuelles de positions dominantes. Parts de marché.
État Providence Le travail salarié et sa protection (depuis la fin du XIXème siècle)	Le marché du travail n'est pas un marché comme un autre. Il doit être protégé.	Lois sur la durée du travail, sur les accidents, le chômage, les retraites. Systèmes d'assurances obligatoires garantissant des droits sociaux.	Statistiques du travail. Salaire, emploi, chômage. Enquêtes par sondage sur les budgets ouvriers. Indices de prix à la consommation. Mesure des inégalités sociales.
État Keynésien La demande globale et ses composantes (depuis les années 1940)	Le marché ne peut fonctionner tout seul sans engendre des crises. Il doit être régulé au niveau global.	Surveillance et pilotage du gap éventuel entre offre et demande, via les politiques monétaires et budgétaires.	Comptabilité nationale. Analyse de la conjoncture. Budget économiques. Modèles macroéconomiques.
État néo-libéral Polycentrisme, incitations, benchmarking (depuis les années 1990)	Un grand marché. Concurrence libre et non faussée. Financiarisation. Démultiplication des centres de décision sous forme de réseaux.	Passage des droits aux incitations. Exemples : bonus-malus, marché des droits à polluer. Transformation des administrations en agences. Contractualisation. Coordination par émulation. Exemple de la méthode ouverte de coordination européenne.	Objectivation d'espaces d'équivalence nouveaux. Objectivation des qualités contractuelles des statistiques. Construction et usage d'indicateurs pour évaluer et classer les performances. Palmarès. Le benchmarking complète ou remplace directives et règlements. Débats sur la quantification du PIB.

À partir du XVIIème siècle, l'État ingénieur est caractérisé par la volonté d'intervention directe et ses statistiques sont tournées vers la planification. Il s'intéresse à la démographie et aux flux de produit. L'État libéral qui nait au XVIIIème siècle théorise à l'inverse une réduction drastique de son intervention dans le but de « libérer » le marché seul producteur de libertés publiques et de richesses. Dans ces conditions, l'outillage statistique d'État n'est d'aucune importance si ce n'est d'étudier les écarts entre la réalité des marchés et les potentialités théoriques. La connaissance du prix est suffisante car il contient toute l'information. À la fin du XIXème siècle, naît l'État Providence. La marchandisation s'étendant désormais au travail lui-même, le but de l'État Providence est de protéger les travailleurs salariés des conséquences de cette évolution du capitalisme : chômage, accident du travail, famille, santé, vieillesse, etc. Les statistiques ont alors pour but de calculer les taux d'emploi et des indices de prix à la consommation. L'État keynésien, après la seconde guerre mondiale, élargit les responsabilités de l'État Providence au pilotage macroéconomique de la nation. La comptabilité nationale devient centrale et les calculs de la période antérieure s'étendent, au-delà des travailleurs salariés, à toute la population. La dernière forme de l'État est, à partir des années 1990, l'État néolibéral. Plutôt que la macroéconomie, l'intérêt du nouveau pouvoir est porté sur les dynamiques microéconomiques. La thématique des incitations devient centrale avec le but de dégager les meilleures pratiques. L'économétrie cherche à partir de ce canevas à isoler les effets propres de la politique publique dans le but d'améliorer des indicateurs de performance désormais critère d'évaluation primordial de l'action de l'État[144].

De cette analyse nous pouvons retenir que le second tournant de la médecine libérale utilise les statistiques propres à l'État néolibéral. Dans les deux cas, des indicateurs de performance ont pour objectif d'inciter à une modification des comportements au niveau microéconomique pour atteindre des valeurs cibles – censées correspondre à des « meilleures pratiques ». L'appareillage statistique qui entoure la régulation de la médecine libérale, s'il est indiscutablement un outil de preuve, est également un instrument de gouvernement au service d'une conception néolibérale de l'État et de l'action publique.

[144] Les dates de naissances de ces différentes formes de l'État chez Alain Desrosières peuvent être rapprochées des différents régimes d'accumulation du capital décrits par la théorie de la régulation (Cf. chapitre 6). L'État libéral, l'État keynésien et l'État néolibéral semblent en effet correspondre respectivement au régime concurrentiel, fordiste, et postfordiste. On peut considérer l'État ingénieur comme une prémices du régime d'accumulation concurrentiel et l'État Providence comme une charnière entre le régime concurrentiel et le régime keynésien. Notons néanmoins que si les régimes d'accumulation naissent et meurent (ils sont bornés chronologiquement), les différentes conceptions de l'État chez Desrosières ne sont pas exclusives les unes aux autres (elles peuvent coexister).

2.2.4. De l'État néolibéral à la gouvernementalité néolibérale

Les statistiques ne sont donc pas qu'un instrument de preuve : intentionnellement ou non, elles relèvent d'une conception particulière du monde ayant une dimension conventionnelle – c'est-à-dire liée à des représentations qui pourraient être différentes mais qui assurent, et c'est là l'essentiel, la coordination entre les acteurs. L'usage des statistiques qui imprègne le second grand tournant de la médecine libérale est d'essence néolibérale. À la suite de Dardot et Laval (2009), il convient dès lors de définir le néolibéralisme en commençant par le distinguer du libéralisme. Le libéralisme correspond à l'idéologie que l'on attribue abusivement à Adam Smith selon laquelle l'État doit laisser faire le marché. Au niveau particulier, laisser libre cours aux penchants naturels des individus à l'échange permet la réalisation, au niveau général, de la prospérité et de l'harmonie sociale. L'État doit assurer ses missions régaliennes et s'interdire toute contrainte à la liberté individuelle. L'impératif premier est de ne pas perturber l'ordre naturel marchand. Si l'histoire économique retient la formule de main invisible, la fable des abeilles de Mandeville éclaire bien plus précisément l'idée libérale[145] : les vices privés font la vertu publique.

L'État libéral est une réalité historique mais, ce n'est pas celle que nous connaissons depuis le fordisme et après l'avènement de la régulation postfordiste. En effet, il est absurde de postuler que la forme actuelle de l'État est libérale dans la mesure où jamais le nombre de règles pensées et mises en place par l'État n'a été aussi important[146].

Le cas de la régulation de la médecine libérale est exemplaire pour comprendre ce qui change entre État libéral et État néolibéral. Jusqu'au début des années 1990, la politique publique en ambulatoire ne s'intéressait qu'aux questions de prix. À ce titre, les règles encadrant l'activité ambulatoire étaient très rares. De nombreux textes sur l'assurance publique obligatoire ont émaillé le XXème siècle de telle sorte que l'on peut parler « en général » d'un État social dans le champ de la santé. La Sécurité sociale porte une dimension révolutionnaire si on se réfère au XIXème siècle. Pourtant, si l'on ne s'intéresse qu'à la médecine ambulatoire « en particulier », force est de constater que la seule règle – il est vrai

[145] En effet, chez Smith, l'identité – divine – de la main invisible est de nature à remettre radicalement en cause l'analogie entre le libéralisme économique et l'auteur de la *Richesse des Nations*.

[146] L'absurdité d'une telle thèse peut se lire dans le paradoxe entre les discours sur le libéralisme d'État et, simultanément, les discours sur l'excès de règlementation provenant de ce même État. Ces deux propositions ne peuvent tenir ensemble : soit l'État est libéral et il ne légifère pas ou peu, soit il est autre chose et il est susceptible de produire une forte quantité de textes à visée règlementaire.

importante mais, assez limitée dans le temps – est une règle de prix : le tarif opposable. En particulier, l'État semble alors, en longue période, plus libéral que social.

Ainsi, d'après Dardot et Laval (2009), le néolibéralisme n'est pas d'abord une politique économique ou une idéologie, il est une rationalité qui structure non seulement l'action des gouvernants (l'État néolibéral et l'entreprise néolibérale) mais aussi la conduite des gouvernés (les sujets néolibéraux). La rationalité néolibérale est avant tout une gouvernementalité, au sens de Foucault – en tant qu'activité cherchant à conduire l'activité des individus à l'aide des instruments étatiques. Gouverner les individus c'est alors conduire les conduites. L'originalité de la gouvernementalité par rapport aux autres formes de gouvernement traditionnelles est que la première se déleste de l'appareillage répressif propre aux secondes. Ainsi, plutôt que d'obtenir une obéissance par la peur d'une sanction, la gouvernementalité instaure un auto-gouvernement des individus – c'est-à-dire une intériorisation des règles engendrant la mise en conformité vis-à-vis de soi-même :

> *« C'est en quoi le gouvernement requiert la liberté comme sa condition de possibilité : gouverner, ce n'est pas gouverner contre la liberté ou malgré elle, c'est gouverner par la liberté, c'est-à-dire jouer activement sur l'espace de liberté laissé aux individus pour qu'ils en viennent à se conformer d'eux-mêmes à certaines normes » (Dardot et Laval, 2009, p. 15).*

Dans ces conditions, le néolibéralisme n'est pas une débureaucratisation au sens d'une réduction drastique du nombre de règles (publiques ou privées). L'État est un acteur de mise en concurrence du monde qui ne connait pas de frontière : la logique concurrentielle doit régner dans le secteur privé comme dans le secteur public. Il s'agit d'imposer et de faire adhérer à un « gouvernement à distance » des comportements grâce, notamment, au couple indicateurs de résultats/incitations. Cela suppose, au cœur de l'action publique, de contrôler étroitement le travail des agents publics (ou qui agissent au nom de l'État) en les soumettant à l'évaluation de citoyens désormais consommateurs. Si la norme concurrentielle est imposée par l'État, elle s'impose par le jeu des acteurs eux-mêmes : chacun intériorise cette norme en devenant son propre surveillant. D'où le paradoxe d'un appel à l'autonomie et à l'engagement dans le travail simultanément à une réduction de l'autonomie par les règles bureaucratiques et la pression concurrentielle.

2.2.5. La convention industrielle propagatrice du néolibéralisme

Le second grand tournant de la médecine libérale marque le changement de la nature de l'État. En abandonnant la régulation par les prix, ce dernier s'est concentré sur les pratiques. Le retour d'un certain libéralisme sur les prix se conjugue avec le développement du

néolibéralisme sur les pratiques. La création des agences de santé, les recommandations de bonnes pratiques, le paiement à la performance médicale, les démarches qualité, etc., viennent entourer le médecin dans un espace de règles provenant en grande partie de l'État lui-même.

La démultiplication de la *soft law*[147] décrite dans le chapitre 7 participe à ce néolibéralisme médical. Au-delà de ses effets de droit – qui restent assez indéterminés malgré une tendance à la reconnaissance du droit mou –, la production de normes de bonnes pratiques agit également par d'autres voies. En effet, même si les normes n'ont pas ou peu de forces contraignantes elles orientent tout de même les actions car « elles apparaissent au sein d'un univers sémantique et instrumental – évaluation, accréditation, indicateurs de 'performance', management, démarche qualité – qui atteste de la montée en puissance d'une gouvernementalité néolibérale » (Rolland et Sicot, 2012, p. 55). Ainsi, les normes sont l'un des maillons de la politique publique de santé visant à « conduire les conduites » :

> *« [Elles] constituent aujourd'hui le socle d'une gouvernementalité, caractérisée par l'information, celle collectée sur les pratiques et celle fournie aux médecins, notamment en matière d'écart à la norme de leurs pratiques. La modification des comportements et l'obtention du calage des pratiques professionnels sur les RBP, souhaitée si ce n'est effective, tient moins au degré de contrainte de l'une ou l'autre des mesures en cours qu'à leur accumulation et à leur maillage, comme le montrent les actions conjointes de l'État et de ses agences d'une part, de l'Assurance maladie d'autre part, au moyen des outils développés par ces deux institutions en association avec des représentants de la profession médicale. La force de cette forme actuelle d'encadrement et d'infléchissement des pratiques tient notamment dans le fait qu'elle ne cherche pas à contraindre mais à persuader dans le cadre d'un 'partenariat' revendiqué entre le régulateur (l'État), le payeur (l'Assurance maladie) et les producteurs (les professionnels). » (Rolland et Sicot, 2012, p. 66).*

La convention de qualité industrielle est un puissant vecteur de la nouvelle définition de ce qu'il convient de faire. C'est le socle normatif indispensable à la gouvernementalité néolibérale. Sans cette convention de qualité des soins il est impensable d'homogénéiser le « bien santé » et de diffuser les valeurs marchandes caractéristiques du néolibéralisme sanitaire.

[147] Rappelons que le droit mou est constitué de tous les textes dont la juridicité est discutée : on ne sait pas s'ils doivent ou non faire l'objet d'une obligation juridique et donc d'une sanction en cas de manquement. C'est particulièrement le cas pour tous les textes portant sur les recommandations professionnelles (recommandations de bonnes pratiques, guidelines, références médicales opposables, indicateurs du paiement à la performance, recommandation des sociétés savantes, etc.).

2.2.6. Le rôle de la convention industrielle dans la privatisation de la santé

L'émergence de la convention de qualité des soins industrielle s'inscrit dans le cadre plus général des privatisations du secteur sanitaire. D'essence néolibérale, la réforme de santé en France trouve sa cohérence autour du concept de concurrence. L'idée générale est de diffuser chez les acteurs du système de soin une rationalité marchande selon laquelle le soin n'est pas un bien particulier et qu'il peut être soumis au principe de concurrence dans l'intérêt de tous. La mise en concurrence généralisée ne peut alors être qu'un vecteur d'efficacité économique et de justice sociale.

La littérature en économie politique de la santé analyse ces transformations des systèmes de santé sous le vocable de privatisation. Cependant, l'utilisation faite de ce terme peut renvoyer à des réalités très différentes. Par exemple, il est possible de parler de privatisation du financement des soins lorsque les assurances complémentaires privées financent une part croissante de la dépense de santé. Mais, il est aussi possible de parler de privatisation lorsque l'hôpital public est invité à se gérer comme une entreprise privée. Dans les deux cas, nous avons affaire à deux formes de privatisation bien différentes.

La cartographie des privatisations réalisée par Batifoulier (2014) et André et al. (2014) permet de remettre de l'ordre dans l'analyse de la réforme sanitaire (Graphique 8.1.). Cette cartographie repose sur une double distinction des privatisations du système de santé. D'une part, la privatisation peut être imposée ou incorporée. Dans le premier cas, il s'agit du moins d'État : l'État se retire d'une activité, l'espace libéré étant occupé par le secteur privé. Les individus doivent alors se soumettre à la privatisation : elle est imposée. Dans le second cas, la privatisation est incorporée au sens où les individus sont invités à changer de rationalité pour devenir des acteurs concurrentiels. Dans les deux, cas c'est le néolibéralisme qui se diffuse puisqu'au nom de la concurrence le « moins d'État » n'est jamais loin du « mieux d'État »[148]. D'autre part, la privatisation peut porter sur le financement des soins (le payeur) ou sur la délivrance des soins (le plus souvent l'hôpital et la médecine ambulatoire).

[148] Il est possible de voir deux grandes manifestations du néolibéralisme dans l'action de l'État (Ramaux, 2010) : (i) une forme de l'action publique qui, tant dans ses services que dans la production de règles pour le secteur privé, promeut la norme concurrentielle. C'est le « mieux d'État ». (ii) un retrait de l'État au profit du secteur privé. C'est le « moins d'État ». Il serait en effet réducteur de penser que le moins d'État n'est que libéral (ou du moins, pas néolibéral) : la cohérence entre les manifestations (i) et (ii) réside dans l'application généralisée du principe de concurrence. Il n'est alors pas étonnant dans certains cas de voir l'État se retirer – tout en réglementant les activités dont il est question – pour favoriser la concurrence.

Graphique 8.1. Une pluralité de privatisations (Batifoulier, 2014)

Privatisation imposée

Développement de l'assurance privée (A)	Abandon et délégation de service public à des opérateurs privés (B)
Financement des Soins	**Délivrance des soins**
Apprentissage du comportement entrepreneurial (C)	Colonisation du public par le privé (D)

Privatisation incorporée

L'industrialisation des soins et le développement de la rémunération à la performance relève de la forme « incorporée » de la privatisation. L'apprentissage du comportement entrepreneurial (Cadrant C) a pour but de modifier la relation médicale elle-même. Les patients sont invités à regarder le système de santé avec les yeux du consommateur cherchant à s'informer sur les prix, les comparer et arbitrer entre les offreurs. L'entreprise de standardisation est essentielle dans ce cadre. Les politiques néolibérales visent aussi à imprégner le secteur public des logiques du privé (Cadrant D). L'hôpital public est souvent cité en exemple de diffusion de la norme privée de l'entreprise (Belorgey, 2011, Mas et al. 2011). On peut également voir dans les logiques d'évaluation au résultat, à l'hôpital comme en médecine de ville, l'émanation de pratiques venant du secteur privé.

Le second grand tournant de la médecine libérale a pour effet majeur de standardiser le soin. La convention de qualité industrielle est à ce titre essentielle dans la mesure où elle permet de définir un produit clairement identifiable là où n'existait qu'un service vaguement défini. La standardisation via la construction de normes et de nomenclatures est la condition nécessaire à la comparaison. Comme sur les marchés de biens de consommation traditionnels, il s'agit désormais d'ouvrir la voie à la comparaison des produits de santé. Si elle autorise la comparaison, l'homogénéisation du bien santé permet également la mise en concurrence des producteurs par les consommateurs. Elle invite alors aussi bien les médecins que les patients à se saisir de cette nouvelle façon de penser le secteur de la santé à l'aune de la norme concurrentielle.

La déprofessionnalisation du médecin conduit à la désingularisation de sa relation avec le patient. Il devient alors possible de comparer les performances médicales. Tant que chaque médecin pouvait faire valoir sa propre conception de la santé, aucune comparaison ne pouvait être faite entre les médecins. L'homogénéisation du produit entraine celle du producteur – qui peut maintenant être évalué et comparé.

L'introduction de la Rémunération sur objectif de santé publique entérine la possibilité matérielle de comparer les médecins entre eux. C'est d'ailleurs ce que fait la Caisse nationale d'assurance maladie lorsqu'elle publie des cartes présentant la performance des médecins selon leur zone d'installation géographique. Plus généralement, les systèmes d'information sont disponibles pour recalculer les rémunérations à la performance par l'intermédiaire d'un mécanisme de *benchmarking*. Il est d'ailleurs envisageable de publier les performances des médecins dans le même esprit que la publication de palmarès pour les hôpitaux.

La modification de la rationalité du médecin n'a pas d'intérêt tant que le patient reste un acteur passif. C'est pour cela que les réformes néolibérales de santé ont porté une attention toute particulière à l'apprentissage d'un comportement entrepreneurial par le patient.

Pour que le patient soit responsabilisé, il faut qu'il soit en mesure d'arbitrer entre plusieurs choix possibles. La comparaison et l'exercice de la souveraineté du consommateur a aussi besoin de faciliter la comparaison : on ne peut pas comparer ce qui n'est pas comparable. Aussi, le standard est un appui précieux à l'étalonnage des professionnels. L'homogénéisation du soin permet alors au patient de se détacher de son médecin en évinçant tout lien personnel. Dès lors, la mise en concurrence des producteurs n'est plus un problème. Le patient consommateur peut arbitrer sans distinction entre les producteurs et changer de médecin en exerçant son pouvoir de marché.

Le développement de cette convention industrielle annonce la disparition de la médecine de proximité (« le médecin de famille »). Les représentations des patients sont amenées à évoluer de façon à faire des médecins des producteurs interchangeables. Bien qu'elle s'adresse directement au médecin, la convention industrielle participe de cette transformation du patient en consommateur de soin. En passant par le médecin, c'est la mutation du patient qui est visée.

Conclusion

Le second grand tournant de la médecine libérale peut s'analyser comme une triple évolution des conventions organisant le système de santé (Tableau 8.3.). Au niveau de la prestation médicale, la convention de qualité industrielle cherche à supplanter la convention inspirée / domestique. Au niveau de l'éthique médicale, les valeurs hippocratiques et la valeur santé sont invitées à s'éclipser au profit d'une définition plus technique et bureaucratique de l'éthique médicale. Au niveau plus général de l'éthique en santé, la santé ne doit plus être vue comme un besoin mais comme une préférence – ce qui défend la mise en concurrence des acteurs plutôt que la mise en commun des ressources. En ce sens, le second grand tournant de la médecine libérale participe du néolibéralisme sanitaire.

Tableau 8.3. Les deux conceptions du soin

	La santé est un bien standard	La santé est un bien singulier
Convention de qualité	Industrielle	Inspirée / domestique
Éthique médicale	Dominante bureaucratique	Dominante hippocratique
Éthique en santé	Concurrence	Commun

Les effets du second grand tournant de la médecine libérale sont d'une portée bien plus générale que la simple rationalité médicale. Cette nouvelle convention change non seulement l'éthique des médecins mais aussi l'éthique en santé – c'est-à-dire les représentations associées à la santé de l'ensemble des acteurs du système de soin. La rationalité interprétative n'est pas la caractéristique exclusive du médecin. Par exemple, le patient se nourrit des représentations qu'il a sur le système de soin pour décider d'aller consulter ou non, d'aller en ambulatoire ou à l'hôpital, etc. Les règles complètes n'existant pas, les acteurs recourent à des conventions pour adopter la « bonne » pratique ou la « bonne » décision .

La convention industrielle fait levier au néolibéralisme sanitaire ayant pour but ultime de faire des soins un bien comme un autre au sens de la théorie économique *mainstream* – c'est-à-dire un bien marchand sur lequel les offreurs et les demandeurs doivent être en concurrence. Il s'agit ici d'une modification majeure des conventions organisant le système de santé. La convention industrielle – qui institue une nouvelle performance médicale et qui va pouvoir être rémunéré avec le paiement à la performance médicale – est un terreau fertile pour une représentation de la santé comme bien marchand. En effet, pour que la concurrence et les

échanges puissent se tenir comme sur un marché de concurrence pure et parfaite, il faut, après avoir évincé les comportements hippocratiques et déclassé le médecin, définir les contours du bien à échanger. La désingularisation mise en œuvre par la nouvelle convention de qualité des soins est le pendant de l'homogénéisation du bien. Elle invite les médecins comme les patients à se caler sur la norme de la concurrence. Au final, nous retrouvons une question transversale à cette thèse mais aussi à la littérature en économie de la santé : est-ce que la santé est un bien comme un autre ?

Conclusion Générale

L'atout majeur de la puissance publique lorsqu'elle cherche à réguler les pratiques médicales réside dans la légitimité de ce combat. Si l'on reprend l'exemple développé dans l'introduction générale, comment peut-on reprocher à la ministre de la santé de vouloir associer les médecins à la lutte contre le tabagisme ? Plus généralement, il est difficile de remettre en cause l'idée selon laquelle le régulateur central doit orienter les médecins de sorte qu'ils participent activement à l'accomplissement des objectifs de santé publique. Il en va de la santé des patients !

Cette reprise en main de la médecine libérale est d'autant plus appréciable que d'autres libertés de la profession ne cessent pas de croître. La liberté tarifaire se développe chez les médecins de secteur 2 comme de secteur 1. Le montant des dépassements n'est plus contrôlé et, dans certaines spécialités, il est difficile de trouver un médecin pratiquant le tarif opposable. En plus des problèmes posés par les pratiques tarifaires, la liberté d'installation conduit à la création de déserts médicaux. Dans certaines zones géographiques, les patients sont isolés de tout, y compris dorénavant, de l'accès au soin.

Dans ce contexte, la Rémunération sur objectif de santé publique (ROSP) et les autres formes de la régulation des pratiques médicales représentent une victoire contre la médecine libérale. En réduisant l'hétérogénéité des pratiques, la ROSP permettrait d'améliorer la qualité des soins et de maîtriser les dépenses de santé, au plus grand bénéfice du patient. Le régulateur ferait ainsi preuve de « pragmatisme » : s'il n'est pas possible de contrôler les prix, le contrôle de la qualité des pratiques reste une conquête appréciable.

Cette nouvelle régulation se caractérise par la promotion d'une modalité particulière d'évaluation du travail : la performance. La puissance publique multiplie les normes de travail quantifiées afin de fixer la définition de ce qu'il convient de faire, quelle que soit la situation. Ces standards n'étant pas en phase avec l'identité libérale de la médecine, la politique publique mobilise l'intérêt lucratif des professionnels pour diffuser la nouvelle norme du « bien ». La ROSP est alors un instrument majeur du projet visant à instituer la performance médicale.

Cependant, notre thèse vient montrer à quel point cette « conquête » est source de dangers. En effet, si la régulation par les pratiques est une problématique légitime, la régulation par les pratiques telle qu'elle s'opère aujourd'hui ne nous semble pas pertinente.

La première partie de cette thèse s'est intéressée à la façon dont la théorie économique *mainstream* s'empare du problème de régulation du médecin. Nous avons mis en évidence un paradoxe propre à l'économie de la santé. Alors que la majorité des chercheurs s'accordent à souligner l'originalité de ce secteur par rapport au reste de l'économie, les travaux sur le médecin ne s'éloignent jamais de la norme de marché de concurrence pure et parfaite. Si les économistes *mainstream* reconnaissent au médecin des « préférences sociales », l'attachement aux outils théoriques forgés pour analyser le marché conduit à produire des résultats paradoxaux. Ainsi, lorsque l'on prête au médecin une éthique médicale ou des motivations intrinsèques, tout en conservant les hypothèses de maximisation sous contrainte et de norme marchande[149], les recommandations de politique publique qui en découlent sont totalement impraticables. Dès lors, la seule recommandation cohérente s'appuie sur une représentation du médecin comme un opportuniste protégé par des rentes informationnelles. Dans ce cadre, se rapprocher de l'idéal marchand (la rémunération à la productivité marginale du travail) suppose de mettre en place un dispositif de paiement à la performance visant à réduire les effets nocifs des asymétries d'information. Au total, il nous semble que si le *mainstream* éprouve des difficultés à décrire et à analyser le monde réel, il constitue de façon évidente une justification théorique aux promoteurs de l'institutionnalisation de la performance.

La deuxième partie de cette thèse s'est concentrée sur l'évaluation empirique du paiement à la performance médicale, comme outil de la nouvelle régulation. Nous avons montré que ce type de dispositif – créant un lien arithmétique entre une part de la rémunération du médecin et l'atteinte d'objectifs quantifiés de performance – est incapable de faire la preuve de son efficacité. Les études empiriques font la preuve de l'absence de preuve : le paiement à la performance médicale n'améliore pas la qualité des soins et ne permet pas de maîtriser les dépenses de santé. En étudiant de façon plus approfondie le cas français (la ROSP), nous avons montré qu'il existe une différence problématique entre la performance mesurée et la définition profane de la performance. Deux lignes directrices ont guidé notre étude empirique sur le cas des médecins franciliens.

D'une part, il n'est pas certain que l'amélioration de la performance mesurée puisse s'interpréter comme un « mieux ». Par exemple, nous avons mis en évidence, qu'en raison d'un effet de taille de la patientèle, il est possible pour un médecin ayant des résultats quantifiés relativement faibles d'obtenir une rémunération à la performance relativement forte. Les gagnants (en termes de prime) ne sont pas forcément les performants (en termes

[149] Dans le cadre de ce que Demsetz (1969) appelle la stratégie du Nirvana.

d'indicateurs) et la ROSP peut s'interpréter comme un paiement à l'acte ou une rémunération à la capitation déguisée. D'autre part, il est certain que l'amélioration de la performance mesurée ne doit pas être systématiquement attribuée à l'action du médecin – c'est-à-dire son mérite (en positif) ou sa responsabilité (en négatif). En effet, la performance du médecin est corrélée à la composition de sa patientèle en termes d'état de santé et de revenus. Au total, les mécanismes de paiement à la performance médicale engendrent des effets inattendus qui sont souvent des effets pervers. Par exemple, en orientant le comportement vers des indicateurs chiffrés, certains aspects du métier sont en voie de disparition du fait de phénomènes de sélection des tâches, de sélection des patients, d'éviction des motivations intrinsèques et du *care*, etc.

Contre une vision naturalisée et décontextualisée du paiement à la performance, nous avons montrée dans la troisième partie de cette thèse en quoi il s'insère dans le mouvement plus général du capitalisme et de l'affrontement entre les principes de socialisation et de libéralisation du financement des soins. La ROSP n'est qu'un outil d'une évolution générale de la politique publique : elle est la dernière étape du second grand tournant de la médecine libérale. Alors qu'entre la fin du XIXème siècle et le début de années 1980 le financement de la santé est allé vers toujours plus de socialisation, la période suivante est marquée par un grand bond en arrière. La régulation par les pratiques est ce qu'il reste lorsque la puissance publique a abandonné la régulation par les prix du côté de l'offre de soin[150]. On a alors démontré qu'il n'est pas possible d'associer la régulation par les pratiques à une forme de « rationalisation » neutre et impartiale. Bien que, dans l'esprit du *New Public Management*, la réforme de santé se présente comme une évidence naturelle face à laquelle toute résistance est un corporatisme, elle relève d'une construction sociale qui est perméable à la critique. Deux axes majeurs ont été mis en évidence.

D'une part, la méthode de l'*Evidence Base Medicine* – qui nourrit la régulation par les pratiques – est hautement contestée au sein même du champ médical. Tous les médecins ne s'accordent pas pour réduire la maladie à une variation quantitative de l'état normal. Si tel n'est pas le cas, alors le problème d'organisation du système de santé change : il ne faut plus des standards de qualité mais une médecine singularisée. Indépendamment de la question de l'épistémologie de la maladie, l'essai clinique randomisé n'est pas une méthode produisant des résultats neutres dans la mesure où elle reste profondément marquée par ses conditions

[150] La politique de prix s'adresse presque exclusivement à da demande de soin – le patient.

sociohistoriques – notamment, par le fait que ce soit l'industrie pharmaceutique qui finance les essais.

D'autre part, l'économie des services met en évidence le fait que le soin n'est pas un produit. Il n'est pas possible d'établir une nomenclature complète des actes et des maladies et il n'est pas pertinent de concevoir le soin sans la participation du malade. Dès lors, il devient douteux de vouloir appliquer les méthodes de l'industrie à l'activité de soin. Comment justifier par exemple que le médecin soit évalué sur le suivi du diabète comme si l'éducation du malade à la santé n'avait pas d'importance ? Le médecin est un élément du système de santé et il est déraisonnable de lui attribuer une performance qui n'est pas (que) la sienne.

Cette façon de voir le travail du médecin est contestable. Elle est aussi contestée au nom d'une autre vision du travail du médecin et de sa performance. On peut lire la politique de régulation par les pratiques comme l'institution d'une forme de performance médicale qui modifie la convention de qualité des soins – d'une convention domestique/inspirée à une convention industrielle. Or, nous avons montré que cette mutation a des conséquences non seulement sur l'éthique médicale mais aussi sur l'éthique en santé. Les valeurs hippocratiques de la médecine sont reléguées au second plan face à une représentation plus administrative de l'activité médicale : on attend désormais du médecin une compétence technique et non un jugement moral. À *priori* marginale, cette désingularisation de la relation patient/médecin s'inscrit pleinement dans le néolibéralisme sanitaire. En effet, la standardisation de l'activité de soin est la condition nécessaire pour activer le comportement concurrentiel tant du côté de l'offreur que du demandeur de soin. À travers l'évolution du métier de médecin, c'est la conception qu'on les patients de la santé qui est visée. En faisant des soins un bien comme un autre, l'imaginaire marchand peut se développer.

La « défaite du patient »

La régulation par les pratiques participe à la construction de la nouvelle place du patient dans le néolibéralisme sanitaire. À partir de revendications légitimes face au paternalisme médical, la puissance publique a doté le patient de nouveaux droits (comme, par exemple, le droit à l'information). La stratégie d'*empowerment* du patient vise à lui donner des outils permettant de contester l'emprise du médecin sur la relation médicale. Mais, plutôt que de valoriser la discussion et la négociation dans le processus thérapeutique, le pouvoir du patient est conçu comme un pouvoir de marché – *exit* plutôt que *voice*, pour reprendre les termes d'Albert Hirschman.

L'homogénéisation des soins en germe dans la régulation par les pratiques est un élément central de cette politique. Tant que la relation de soin est fondée sur la singularité des personnes qui la font vivre, la mise en concurrence est impossible. Pour comparer les prix, il faut homogénéiser les produits. En France, la ROSP est le premier pas vers la standardisation et la certification des soins en médecine ambulatoire – dans le champ hospitalier la logique de *benchmarking* et de palmarès est déjà ancienne (Pierru, 2007). Le rôle des agences indépendantes de santé n'est donc pas tant d'améliorer la qualité des soins que d'assurer une homogénéisation des soins synonyme d'équivalence entre offreurs.

Une fois armé de droits, le patient est redevable de devoirs. La « démocratie sanitaire » n'est en effet que la contrepartie nécessaire de la responsabilisation financière du patient (Domin, 2006). Puisque désormais la santé est un bien comme un autre et que le patient peut arbitrer entre les offreurs, alors ce dernier doit payer le prix de ses préférences. Or, la mise à contribution financière des patients engendre des inégalités d'accès aux soins qui se cumulent avec des inégalités sociales de santé : les plus pauvres, qui sont aussi les plus malades, ont de moins en moins accès aux soins.

Avec la régulation par les pratiques, ces inégalités relativement classiques se doublent d'inégalités face aux indicateurs de performance. Si l'on considère que l'amélioration des indicateurs de performance n'est pas toujours un « mieux » du point de vue du patient, alors les patients les plus pauvres sont aussi les plus démunis face à la performance.

La stratégie la plus simple pour contourner les indicateurs de performance est de consulter un médecin qui n'y est pas dépendant. Dans la mesure où la part de la ROSP dans la rémunération totale est amenée à s'élever, seuls les médecins exerçant en secteur 2 pourront à terme négliger le respect des indicateurs – en contrepartie du dépassement d'honoraire. Ces médecins là pourront plus facilement se permettre financièrement d'adopter une position critique face aux indicateurs. Les médecins détachés des indicateurs et de la performance standardisée seront moins rémunérés que les autres et ne recevront pas de prime (si tant est que l'indicateur arrive à rémunérer véritablement les performants, ce qui est loin d'être le cas comme on l'a montré). Mais, s'ils se payent en dépassement d'honoraires, désormais libéralisés, leur revenu n'en sera pas affecté. L'échange de la liberté tarifaire contre le contrôle renforcé des pratiques sera neutre pour ces médecins mais certainement pas pour le patient. La différentiation de la qualité des soins entre une médecine standardisée et une médecine personnalisée risque alors de se conjuguer avec la segmentation tarifaire des médecins. Le risque est alors que les patients les plus pauvres ne puissent pas avoir accès à

cette forme de médecine. En effet, les dépassements d'honoraires peuvent devenir un facteur d'exclusion à une médecine plus personnalisée.

Dans les cas où le patient et le médecin sont en conflit sur ce qu'il convient de faire (respecter ou non l'indicateur), les inégalités prennent une autre forme. Comme l'ont montré Bruno et Didier (2013) dans les cas de la police, de l'hôpital et de l'université, l'évaluation chiffrée n'est pas un principe transcendant les rapports de domination traditionnels. Les niveaux hiérarchiques les plus élevés contrôlent sans être contrôlés. Lorsque, pour ces catégories supérieures, l'évaluation chiffrée existe et que les résultats sont médiocres, la position sociale dominante permet aux individus de relativiser leur performance – ce qui n'est pas possible pour les autres. Ce que montrent ces travaux est l'extraordinaire souplesse de l'interprétation de l'évaluation chiffrée selon la position sociale. La légitimité de la critique n'est pas la même selon le niveau hiérarchique de celui qui parle. Les travaux de Spire (2012) sur l'histoire sociale de l'impôt permettent de rapprocher cette idée du cas de la relation médicale. L'auteur montre que selon la position sociale des individus (« faible » ou « puissant ») les rapports avec les agents du fisc diffèrent radicalement : pour les puissants, la traque au fraudeur laisse place à la culture du dialogue. Les conditions d'application de la loi fiscale sont en grande partie socialement déterminées. On peut alors s'attendre aux mêmes types d'inégalités dans la relation de soin avec, d'un côté, des patients « puissants » en mesure de négocier le non-respect des critères de performance et, d'un autre côté, des patients soumis à l'impératif de la performance, parce que dominés socialement.

Ces effets néfastes sur les patients invitent à se poser des questions sur leur place et leur rôle dans le système de santé. Il apparait alors un déficit démocratique permettant certainement d'expliquer la tournure de la réforme de santé. L'État a depuis longtemps repris le contrôle sur la démocratie sociale et jamais les patients ne sont invités à prendre des décisions sur l'avenir du système de santé. Cette situation créée un consensus sur ce qu'il convient de faire (le néolibéralisme sanitaire) et un nombre important de sujets pourtant centraux sont évincés.

Perspectives pour la médecine ambulatoire

L'un des sujets important mis à l'écart par l'attention donnée à la performance médicale porte sur l'organisation de la médecine française. Le libéralisme de la profession médicale n'est pas une situation naturelle à laquelle l'on doit s'accommoder. Le sujet primordial,

seulement effleuré par la politique de contrôle des pratiques, est la redéfinition des libertés médicales et de la place du médecin dans le système de santé.

Le passage de la régulation par les prix à la régulation par les pratiques se traduit en termes de libertés médicales par l'abandon des libertés thérapeutique et de prescription au profit des libertés tarifaire et d'installation. Or, comme l'a montré cette thèse, dans le cadre de la convention de qualité domestique/inspirée, si les caractéristiques de la relation médicale peuvent justifier les libertés thérapeutique et de prescription, rien ne permet de justifier la liberté tarifaire et la liberté d'installation.

Dans cette perspective, la condition de la qualité des soins n'est pas tant le respect du standard que le jugement – c'est-à-dire la résolution du problème de passage des connaissances générales au cas particulier. Pour s'assurer de la qualité du jugement, il faut s'assurer que le médecin ait les moyens d'un jugement indépendant – de l'État et du marché. Cependant, contrairement à ce qu'assurent les représentants de la profession, la solution libérale n'est pas viable car, tant que le médecin doit arbitrer entre le nombre d'acte par jour et le temps par acte, il est enchaîné à sa contrainte financière. Le libéralisme médical n'est pas une situation d'indépendance du jugement.

Une vraie solution pour rendre le jugement indépendant de la contrainte financière est de déconnecter la rémunération des indicateurs de production (en volume ou en performance), en prenant au sérieux l'existence d'une éthique médicale ou de motivations intrinsèques. Si l'on part du principe qu'en général et la plupart du temps les médecins veulent bien faire leur travail, alors une rémunération non liée à la productivité est une solution viable.

Cependant, il ne faut pas s'en tenir là. Le contrôle est légitime ; mais il peut prendre d'autres formes que ce que propose actuellement la puissance publique. La première source de contrôle pour une profession est celle des pairs. L'histoire de l'Ordre des médecins a néanmoins montré qu'il ne faut pas attendre de lui autre chose que la défense des intérêts matériels de la profession. Une raison de la faiblesse du contrôle par les pairs en France est l'absence de collectifs de travail. La médecine libérale est une médecine individuelle. Or, pour qu'il y ait un contrôle par les pairs rigoureux, il faudrait des interactions entre médecins sur leur lieu de travail. Une nouvelle forme de régulation devrait prendre le soin de créer des collectifs de travail. D'autres formes du contrôle sont envisageables et surtout celle du contrôle démocratique : il est surprenant que les patients et les citoyens, en tant que parties

prenantes du système santé, ne soient jamais convoqués lorsqu'il s'agit de définir l'orientation du travail médical.

Une réorganisation radicale de la médecine ambulatoire peut s'appuyer sur les deux principes que nous venons d'énoncer : liberté du jugement (du médecin) et le contrôle (par les pairs et démocratique). Pour ce faire, il existe une institution qui a fait ses preuves : le salaire (Friot, 2012a et b). Le salaire est la forme de rémunération par excellence qui laisse le travail à mi-distance entre liberté et contrôle. Si elle existe, la contrainte financière n'est pas aussi importante que dans le cas des rémunérations variables (à l'acte ou à la performance). Plus rigoureux que dans le cas du paiement à l'acte, le contrôle est moins impératif que dans le cas de la rémunération à la performance.

En outre, l'exercice salarié en centre de santé libère des temps là où le libéralisme contraint à la rentabilité. Si l'on pense à la formation professionnelle, il est bien plus facile de l'organiser dans un centre de santé que dans un cabinet libéral : pendant l'absence du professionnel en formation, il y a toujours un médecin pour s'occuper des patients et le temps de formation n'est pas un coût d'opportunité. Plus généralement, le salariat, en créant des temps gratuits, rémunère ce qui est essentiel (la formation, l'attention accordée au patient, la transmission des dossiers, etc.).

Bien que cela ne soit pas l'objectif principal, le salariat permet également d'assurer une maîtrise des dépenses plus rigoureuse. Contrairement au paiement à l'acte, il deviendrait désormais possible d'anticiper l'évolution de la masse salariale. Un autre des avantages de la solution salariale est de lutter efficacement contre les inégalités d'accès aux soins. Le régulateur éviterait de financer des dispositifs de filet de sécurité coûteux (comme les dispositifs ALD, CMUC, etc.) dans la mesure où la gratuité de la consultation consécutive à la mise en place du salariat résoudrait par construction le problème de non recours aux soins pour raison financières. Enfin, en évitant les reports de soin par la facilitation de l'accès aux soins, la puissance publique économiserait les dépenses liées à l'aggravation des états de santé.

Si le salariat est une solution pour garantir l'indépendance du jugement, c'est aussi une solution pour le contrôle. En effet, la salarisation doit être pensée avec le principe de centre de santé – et donc de création d'un véritable collectif professionnel. C'est dans ce cadre collectif que doit se concevoir le contrôle des pairs. L'individualisme de la médecine libérale éloigne les praticiens les uns des autres. En situation d'éloignement, il n'est donc pas étonnant que les bonnes pratiques, même lorsqu'elles sont consensuelles, ne se diffusent pas ou peu et que les

mauvaises pratiques restent impunies. Le travail collectif impose une émulation et un contrôle social bien plus puissant que n'importe quelle norme de bonne pratique.

Le denier échelon du contrôle peut alors être un échelon démocratique. S'il est difficile de mettre un citoyen derrière chaque médecin, il est bien plus simple d'élire des représentants au pilotage des centres de santé. Evidemment, leur rôle et leur pouvoir reste à définir mais, le principe démocratique semble être un rouage essentiel dans la définition des systèmes de santé.

Ces perspectives pour la profession médicale peuvent paraître irréalisables en raison de l'opposition de la médecine libérale. Ce serait oublier deux points essentiels de l'évolution du système de santé actuel. D'une part, les jeunes médecins sont très peu séduits par l'exercice libéral. Au contraire, ils ont une appétence pour le salariat que n'avaient pas leurs ainés (Barlet et al. 2011). D'autre part, l'évolution du capitalisme semble inéluctablement menacer le statut libéral de la profession médicale. En effet, avec la croissance des financements privés, la médecine française court le risque d'un « scénario à l'américaine » (Pierru, 2007) : sous peine de ne pas vouloir s'accorder avec la puissance publique, les médecins français risquent de se voir imposer une régulation bien plus sévère de leurs pratiques et de leurs tarifs provenant du secteur privé, à l'image de ce qui est arrivé aux médecins américains.

Les deux logiques de la performance

Comme nous l'annoncions au début de cette thèse, le médecin est un exemple exemplaire des raisons pour lesquelles la théorie *mainstream* ne permet pas de comprendre les phénomènes économiques. L'étude de la relation médicale, dans une approche institutionnaliste, permet de comprendre ce qui est en jeu dans la montée en puissance de la « performance totale » (Jany-Catrice, 2012). Il est possible de donner une clef de lecture générale à partir de ce que l'on propose d'appeler les deux logiques de la performance.

- La logique prométhéenne : Benoît Heilbrunn (2004) définit la performance comme une idéologie prométhéenne d'après laquelle l'homme individualisé est capable de peser seul sur le cours des choses. On peut étendre cette définition aux cas où une organisation sociale (entreprise, État ou autre) prétend que ses résultats agrégés dépendent de la performance individuelle de ses membres. Comme si le tout était la somme des parties. Améliorer la performance des médecins serait la clef de voute de l'amélioration du système de santé. Améliorer la performance de l'entreprise serait la clef de voute de l'augmentation du profit.

Or, cette logique est fausse. Il n'est pas possible d'attribuer aux individus la responsabilité ou les mérites de résultats qui sont avant tout collectifs. Cette contradiction, en plus de générer des inefficacités, semble être à l'origine des phénomènes de souffrance au travail.

- La logique néolibérale : L'attention portée aux problèmes de performance individuelle a pour effet majeur de détourner les regards de sujets impensés. Dans le cas de la politique publique, puisque tous les enjeux portent sur la performance des fonctionnaires, il n'est plus question de proposer au débat, par exemple, l'augmentation des dépenses et des recettes publiques. L'hôpital ou l'université ne seraient pas victimes de sous financement mais de trop faibles performances individuelles. Dans le cas de l'entreprise, les mauvaises performances individuelles sont instrumentalisées pour justifier les licenciements pour motif personnel ou pour expliquer les mauvaises performances collectives. Dans le même registre, les débats sur la croissance permettent d'évincer ceux sur la répartition et sur la production.

En naturalisant le problème de la performance individuelle, les classes dominantes (économiques ou politiques) se défaussent de leurs responsabilités dans la crise actuelle. L'institution de ce régime de la performance totale désigne l'individu au travail comme seul responsable des problèmes qui résultent en réalité des contradictions du capitalisme néolibéral.

Références bibliographiques

Aballea P., Bartoli F., Eslous L. et Yeni I. (2007), *Les dépassements d'honoraires médicaux*, Rapport n° RM 2007-054P, Inspection Générale des Affaires Sociales.

Abecassis et Domin (2008), « Comment parlent les conventions médicales ? », dans Batifoulier P., Gadreau M. et Vacarie I. (Eds), *La gouvernance de l'assurance maladie : l'orientation marchande et ses paradoxes*, Rapport final pour la Mire-DREES, pp.78-102.

ACDC (2012), *Chômage et précarité à un niveau record*, Note n°10, 26 avril.

Akerlof G. (1970), « The market for 'lemons': Quality uncertainty and the market mechanism », *Quarterly Journal of Economics*, Volume 84, Numéro 3, pp. 488-500.

Akerlof G. (1982), « Labour contracts as partial gift exchange », *The Quarterly Journal of Economics*, Volume 97, Numéro 4, pp. 543-569.

Akerlof G. et Kranton R. (2000), « Economics and identity », *The Quarterly Journal of Economics*, Volume 115, Numéro 3, pp. 715-753.

Akerlof, G. et Yellen, J. (1990), « The fair wage-effort hypothesis and unemployment », *The Quarterly Journal Economics*, Volume 105, Numéro 2, pp. 255-283.

Albouy V. et Déprez M. (2009), « Mode de rémunération des médecins », *Économie & Prévision*, Volume 2, Numéro 188, pp. 131-139.

Allen J. (1993), « Economics and the health care system : changing physicians' practices », *Transactions of the American Clinical and Climatological Association*, Volume 105, Numéro 1, pp. 1-18.

Allouache A. et Vacarie I. (2008), « La conception de la règle », dans Batifoulier P., Gadreau M. et Vacarie I. (Eds), *La gouvernance de l'assurance maladie : l'orientation marchande et ses paradoxes*, Rapport final pour la Mire-DREES, pp. 59-70.

Alshamsan R., Majeed A., Ashworth M., Car J. et Millett C. (2010), « Impact of pay for performance on inequalities in health care: systematic review », *Journal of Health Services Research & Policy*, Volume 15, Numéro 3, pp. 178-184.

Ammi M. (2011), *Analyse économique de la prévention. Offre de prévention, incitation et préférence en médecine libérale*, thèse de doctorat en science économique, université de Bourgogne.

Andersen L. (2009), « Qu'est ce qui détermine le comportement et la performance des professionnels de santé ? La motivation de service public, les normes professionnelles et/ou les incitants financiers ? » *Revue Internationale des Sciences Administratives*, Volume 74, Numéro 1, pp. 87-108.

André C. Batifoulier P. et Jansen M. (2014), « Les privatisations du soin en Europe. De la théorie à la pratique », *Communication au 4ème Congrès de l'Association française d'économie politique*, juillet 2014.

Ariely D, Bracha A. et Meier S. (2009), « Doing good or doing well? Image motivation and monetary incentives in behaving prosocially », *American Economic Review*, Volume 99, Numéro 1 pp. 544-555.

Arrow, K. (1962). « Economic welfare and the allocation of resources for invention », dans *The rate and direction of inventive activity: Economic and social factors*, National Bureau of Economic Research, pp. 609-626.

Arrow K. (1963), « Uncertainty and the welfare economics of medical care », *American Economic Review*, Volume 53, Numéro 5, pp. 941-973.

Arrow K. (1968), « The Economics of Moral Hazard: Further Comment », *American Economic Review*, Volume 58, Numéro 3, pp. 537-539.

Arrow K. (1970), « Political and economic evaluation of social effects and externalities », dans Margolis J. (Ed), *The analysis of public output*, UMI, pp. 1-30.

Arrow K. (1972), « Gifts and Exchanges », *Philosophy and Public Affairs*, Volume 1, Numéro 4, pp. 343-362.

Arrow K. (1974), *The limits of organization*, Norton, New York.

Ashworth M., Lloyd D., Smith R., Wagner A. et Rowlands G. (2007a) « Social deprivation and statin prescribing: a cross-sectional analysis using data from the new UK general practitioner 'Quality and Outcomes Framework' », *Journal of Public Health*, Volume 29, Numéro 1, pp. 40-7.

Ashworth M, Seed P, Armstrong D, Durbaba S et Jones R. (2007b), « The relationship between social deprivation and the quality of primary care: a national survey using indicators from the UK Quality and Outcomes Framework », *British Journal of General Practice*, Volume 57, Numéro 9, pp. 441-8.

Atkinson J. (1964), *An introduction to motivation*, Van Nostrand, Oxford.

Austin R. et Gittell J. (1999), « Anomalies of High Performance: reframing economic and organizational theories of performance measurement », Harvard Business School, Harvard University.

Azariadis C. (1975), « Implicit contracts and underemployment equilibria », *Journal of Political Economy*, Volume 83, Numéro 6, pp. 1183-1202.

Bachimont J., Cogneau J. et Letourmy A. (2006), « Pourquoi les médecins généralistes n'observent-ils pas les recommandations de bonnes pratiques ? L'exemple du diabète de type 2 », *Sciences Sociales et Santé*, Volume 24, Numéro 2, pp. 75-104.

Baker G. (1992), « Incentive contracts and performance measurement », *Journal of Political Economy*, Volume 100, Numéro 3, pp. 598-614.

Ballet J., Barillot S., et Mahieu F-R. (2007), « Le dilemme du Samaritain, une mise en perspective », *Mondes en développement*, Volume 1, Numéro 137, pp. 11-22.

Ballet J. et Bazin D. (Eds) (2006), *Positive ethics in economics*, NJ: Transaction Publishers, London and New Brunswick.

Bardey D. (2002), « Demande induite et réglementation de médecins altruistes », *Revue Économique*, Volume 53, Numéro 3, pp. 581-588.

Bardey D. et Lesur R. (2006), « Optimal regulation of health system with induced demand and ex post moral hazard », *Annales d'Économie et de Statistique*, Volume 83/84, pp. 279-293.

Barlet M., Bellamy V., Guillaumat-Tailliet F. et Jakoubovitch S. (2011), « Médecins généralistes : que pensent-ils de leur rémunération ? », *Revue Française des Affaires Sociales*, Volume 2/3, Numéro 2/3, pp. 123-155.

Bartoli A. (2005), *Le management dans les organisations publiques*, Dunod, Paris.

Batifoulier P. (Ed) (2001), *Théorie des conventions*, Economica, Paris.

Batifoulier P. (2009), *La valeur des valeurs. Économie institutionnaliste de la santé*, Habilitation à diriger des recherches en sciences économiques, Université Paris Ouest Nanterre la Défense.

Batifoulier P. (2011), « Valeurs et convention. Une économie politique de la santé », dans Batifoulier P., Buttard A., Domin J-P. (Eds), *Santé et politiques sociales. Entre efficacité et justice, Autour des travaux de Maryse Gadreau*, Eska, Paris.

Batifoulier P. (2013), « Faire payer le patient : une politique absurde », *Revue du MAUSS*, Volume 41, Numéro 1, pp. 77-92.

Batifoulier P. (2014), *Capital santé. Quand le patient devient client*, La Découverte, Paris.

Batifoulier P. et Biencourt O. (2005), « La coloration 'libérale marchande' de l'éthique médicale. Une analyse conventionnaliste », dans Batifoulier P. et Gadreau M. (Eds), *Éthique médicale et politique de santé*, Economica, Paris, pp. 181-198.

Batifoulier P., Braddock L. et Latsis J. (2013), « Priority Setting in Health Care: From Arbitrariness to Social Values », *Journal of Institutional Economics*, Volume 9, Numéro 1, pp. 61–80.

Batifoulier P., Braddock L. et Latsis J. (2014), « A la recherche de la justification perdue des ordres de priorités en santé », *Les Tribunes de la Santé*, Volume 5, Numéro 3, pp. 75-87.

Batifoulier P. et Da Silva N. (2014), « The economic behaviour of doctors: medical altruism without an ethic? », *Document de travail EconomiX*, Numéro 2014.

Batifoulier P. et Da Silva N. (2014), « Medical Altruism in Mainstream Health Economics: Theoretical and Political Paradoxes », *Review of Social Economy*, Volume 72, Numéro 3, pp. 261-279.

Batifoulier P. Domin J-P. et Abecassis P. (2010), « Les effets pervers de la politique incitative : l'exemple de l'assurance maladie complémentaire », *Économie Appliquée*, Volume 63, Numéro 4, pp. 171-194.

Batifoulier P. Domin J-P et Gadreau M. (2007), « La gouvernance de l'assurance maladie au risque d'un État social marchand », *Économie appliquée*, Volume 60, Numéro 2, pp. 101-126.

Batifoulier P., Domin, J. P., et Gadreau, M. (2008), « Mutation du patient et construction d'un marché de la santé. L'expérience française », *Revue Française de Socio-Économie*, Volume 1, Numéro 1, pp. 27-46.

Batifoulier P., Domin J-P. et Gadreau M. (2011), « Market empowerment of the patient : the French experience », *Review of Social Economy*, Volume 69, Numéro 2, pp. 143-162.

Batifoulier P., Eymard-Duvernay F. et Favereau O. (2007), « État social et assurance maladie. Une approche par l'économie des conventions », *Économie Appliquée*, Volume 60, Numéro 1, pp. 203-229.

Batifoulier P. et Gadreau M. (2007), « Régulation et coordination du système de santé. Des institutions invisibles à la politique économique », dans Eymard-Duvernay F. (Ed), *L'économie des conventions, méthodes et résultats, Tome II*, La Découverte, Paris, pp. 453-468.

Batifoulier P, Gadreau M. et Lievaut J. (2009) « La décision tarifaire : des actes gratuits aux actes chers. Les effets pervers de la politique publique », *Journal d'Économie Médicale*, Volume 27, Numéro 4, pp. 222-234.

Batifoulier P. et Gadreau M. (2006), « Régulation et coordination du système de santé. Des institutions invisibles à la politique économique », dans Eymard-Duvernay F., *L'économie des conventions méthode et résultats*, La découverte, Recherches, Paris.

Batifoulier P., Latsis J. et Merchiers J. (2010), Les priorités de la prise en charge financière des soins. Une approche par la philosophie », *Document de travail EconomiX*, Numéro 2010-02.

Batifoulier P. et Thévenon O. (2003), « L'éthique (médicale) est-elle soluble dans le calcul économique ? », *Économie appliquée*, Volume 2, Numéro 3, pp. 161-186.

Baumister R. et Leary M. (1995), « The Need to Belong : Desir for Interpersonal Attachements as a Fundamental Human Motivation », *Psychological Bulletin*, Volume 117, Numéro 3, pp. 497-529.

Beauvallet M. (2010), *Les stratégies absurdes. Comment faire pire en croyant faire mieux*, Points, Paris.

Becker, G. (1981), « Altruism in the Family and Selfishness in the Market Place », *Economica*, Volume 48, Numéro 189, pp. 1-15.

Belorgey, N. (2010), *L'hôpital sous pression: Enquête sur le nouveau management public*, La Découverte, Paris.

Belorgey N. (2011), « Réduire le temps d'attente et de passage aux urgences. Une entreprise de "réforme" d'un service public et ses effets sociaux », *Actes de la Recherche en Sciences Sociales*, Volume 4, Numéro 189, pp. 16-33.

Ben-Her A. et Putterman L. (1998), « Values and Institutions in Economic Analysis », dans *Economics, Values and Organization*, Ben-Her A. and Putterman L. (Eds), Cambridge University Press, Cambridge, pp. 3–69.

Bénabou R. et Tirole J. (2003), « Intrinsic and Extrinsic Motivation », *Review of Economic Studies*, Volume 70, Numéro 3, pp. 489-520.

Bénabou R. et Tirole J. (2006), « Incentives and Prosocial Behavior », *American Economic Review*, Volume 96, Numéro 5, pp. 1652- 1678.

Benamouzig D. (2009), « Les agences sanitaires : entre réforme et technocratie », dans Meimon J. (Ed), *Les réorganisations administratives : Bilan et perspectives en France et en Europe*, Institut de la gestion publique et du développement économique, Paris, pp. 73-93.

Benamouzig D. et Besançon J. (2005), « Administrer un monde incertain : les nouvelles bureaucraties techniques. Le cas des agences sanitaires en France », *Sociologie du Travail*, Volume 47, Numéro 3, pp. 301-322.

Benavent J., Juan C., Clos J., Sequeira E., Gimferrer N. et Vilaseca, J. (2009), « Using pay-for-performance to introduce changes in primary healthcare centres in Spain: first year results », *Quality in Primary Care*, Volume 17, Numéro 2, pp. 123-131.

Berle A. et Means G. (1932), *The modern corporation and private property*, McMillan, New York.

Berthe B. (2001), *L'effort au Travail. Analyse d'un Concept Économique*, Presses Universitaires de Rennes, Rennes.

Berwick D., DeParle N-A, Eddy D., Ellwood P., Enthoven A., Halvorson G., Kizer K., McGlynn E., Reinhardt U., Reischauer R., Roper W., Rowe J., Schaeffer L., Wennberg J. et Wilensky G. (2003), *Health Affairs*, Volume 22, Numéro 6, pp. 8-10.

Besley T. et Ghatak M. (2005), « Competition and incentives with motivated agents », *American Economic Review*, Volume 95, Numéro 3, pp. 616-636.

Bessis F. (2008), « Conventions de qualité et réflexivité : que faut-il attendre (et ne pas attendre) du marché ? », dans *Les avocats entre ordre professionnel et ordre marchand. Concurrence par la qualité et socio-économie d'une réglementation professionnelle*, Rapport pour le Conseil National des Barreaux, pp.103-140

Bessis F. et Favereau O. (2010), « Le marché contre les professions », dans Batifoulier P., Buttard A. et Domin J-P., *Santé et politiques sociales : entre et efficacité et justice. Autour des travaux de Maryse Gadreau*, Editions ESKA, Paris, pp. 82-92.

Bewley T. (1995), « A depressed labor market as explained by participants », *American Economic Review*, Volume 85, Numéro 2, pp. 250-254.

Bezes P. (2005), « Le renouveau du contrôle des bureaucraties », *Informations Sociales*, Numéro 6, Numéro 126, pp. 26-37.

Bidet A. et Vatin F. (2009), « Mesure et acteurs au travail », dans Steiner P. et Vatin F. (Eds), *Traité de sociologie économique*, Presses Universitaires de France, Quadrige, Paris.

Bien F. (2001), *Essais en Économie de la Santé et Assurance*, Thèse pour le doctorat de Sciences Économiques, Université Paris X-Nanterre.

Biencourt O., Chaserant C. et Rebérioux A. (2001), « L'économie des conventions : l'affirmation d'un programme de recherche », dans Batifoulier P. (Ed), *Théorie des conventions*, Economica, Paris, pp. 193-218.

Billaudot B., (2008), « Une théorie de l'État social », *Revue de la régulation. Capitalisme, institution, pouvoirs*, Numéro 2, http://regulation.revues.org/2523, consulté le 15/10/14.

Blaug M. (1998), « Where are we now in British health economics? », *Health Economics*, Volume 7, S1, pp. 63-78.

Blavier P. et Montialoux C. (2009), « Ce qui ne va pas dans le système de santé américain » *Regards Croisés sur l'Économie*, Volume 5, Numéro 1, pp. 138-140.

Bloche M. (2001), « The market for medical ethics », *Journal of Health Politics, Policy and Law*, Volume 26, Numéro 5, pp. 1099-1112.

Blomqvist A. (1991), « The doctor as double agent : information asymmetry, health insurance, and medical care », *Journal of Health Economics*, Volume 10, Numéro 4, pp. 411-432.

Bloy G. et Rigal L. (2012), « Avec tact et mesure ? Les médecins généralistes français aux prises avec les évaluations chiffrées de leur pratique », *Sociologie du Travail*, Volume 54, Numéro 4, pp. 433-456.

Bohnet I., Frey B. et Huck S. (2001), « More order with less law: On contract enforcement, trust, and crowding », *American Political Science Association*, Volume 95, Numéro 1, pp. 131-144.

Boisard P. et Letablier M. (1987), « Le camembert: normand ou normé. Deux modèles de production dans l'industrie fromagère », dans *Entreprises et produits*, Cahier du CEE, Paris, pp. 1-30.

Boltanski L. et Chiapello E. (1999), *Le nouvel esprit du capitalisme*, Gallimard, Paris.

Boltanski L. et Thévenot L. (1991), *De la justification. Les économies de la grandeur*, Gallimard, Paris.

Boyer R. (1986), *La théorie de la régulation: une analyse critique*, La découverte, Paris.

Boyer R. (2004), *Théorie de la régulation 1. Les fondamentaux*, La Découverte, Paris.

Boyer R. et Saillard Y. (1995), « Un précis de la régulation », dans Boyer, R. et Saillard Y. (Eds.), *Théorie de la régulation. L'état des savoirs*, La Découverte, Paris, pp. 58-68.

Bras P-L. (2011), « Le contrat d'amélioration des pratiques individuelles (CAPI) à la lumière de l'expérience anglaise », *Journal d'Économie Médicale*, Volume 29, Numéro 5, pp. 216-230.

Brekke K., Kverndokk S. et Nyborg K. (2003) « An economic model of moral motivation », *Journal of Public Economics*, Volume 87, Numéro 9/10, pp. 1967-1983.

Brennan, G. et Buchanan J. (1983), « Predictive power and the choice among regimes », *The Economic Journal*, Volume 93, Numéro 369, pp. 89-105.

Brown L. (2011), « 'Obamacare' : où en sommes-nous ? », *Les Tribunes de la Santé*, Volume 30, Numéro 1, pp. 39-55.

Bruni M., Nobilio L. et Ugolini C. (2008), « Economic incentives in general practice: the impact of pay-for-participation and pay-for-compliance programs on diabetes care », *Health Policy*, Volume 90, Numéro 2/3, pp. 140-148.

Bruno I. et Didier E. (2013), *Benchmarking: l'État sous pression statistique*, La Découverte, Paris

Buchanan J. (1975), « The Samaritan's dilemma », dans *Altruism, morality and economic theory*, Phelps E. (Ed), Russel sage foundation, New York, pp. 71-85.

Buetow S. (2008), « Pay-for-performance in New Zealand primary health care », *Journal of Health Organization and Management*, Volume 22, Numéro 1, pp. 36-47.

Bureau D. et Mougeot M. (2007), *Performance, incitations et gestion publique, Conseil d'analyse économique*, La documentation française.

Cahuc P. (1990), « La théorie des négociations salariales : une revue de la littérature », *Économie & Prévision*, Volume 92, Numéro 1-2, pp. 21-30.

Caille A. (2005), *Dé-penser l'économique*, La Découverte, Paris.

Caillé A. (2009), *Théorie anti-utilitariste de l'action, fragments d'une sociologie générale*, Edition La Découverte, Paris.

Caisse nationale d'assurance maladie des travailleurs-salariés (2013), *Rémunération sur objectif de santé publique : une mobilisation des médecins et de l'assurance maladie en faveur de la qualité des soins*, Dossier de presse, 11 avril.Laval, 2007

Camerer C., Lowenstein G. et Rabin, M. (Eds) (2011), *Advances in Behavioural Economics*, Princeton university press, Princeton.

Cameron P., Kennedy M., Mcneil J. (1999), « The effects of bonus payments on emergency service performance in Victoria », *Medical Journal of Australia*, Volume 171, Numéro 5, pp. 243-246.

Campbell S., Reeves D., Kontopantelis E., Sibbald B., Roland M. (2009), « Effects of pay for performance on the quality of primary care in England », *New England Journal of Medicine*, Volume 361, Numéro 4, pp. 368-378.

Canadian Medical Association (2010), *Health care transformation in Canada. Change that works, care that lasts*, Canadian Medical Association, Ottawa, Ontario.

Canguilhem G. (1966), *Le normal et le pathologique*, PUF, Paris.

Cardenas J., Stranlund J. et Willis C. (2000), « Local environmental control and institutional crowding-out », *World Development*, Volume 28, Numéro 10, pp. 1719-1733.

Castel R. (1995), *Les métamorphoses de la question sociale: une chronique du salariat*. Fayard, Paris.

Castel P. et Merle I. (2002), « Quand les normes de pratiques deviennent une ressource pour les médecins », *Sociologie du Travail*, Volume 44, Numéro 3, pp. 337-355.

Castel P. et Dalagarrondo S. (2005), « Les dimensions politiques de la rationalisation des pratiques médicales », *Sciences Sociales et Santé*, Volume 23, Numéro 4, pp.6-40.

Chaix-Couturier C., Durand-Zaleski I., Jolly D. et Durieux P. (2000), « Effects of financial incentives on medical practice: results from a systematic review of the literature and methodological issues », *International Journal for Quality in Health Care*, Volume 12, Numéro 2, pp. 133-142.

Champy F. (2009), *La sociologie des professions*, PUF, Paris.

Champy F. (2011), *Nouvelle théorie sociologique des professions*, PUF, Paris.

Chanial P. (Ed) (2008), *La société vue du don*, La Découverte, Paris.

Chanial P. (2010), « Le New Public Management est-il bon pour la santé ? », *Revue du MAUSS*, Volume 35, Numéro 1, pp. 135-50.

Chavance B. (2007), *L'économie institutionnelle*, La Découverte, Paris.

Clerc I., Paraponaris A., Ventelou B., Verger P. et le Comité de pilotage du Panel de médecins généralistes de ville (2011), « Rôle des attitudes a priori des médecins généralistes français vis-à-vis des recommandations de bonnes pratiques » *Revue d'Epidémiologie et de Santé Publique*, Volume 59, Numéro 2.

Clerc I., Ventelou B., Guerville M., Paraponaris A. et P. Verger (2011), « General practitioners and clinical practice guidelines: a reexamination », *Medical Care Research and Review*, Volume 68, Numéro 4, pp. 504-518.

Clerc I., B. Ventelou et P. Verger (2010), « Comportement des médecins généralistes face aux recommandations de bonnes pratiques », *Revue d'Epidémiologie et de Santé Publique*, Volume 58, Numéro 2.

Cohu S. et Lequet-Slama D. (2007), « Le système d'assurance santé aux États-Unis. Un système fragmenté et concurrentiel », *Etudes et Résultats*, Numéro 600.

Coleman K., Reiter K., Fulwiler D. (2007), « The impact of pay-for-performance on diabetes care in a large network of community health centers », *Journal of Health Care for the Poor and Underserved*, Volume 18, Numéro 4, pp. 966-983.

Commission des comptes de la sécurité sociale (2011), *Les comptes de la sécurité sociale. Résultats 2010*.

Commission des comptes de la sécurité sociale (2012), *Les comptes de la sécurité sociale. Résultats 2011*.

Commission des comptes de la sécurité sociale (2013), *Les comptes de la sécurité sociale. Résultats 2012*.

Conseil National de l'Ordre des Médecins (2009), *Circulaire du CNOM adressée aux conseils départementaux et régionaux*, http://www.ordmed31.org/corbeille/capi/, consulté le 15/10/14.

Conseil National de la Résistance (1944), *Les jours heureux*, http://fr.wikisource.org/wiki/Programme_du_Conseil_national_de_la_R%C3%A9sistance, consulté le 15/10/14.

Convention médicale (2011), *Convention nationale organisant les rapports entre les médecins libéraux et l'assurance maladie signée le 26 juillet 2011*, http://www.csmf.org/upload/File/Conv_med/conv_med_annexes_110726.pdf, consulté le 15/10/14.

Cordier A. (2003), *Éthique et professions de santé*, Rapport au Ministre de la santé, de la famille et des personnes handicapées, Paris.

Cordonnier L. (2000), *Pas de pitié pour les gueux : sur les théories économiques du chômage*, Raison d'agir, Paris.

Cowen E. (1994), « The enhancement of psychological wellness : challenges and opportunities », *American Journal of Community Psychology*, Volume 22, Numéro 2, pp. 149-179.

Crozet Y. (1997), *Analyse économique de l'État*, Armand Colin, Paris.

Curtin K., Beckman H., Pankow G., Milillo Y., Greene R. (2006), « Return on investment in pay for performance: a diabetes case study », *Journal of Healthcare Management*, Volume 51, Numéro 6, pp. 365-376.

Da Silva N. et Rauly A. (2014), « Le difficile déploiement de la télémédecine : une lecture par l'économie des conventions », dans Lamotte B., Le Roy A., Massit C., Puissant E.

(Eds), In*novations sociales, innovations économiques, XXXIVes Journées de l'Association d'économie sociale*, Presses Universitaires de Louvain, Louvain, pp. 145-162.

Damberg C., Raube K., Teleki S. et de la Cruz E. (2009), « Taking stock of pay-for-performance: a candid assessment from the front lines », *Health Affairs*, Volume 28, Numéro 2, pp. 517-525.

Darbon S. et Letourmy A. (1983), « La microéconomie des soins médicaux doit-elle nécessairement être d'inspiration néoclassique ? », *Sciences Sociales et Santé*, Volume1, Numéro 2, pp. 31-77.

Dardot P. et Laval C. (2009), *La nouvelle raison du monde. Essai sur la société néolibérale*, La découverte, Paris.

Davis B. et Capenter C. (2009), « Proximity of fast-food restaurants to schools and adolescent obesity », *American Journal of Public Health*, Volume 99, Numéro 3, pp. 505-510.

Davis J. (2001), (Ed), *The social economics of health care*, Routledge, London.

Davis J. et McMaster R. (2007), « The individual in mainstream health economics : A case of persona non-grata », *Health Care Analysis*, Volume 15, Numéro 3, pp. 195-210.

Davis J. et McMaster R. (2013), « Situating care in health economics », *Working Paper 2013/02*, Adam Smith Research Foundation.

De Brantes F. et D'Andrea B. (2009), « Physicians respond to pay-for-performance incentives: larger incentives yield greater participation », *American Journal of Managed Care*, Volume 15, Numéro 5, pp. 305-310.

De Bruin S., Baan C. et Struijs J. (2011), « Pay for performance in disease management: a systematic review of the literature », *BMC Health Service Research*, Volume 11, Numéro 272, pp. 1-14.

De Gaulejac V. (2009), *La société malade de la gestion. Idéologie gestionnaire, pouvoir managérial et harcèlement social*, Points Économie, Paris.

De Gaulejac V. (2011), *Travail, les raisons de la colère*, Seuil, Paris.

De Jaegher K. et Jegers M. (2000), « A model of physician behaviour with demand inducement », *Journal of Health Economics*, Volume 19, Numéro 2, pp. 231-258.

Deci E. (1971), « Effects of externally mediated rewards on intrinsic motivation », *Journal of Personality and Social Psychology*, Volume 18, Numéro 1, pp. 105-115.

Deci E., Koestner R. et Ryan R. (1999), « A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation », *Psychological Bulletin*, Volume 125, Numéro 6, pp. 627-668.

Deci E. et Ryan R. (1985), *Intrinsic motivation and self-determination in human behaviour*, Plenum Press, New York.

Deci E. et Ryan R. (2000), « The 'What' and 'Why' of Goal Pursuits : Human Needs and the Self-Determination of Beavior », *Psychological Inquiry*, Volume 11, Numéro 4, pp. 227-268.

Deci E. et Ryan R. (2002), *Handbook of Self-Determination Research*, The University of Rochester Press, Rochester.

Degli A. (2009), « Intrinsic vs. Extrinsic Motivations to Volunteer and Social Capital Formation », *Kyklos*, Volume 62, Numéro 3, pp. 359-370.

Dejours C. (1998), *Souffrance en France: la banalisation de l'injustice sociale*, Seuil, Paris.

Dejours C. (2003), *L'évaluation du travail à l'épreuve du réel: critique des fondements de l'évaluation*, Editions Quae, Paris.

Delattre E. et Dormont B. (2003), « Fixed fees and physician-induced demand: A panel data study on French physicians », *Health Economics*, Volume 12, Numéro 9, pp. 741-754.

Delfgaauw J. et Dur R. (2007), « Signaling and screening of worker Motivation », *Journal of Economic Behavior and Organization*, Volume 62, Numéro 4, pp. 605-624.

Delfgaauw J. et Dur R. (2008), « Incentives and workers' motivation in the public sector », *Economic Journal*, Volume 118, Numéro 525, pp. 171-91.

Delivering the NHS plan (2002), National Health Service, April.

Demsetz H. (1969), « Information and Efficiency : Another Viewpoint », *Journal of Law and Economics*, Volume 12, Numéro 1, pp. 1-22.

Desprès C., Guillaume S., et Couralet P. (2009), *Le refus de soins à l'égard des bénéficiaires de la Couverture maladie universelle complémentaire à Paris*, La documentation française, Paris.

Desrosières A. (2008), *Pour une sociologie historique de la quantification : L'Argument statistique I et II*, Presses de l'École des Mines de Paris, Paris.

Desrosières A. (2010), *La politique des grands nombres. Histoire de la raison statistique*, La Découverte, Paris.

Devetter F-X. et Jany-Catrice F. (2010) « L'invention d'un secteur et ses conséquences socio-économiques : objectifs et bilan des politiques de soutien aux services à la personne », *Politiques et Management Public*, Volume 27, Numéro 2, pp. 75-102.

Di Vittorio P., Minard M., et Gonon F. (2013), « Les virages du DSM : enjeux scientifiques, économiques et politiques », *Hermès, La Revue*, Volume 66, Numéro 2, pp. 85-92.

Dickinson A. (1989), « The detrimental effects of extrinsic reinforcement on "intrinsic motivation" », *The Behavior Analyst*, Volume 12, Numéro 1, pp. 1-15.

Diener E. (2000), « Subjective Well-Being. The Science of Happiness and a Proposal for a National Index », *Amercian Psychologist*, Volume 55, Numéro 1, pp. 34-43.

Domenighetti G. (1995), « Incertitude médicale : justification éthique de la demande induite », *Journal d'économie médicale*, Volume 13, Numéro 7-8, pp.353-362.

Domin J-P. (2005), « Le palmarès hospitalier ou la marche forcée vers le marché de la santé », *Mouvements*, Volume 3, Numéro 39-40, pp. 158-162.

Domin J-P (2006), « La démocratie sanitaire participe-t-elle à la construction d'un consommateur de soins ? », *Journal d'Économie Médicale*, Numéro 7, Volume 7/8, 427-438.

Domin J-P. (2013)a, *Une histoire économique de l'hôpital (XIXe-XXe siècles). Une analyse rétrospective du développement hospitalier, Tome II (1945-2009)*, CHSS/La Documentation française, Paris.

Domin J-P. (2013)b, « Quand l'hôpital se moque de la charité, son compromis fondateur est rompu », *Revue du MAUSS*, Volume 41, Numéro 1, pp. 35-46.

Domin J.-P. (2014), « Une mesure économique des changements en matière de politiques de socialisation de la santé (1890-1938) », *Économies et Sociétés*, Volume 48, Numéro 2, pp. 247-273.

Doran T., Fullwood C., Gravelle H., Reeves D., Kontopantelis E., Hiroeh U. et Roland M. (2006), « Pay for performance programs in family practices in the United Kingdom », *New England Journal of Medicine*, Volume 359, Numéro 1, pp. 274-284.

Doran T., Fullwood C., Reeves D., Gravelle H. et Roland M. (2008), « Exclusion of patients from pay-for-performance targets by English physicians », *New England Journal of Medicine*, Volume 359, Numéro 3, pp. 274-284.

Dourgnon P., Jusot F. et Fantin R., (2012), « Payer nuit gravement à la santé : une étude de l'impact du renoncement financier aux soins sur l'état de santé », *Économie Publique*, Volume 28-29, pp. 123-147.

Dranove D. (1988), « Demand inducement and the physician/patient relationship », *Economic Inquiry*, Volume 26, Numéro 2, pp. 251-298.

Dranove D. et White W. (1987), « Agency and the organization of health care delivery », *Inquiry*, Volume 24, Numéro 4, pp. 405-415.

Dreyfus M., Ruffat M., Viet V. et Voldman D. (2006), *Se protéger, être protégé: une histoire des assurances sociales en France*, Presses universitaires de Rennes, Rennes.

Dubar C. et Tripier P. (1998), *Sociologie des professions*, Armand Colin, Paris.

Duckett S., Daniels S., Kamp M., Stockwell A., Walker G. et Ward M. (2008), « Pay for performance in Australia: Queensland's new clinical practice improvement payment », *Journal of Health Services Research & Policy*, Volume 13, Numéro 3, pp. 174-177.

Duguet A-M. (2012), « Conflits d'intérêts et expertises dans le domaine de la santé : l'annulation par le Conseil d'État d'une recommandation de l'HAS », *Médecine & Droit*, Volume 112, pp. 10-15.

Dumesnil J. (2009), *Art médical et normalisation du soin*, Souffrance et Théorie, Paris.

Duval J. (2007), *Le mythe du "trou de la Sécu"*, Raison d'agir, Paris.

Eber N. (2006), *Le dilemme du prisonnier*, Repères, La Découverte, Paris.

Eggleston K. (2000), « Risk selection and optimal health insurance-provider payment systems », *Journal of Risk and Insurance*, Volumue 67, Numéro 2, pp. 173-19.

Eggleston K. (2005), « Multitasking and mixed systems for provider payment », *Journal of Health Economics*, Volume 4, Numéro 1, pp. 211-223.

Eijkenaar F. (2012), « Pay for performance in health care : An international overview of initiatives », *Medical Care Research and Review*, Volume 69, Numéro 3, pp. 251-276.

Eijkenaar F. (2013), « Key issues in the design of pay for performance programs », *The European Journal of Health Economics*, Volume 14, Numéro 1, pp. 117-131.

Eijkennar F., Emmert M., Scheppach M. et Schöffski O. (2013), « Effects of pay for performance in health care : A systematic review of systematic reviews », *Health Policy*, Volume 110, Numéro 2/3, pp. 115-130.

Eisenberg J. (1986), *Doctors' decisions and the cost of medical care*, Health administration press, Chicago.

Ellis R. et McGuire T. (1990), « Optimal payment systems for health services », *Journal of Health Economics*, Volume 9, Numéro 4, pp. 375-396.

Ellingsen T. et Johannesson M. (2007), « Paying respect », *Journal of Economic Perspective*, Volume 21, Numéro 4, pp. 135-49.

Ellingsen T. et Johannesson M. (2008), « Pride and prejudice: The human side of incentive theory », *American Economic Review*, Volume 98, Numéro 3, pp. 990-1008.

Elster J. (1985), *Sour grapes: Studies in the subversion of rationality*, Cambridge University Press, Cambridge.

Elster J. (1998), « Emotions and economic theory », *Journal of Economic literature*, Volume 36, Numéro 1, pp. 47-74.

Elster J. (2009), *Le désintéressement. Traité critique de l'homme économique I*, Édition du seuil, Paris.

Emmert M., Eijkenaar F., Kemter H., Esslinger A., Schöffski O. (2012), « Economic evaluation of pay for performance in health care : a systematic review », *European Journal Health Economics*, Volume 13, Numéro 6, pp. 755-767.

Emons W. (1997), « Credence goods and fraudulent experts », *Rand Journal of Economics*, Volume 28, Numéro 1, pp. 107-119.

Epstein A. (2006), « Paying for performance in the United States and abroad », *New England Journal of Medicine*, Volume 355, Numéro 4, pp. 406-408.

Etile F. Milcent C. (2006), « Income-related reporting heterogeneity in self-assessed health: evidence from France », *Health economics*, Volume 15, Numéro 9, pp. 965-981.

Evans R. (1974), « Supplier-induced demand : some empirical evidence and implications », dans Perlman M. (Ed.), *The economics of health and medical care*, Macmillan, pp. 162-173.

Eymard-Duvernay F. (1989), « Conventions de qualité et formes de coordination », *Revue Économique*, Numéro 40, Volume 2, p. 329-359.

Eymard-Duvernay F. (1993), « La négociation de la qualité », *Économie Rurale*, Volume 217, Numéro 1, pp. 12-17.

Eymard-Duvernay F. (2002), « Les qualifications des biens », *Sociologie du Travail*, Volume 44, Numéro 2, pp. 267-272.

Eymard-Duvernay F. (2004), *Économie politique de l'entrerpsie*, Repères, Paris.

Eymard-Duvernay F. (Ed) (2006), *L'économie des conventions, méthodes et résultats*, La Découverte, Paris.

Falk A. et Kosfeld M. (2006), « The hidden costs of control », *American Economic Review*, Volume 96, Numéro 5, pp. 1611-1630.

Fargeon V. (2009), *Introduction à l'économie de la santé*, Presses Universitaire de Grenoble, Grenoble.

Faure O. (1984), « La médecine gratuite au XIXème siècle : de la charité l'assistance », *Histoire, économie et société*, Volume 3, Numéro 3-4, pp. 593-608.

Favereau O. (1986), « La formalisation du rôle des conventions dans l'allocation des ressources », dans Salais R. et Thévenot L. (Eds), *Le travail. Marchés, règles, conventions*, Economica, Paris, pp. 249-268.

Favereau O. (1989), « Marchés internes, marchés externes», *Revue économique*, Volume 40, Numéro 2, pp. 273-328.

Favereau O. (1995), « Convention et régulation », dans Boyer R. et Saillard Y. (Eds), *Théorie de la régulation. L'état des savoirs*, La Découverte, Paris, p. 511-520.

Favereau O. (1999), « Salaire, emploi et économie des conventions », *Cahier d'Économie Politique*, Volume 34, Numéro 34, pp. 163-194.

Favereau O. (2004), « Trois considérations critiques sur les rapports entre l'éthique et la théorie économique » dans M. Canto Sperber (Ed), *Éthiques d'aujourd'hui*, PUF, Paris, pp. 35-36.

Favereau O. (2005), « Quand les parallèles se rencontrent : Keynes et Wittgenstein, l'économie et la philosophie », *Revue de Métaphysique et de Morale*, Volume 3, Numéro 47, pp. 403-427.

Favereau O. (2010), « La place du marché » dans Hatchuel A., Favereau O. et Aggeri F. (Eds), *L'activité marchande sans le marché ?*, Colloque de Cerisy, Presses des Mines, Paris., p. 111-131.

Fehr E. et Fischbacher U. (2003), « The nature of human altruism », *Nature*, Volume 426, Numéro 6960, pp. 785–791.

Fehr E. et Gächter S. (2000), « Fairness and retaliation: The economics of reciprocity », *Journal of Economic Perspectives*, Volume 14, Numéro 3, pp. 159-181.

Fenouillet F. (2012), *Les théories de la motivation*, Dunod, Paris.

Fiorentini G., Iezzi E., Lippi Bruni M. et Ugolini C. (2011), « Incentives in primary care and their impact on potentially avoidable hospital admissions », *European Journal of Health Economics*, Volume 12, Numéro 1, pp. 297-309.

Fiorillo D. (2011), « Do monetary rewards crowd out the intrinsic motivation of volunteers? Some empirical evidence for Italian volunteers », *Annals of Public and Cooperative Economics*, Volume 82, Numéro 2, pp. 139-165.

Flora S. (1990), « Undermining intrinsic interest from the standpoint of a behaviorist », *Psychological Record*, Volume 40, Numéro 3, pp.323-346.

Folbre N. et Goodin R. (2004), « Revealing altruism », *Review of Social Economy*, Volume 62, Numéro 1, pp. 1–25.

Forest V. (2008), « Performance-related pay and work motivation: theoretical and empirical perspectives for the French civil service », *International Review of Administrative Sciences*, Volume 74, Numéro 2, pp. 325-339.

Forest V. (2008), *Nouvelle gestion publique et comportements au travail : l'exemple des médecins de la fonction publique hospitalière*, Thèse de doctorat de science de gestion, Université Lumière Lyon 2.

Fouquet A., (2010), « L'usage des statistiques : de l'aide à la décision à l'évaluation des politiques publiques », *Revue Française des Affaires Sociales*, Volume 1, Numéro 1/2, pp. 307-322.

Franc C., Renard E., Le Vaillant M. et Pelletier-Fleury N. (2009), « La perception par les pédiatres des contraintes économiques. Quelle incitation dans la prise en charge de l'obésité infantile ? », *Revue Économique*, Volume 60, Numéro 2, pp. 507-520.

Frech H. (1996), *Competition and monopoly in medical care*, AEI Press, Washington.

Frey B. (1993), « Shirking or work moral ? The impact of regulation », *European Economic Review*, Volume 37, Numéro 8, pp. 1523-1532.

Frey B. (1997a), « A Constitution for Knaves Crowds Out Civic Virtues », *The Economic Journal*, Volume 104, Numéro 443, pp. 1043-1053.

Frey B. (1997b), « From the Price to the Crowding Effect », Sw*iss Journal of Economics and Statistics*, Volume 133, Numéro 2, pp. 325-350.

Frey B. (1997c), « On the relationship between intrinsic and extrinsic work motivation », *International Journal of Industrial Organization*, Volume 15, Numéro 4, pp. 427-439.

Frey B. (1997d), *Not just for the money*, Edward Elgard Publishing, Chelthenham.

Frey B. et Benz M. (2002), « From Imperialism to Inspiration : Asurvey of Economics and Psychology », *Working Paper Series*, No. 118.

Frey B. et Götte L. (1999), « Does Pay Motivate Volunteers? » *Unpublished Manuscript*, Institute for Empirical Economic Research. University of Zurich.

Frey B. and Jegen R. (2001), « Motivation Crowding Theory », *Journal of Economic Surveys*, Volume 15, Numéro 5, pp. 589-611.

Frey B. et Oberholzer-Gee F. (1997), « The cost of price incentives : An empirical Analysis of Motivation Crowding-Out », *American Economic Review*, Volume 87, Numéro 4, pp. 746-755.

Frey B., Oberholzer-Gee F. et Eichenberger R. (1997), « The Old Lady Visits Your Backyard : A Tale of Morals and Markets », *Journal of Political Economy*, Volume 104, Numéro 6, pp. 1297-1313.

Frey B. et Osterloh M. (2005), « Yes, Managers Should Be Paid Like Bureaucrats », *Journal of Management Inquiry*, Volume 14, Numéro 1, pp. 96-111.

Friot B. (2012)a, *Puissances du salariat*, La dispute, Paris.

Friot B. (2012)b, *L'enjeu du salaire*, La dispute, Paris.

Frohlich N. et Oppenheimer J. (1998), « Some consequences of e-mail vs. face-to-face communication in experiment », *Journal of Economic Behavior & Organization*, Volume 35, Numéro 3, pp. 389-403.

Fuchs V. (1978), « The supply of surgeons and the demand of operations », *Journal of Human Ressources*, Volume 13, Numéro 1, pp. 35-36.

Gadreau M. (1978), *Une mesure de la santé*, Sirey, Paris.

Gadreau M. (2009), « Économie et éthique en santé : Du calcul économique aux jugements de valeurs », Revue de philosophie économique, Volume 10, Numéro 1, pp. 3-13.

Gadrey J., (1991), « Le service n'est pas un produit : quelques implications pour l'analyse économique et pour la gestion », *Politiques et Management Public*, Volume 9, Numéro 1, pp. 1-24.

Gadrey J. (2008), « La crise écologique exige une révolution de l'économie des services. Développement durable et territoires », *Économie, Géographie, Politique, Droit, Sociologie*, http://developpementdurable.revues.org/6423, consulté le 15/10/14.

Gallois F. (2012), *Une approche régulationniste des mutations de la configuration institutionnelle française des services à la personne*, Thèse de doctorat en sciences économiques, Université de Reims Champagne-Ardenne.

Galvin R. (2006), « Pay-for-Performance : too much of a good thing ? A conversation with Martin Roland », *Health Affairs*, Volume 25, Numéro 5, pp. 412-419.

Galvis-Narinos F. et Montélimard A. (2009), « Le système de santé des États-Unis », *Pratiques et Organisation des Soins*, Volume 40, Numéro 4, pp. 309-315.

Gaynor M. (1994), « Issues in the industrial organization of the market for physician services », *Journal of Economics & Management Strategy*, Volume 3, Numéro 1, pp. 211-255.

Gene-Badia J., Escaramis-Badiano G., Sans-morales M., Sampietro-colom L., Aguadi-menguy F., Cabezas-Pena C., Gallo de Puelles P. (2007), « Impact of economic incentives on quality of professional life and on end-user satisfaction in primary care », *Health Policy*, Volume 80, Numéro 1, pp. 2-10.

Gilardone M. (2009), « Inégalités de genre et approche par les capabilités: quelle mise en dialogue chez Sen? », *Revue Tiers Monde*, Volume 198, Numéro 2, pp. 357-371.

Gillon R. (1994), « Medical ethics: four principles plus attention to scope », *British Medical Journal*, Numéro 309, Volume 6948, pp. 184-190.

Gintis H., Bowles S., Boyd R. et Fehr E. (2003), « Explaining altruistic behaviour in humans », *Evolution and Human Behaviour*, Volume 24, Numéro 3, pp. 153-172.

Giuffrida A., Gosden T., Forland F., Kristiansen I., Sergison M., Leese B., Pedersen L. et Sutton M. (1999), *Target payments in primary care: effects on professional practice and health care outcomes*, The Cochrane Collaboration.

Glannon W. et Ross L. (2002), « Are doctors altruistic ? », *Journal of Medical Ethics*, Volume 28, Numéro 2, pp. 68-69.

Gneezy U. et Rustichini A. (2000a), « Pay Enough or Don't Pay at All », *Quaterly Journal of Economics*, Volume 115, Numéro 3, pp. 791-810.

Gneezy U. et Rustichini A. (2000b), « A fine is a price », *Journal of Legal Studies*, Volume 29, Numéro 1, pp. 1-17.

Gonon F. (2013), « Quel avenir pour les classifications des maladies mentales ? Une synthèse des critiques anglo-saxonnes les plus récentes », *L'information psychiatrique*, Volume 89, Numéro 4, pp. 285-294

Gori R., Cassin B., et Laval C. (2009), *L'appel des appels. Pour une insurrection des consciences*, Mille et une nuits, Paris.

Grignon M., (2009), « Pourquoi les systèmes de santé sont-ils organisés différemment ? ». *Revue Économique*, Volume 60, Numéro 2, pp. 545-558.

Grignon M., Paris V., Polton D., Couffinhal A. et Pierrard, B. (2002), *L'influence des modes de rémunération des médecins sur l'efficience du système de soins*. Commission sur l'avenir des soins de santé au Canada.

Grinzberg E. (1953), « Health, medicine, and economics welfare », *Journal Mount Sinai Hospital*, Volume 19, pp. 734-743.

Grinzberg E. (1954), « What every economist should know about health and medicine », *American Economic Review*, Volume 44, Numéro 1, pp. 104-119.

Gross R. (2003), « Implementing health care reform in Israel: organizational response to perceived incentives », *Journal of Health Politics, Policy and Law*, Volume 28, Numéro 4, pp. 659-692.

Gross R., Elhaynay A., Friedman N. et Buetow, S. (2008), « Pay-for-performance programs in Israeli sick funds », *Journal of Health Organization and Management*, Volume 22, Numéro 1, pp. 23-35.

Grossieux P. (2003), *Principe de précaution et sécurité sanitaire*, Presses Universitaires d'Aix-Marseille, Marseille.

Halimi S. (2004), *Le grand bon en arrière*, Agone, Elément, Paris.

Hamilton W. (1932), « Personal Statement » dans, *Medical care for the American people*, University of Chicago Press, Chicago, pp. 189-200.

Harsanyi J. (1955), « Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility », *Journal of Political Economy*, Volume 63, Numéro 4, pp. 309-321.

Hassenteufel P. (1992), « Représentation des médecins et redéfinition de l'action publique: la régulation des dépenses de médecine ambulatoire en France, en RFA et en Autriche », *Politiques et Management Public*, Numéro 10, Volume 3, pp. 31-60.

Hassenteufel P. (1997), *Les médecins face à l'État. Une comparaison européenne*, Presses de Science Po, Paris.

Hassenteufel P. et Palier B. (2005), « Le trompe l'œil de la 'gouvernance' de l'assurance maladie », *Revue Française d'Administration Publique*, Volume 113, Numéro 1, pp. 13-27.

Hatzfeld H. (1963), *Le grand tournant de la médecine libérale*, Économie et Humanisme. Les Editions Ouvrière, Paris.

Hatzfeld H. (1971), *Du paupérisme à la sécurité sociale : essai sur les origines de la Sécurité sociale en France (1850-1940)*, Colin, Paris.

Hausman D. et McPherson M. (1993), « Taking ethics seriously: Economics and contemporary moral philosophy », *Journal of Economic Literature*, Volume 31, Numéro 2, pp. 671–731.

Heilbrunn B. (Ed) (2004), *La performance, une nouvelle idéologie ? Critique et enjeux*, La Découverte, Paris.

Herzberg F. (1971), *Le travail et la nature de l'homme*, EME, Paris

Heyes A. (2007), « The economics of vocation or why is a badly paid nurse a good nurse?" *Journal of Health Economics*, Volume 24, Numéro 3, pp. 561-569.

Hibou B. (2012), *La bureaucratisation du monde à l'ère néolibérale*, La Découverte, Paris.

Hilman A., Ripley K., Goldfarb N., Nuamah I., Weiner J. et Lusk E. (1998), « Physician financial incentives and feedback : failure to increase cancer screening in medical managed care », *American Journal of Public Health*, Volume 88, Numéro 11, pp. 1699-1701.

Hodgson G. (2009), « Towards an Alternative Economics of Health Care », *Health Economics, Policy and Law*, Volume 4, Numéro 1, pp. 99-114.

Hodgson G. (2013), *From pleasure machines to moral communities: an evolutionary economics without homo economicus*, University of Chicago Press, Chicago.

Hoerger T. et Howard L. (1995), « Search behavior and choice of physician in the market for prenatal care », *Medical Care*, Volume 33, Numéro 4, pp. 332-349.

Holmås T., Kjerstad E., Lurås H. et Straume O. (2010), « Does monetary punishment crowd out pro-social motivation? A natural experiment on hospital length of stay », *Journal of Economic Behavior & Organization*, Volume 75, Numéro 2, pp. 261-267.

Holmstrom B. et Milgrom P. (1991), « Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design », *Journal of Law, Economics, & Organization*, Volume 7, Numéro 1, pp. 24-52.

Hood C. (1995), « The 'New Public Management' in the 1980s: variations on a theme », *Accounting, Organization and Society*, Volume 20, Numéro 2-3, pp. 93-109.

Hood C. et Peters G. (2004), « The middle aging of new Public management: into the age of paradox? », *Journal of Public Administration Research and Theory*, Volume 14, Numéro 3, pp. 267-282.

Husson M. (2010), « Le partage de la valeur ajoutée en Europe », La Revue de l'Ires, Volume 1, pp. 47-91.

Ichino A. et Maggi G. (2000), « Work environnement and individual back-ground : Explaining regional shirking differentials in a large Italian firm », *Quarterly Journal of Economics*, Volume 115, Numéro 3, pp. 1057-1090.

IGAS (2002), *Evaluation du système d'information des professionnels de santé*, Rapport Numéro 2002-142.

Isaacs S. et Schroeder S. (2008), « California Dreamin' – State health care reform and the prospect for national change », *New England Journal of Medicine*, Volume 358, Numéro 15, pp. 1537-1540.

INSEE (2011), « Légère baisse du taux de chômage au quatrième trimestre 2010 », *Informations Rapides*, Numéro 60.

Institut of Medicine (2001), *Crossing the quality chasm*, Report brief.

Institut of Medicine (2006), *Rewarding provider performance : Aligning incentives in medicare*, Report brief.

IRDES (2006), *Faisabilité d'un système d'information public sur la médecine de ville*, Rapport Numéro 535.

Jack W. (2003), « Contracting for health services: an evaluation of recent reforms in Nicaragua », *Health Policy and Planning*, Volume 18, Numéro 2, pp. 195-204.

Jack W. (2005), « Purchasing health care services from providers with unknown altruism », *Journal of Health Economics*, Volume 24, Numéro 1, pp. 73–93.

James Jr H. (2005), « Why did you do that ? An economic examination of the effect of extrinsic compensation on intrinsic motivation and performance », *Journal of Economic Psychology*, Volume 26, Numéro 4, pp. 549-566.

Jany-Catrice F. (2010), « La construction sociale du 'secteur' des services à la personne : une banalisation programmée ? », *Sociologie du Travail*, Volume 52, Numéro 4, pp. 521-537.

Jany-Catrice F. (2012), *La performance totale : nouvel esprit du capitalisme ?*, Presses Universitaire du Septentrion, Villeneuve-d'Ascq.

Jany-Catrice F. (2012), « De l'évaluation des politiques publiques à la 'performance totale' », *Économie et Institutions*, n°18-19, 1er et 2e semestres, pp. 17-34.

Jany-Catrice F. et Méda D. (2013), « Les nouvelles mesures des performances économiques et du progrès social. Le risque de l'économicisme », *Revue du MAUSS*, Volume 41, Numéro 1, pp. 371-397.

Jaunait A. (2005), *Comment pense l'institution médicale ? Une analyse des codes français de déontologie médicale*, Editions Dalloz, Paris.

Jensen M. and Meckling W. (1976), « Theory of the Firm : Managerial Behavior, Agency Costs and Ownership Structure », *Journal of Financial Economics*, Volume 3, Numéro 4, pp. 163-261.

Journal of Health Politics, Policy and Law, Special : Kenneth Arrow and the Changing Economics of Health Care (2001), Volume 26, Numéro 5.

Journal Officiel de la République Française (1960), 13 mai 1960, Numéro 6, p. 4357.

Journal Officiel de la République Française (2000), Loi n°2000-1257 du 23 décembre 2000 de financement de la sécurité sociale pour 2001.

Journal Officiel de la République Française (2002), Loi n°2002-1487 du 20 décembre 2002 de financement de la sécurité sociale pour 2003.

Journal Officiel de la République Française (2009), Décision du 9 mars 2009 de l'Union nationale des caisses d'assurance maladie relative à la création d'un contrat type d'amélioration des pratiques à destination des médecins libéraux conventionnés, n°0093 du 231 avril 2009 page 6839 texte n°34.

Joyce R. (2007), *The evolution of morality*, MIT Press, Boston.

Kahn III C., Ault T., Isenstein H., Potetz L. et Van Gelder S. (2006), « Snapshot of hospital quality reporting and pay for performance under medicare », *Health Affairs*, Volume 25, Numéro 1, pp. 148-162.

Kahneman D. (2011), *Thinking, fast and slow*, Macmillan, New York.

Kambia-Chopin B., Perronnin M., Pierre A. et Rochereau T. (2008), « La complémentaire santé en France en 2006 : un accès qui reste inégalitaire », *Question d'économie de la santé*, Numéro 132.

Kaiser Family Foundation (2012), en ligne (kff.org).

Karpik L. (1989), « Le désintéressement », *Annales. Histoire, Sciences Sociales*, 44è année, Numéro 3, pp. 733-751.

Karpik L. (2003), « Les dispositifs de l'économie des singularités », dans Dubuisson-Quellier S et Neuville J-P. (Ed), *Juger pour échanger*. Editions de la MSH et de l'INRA, Paris, pp. 217-224.

Keel O. (2011), *La médecine des preuves. Une histoire de l'expérimentation thérapeutique par essais cliniques contrôlés*, Champ Libre, Québec.

Kelman S. (1992), « Adversary and Cooperationist Institutions for Conflict Resolution in Public Policymaking », *Journal of Policy Analysis and Management*, Volume 11, Numéro 2, pp. 178-206.

Kerleau M. (1998), « L'hétérogénéité des pratiques médicales, enjeu des politiques de maîtrise des dépenses de santé », *Sciences Sociales et Santé*, Volume 16, Numéro 4, pp. 5-32.

Kessel D. (1958), « Price discrimination in medicine », *Journal of Law and Economics*, Volume 1, Numéro 1, pp. 20-53.

Keynes J. (2002), *La pauvreté dans l'abondance*, Tel Gallimard, Paris.

Khalil E. (2004), « What is altruism ? », *Journal of Economic Psychology*, Volume 25, Numéro 1, pp. 97–123.

Kouides R., Bennett M., Lewis B., Cappuccio J., Barker W. et LaForce M. (1998), « Performance-based physician reimbursement and influenza immunization rates in the elderly: the primary-care physicians of Monroe County », *American Journal of Preventive Medicine*, Volume 14, Numéro 2, pp. 89-95.

Kravitz R. et Greenfield S. (1995), « Variations in resources utilization among medical specialities and systems of care », *Annual Review of Public Health*, Volume 16, Numéro 1, pp. 431-445.

Kreps D. (1997), « Intrinsic Motivation and Extrinsic Incentives », *American Economic Review*, Volume 87, Numéro 2, pp. 359-364.

Kuhmerker K. et Hartman T. (2007), *Pay for performance in state medicaid programs. À survey of state medicaid directors and programs*, IRPO, http://www.amcp.org/WorkArea/DownloadAsset.aspx?id=11786, consulté le 15/10/14.

Kunreuther H. et Easterling D. (1990), « Are risk-benefit tradeoffs possible in siting hazardous facilities? », *American Economic Review*, Volume 80, Numéro 2, pp. 252-256.

Labrousse A., (2010). « Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d'un outil de preuve et de gouvernement », *Revue de la Régulation*, http://regulation.revues.org/7818, consulté le 15/10/14.

Laguardia J. et Ryan R. (2000), « Buts Personnels, Besoins Psychologiques Fondamentaux et Bien Être : Théorie de l'Auto Détermination et Applications », *Revue Québécoise de psychologie*, Volume 21, Numéro 2, pp. 281-304.

Laude A. (1998), « La force juridique des références médicales opposables », *Médecine & Droit*, Volume 28, pp. 1-3.

Laude A. (2005), « Les médecins et les malades face au droit », *Revue Française d'Administration Publique*, Volume 1, Numéro 113, pp. 113-120.

Laude A. (2009), « Dix ans de droit de la santé », *Les Tribunes de la Santé*, Volume 4, Numéro 25, pp. 51-62.

Laurent C., Baudry J., Berriet-Solliec M., Kirsch M., Perraud D., Tinel B., Trouvé A., Allsopp N., Bonnafous P., Burel F., Carneiro M., Giraud C., Labarthe P., Matose F. et Ricroch A. (2009), « Pourquoi s'intéresser à la notion d'"evidence-based policy"? », *Revue Tiers Monde*, Volume 4, Numéro 1, pp. 853-873.

Lazear E. (2000a), « Performance Pay and Productivity », *American Economic Review*, Volume 90, Numéro 5, pp. 1346-1361.

Lazear E. (2000b), « The Power of Incentives », *American Economic Review*, Volume 90, Numéro 2, pp. 410-414.

Lazear E. (2007), « Personnel Economics : The Economicst's View of Human Resources », *Journal of Economic Perspectives*, Volume 21, Numéro 4, pp. 91-114.

Le Pape A. et Sermet C. (1998), « Les références médicales opposables sur le médicament: bilan de trois années d'application », *Questions d'Économie de la Santé*, Volume 14, pp. 1-6.

Lee T.-T., Cheng S.-H., Chen C.-C. et Lai, M.-S. (2010), « A pay-for-performance program for diabetes care in Taiwan: A preliminary assessment », *American Journal of Managed Care*, Volume 16, Numéro 1, pp. 65-69.

Léonard C. (2010), *Le parcours du patient dans les soins de santé. Refonder la solidarité grâce à la 'responsabilisation capacitante'*, Thèse présentée en vue de l'obtention du titre de Docteur en Sciences Médicales, Université Catholique de Louvain.

Léonard C., Stordeur S. et Roberfroid D. (2009), « Association between physician density and health care consumption : a systematic review of the evidence », Health Policy, Volume 91, Numéro 2, pp. 121-134.

Lepper M. et Green D. (1978), *The hidden cost of reward: New perspectives on the psychology of human motivation*, Erlbaum, Hillsdale, New Jersey.

Lepper M., Greene D. et Nisbett R. (1973), « Undermining children's intrinsic interest with extrinsic reward: A test of the overjustification hypothesis », *Journal of Personality and Social Psychology*, Volume 28, Numéro 1, pp. 129-137.

Levaggi R. et Rochaix L. (2003), « Optimal payment schemes for physicians », *Portuguese Economic Journal*, Volume 2, Numéro 2, pp. 87-107.

Levy-Garboua L., Meidinger C. et Rapoport B. (2006), « The formation of social preferences: some lessons from psychology and biology », dans *Handbook of economics of giving, reciprocity and altruism*, Kolm S. et Mercier-Ythier J. (Eds), North Holland, Amsterdam, pp. 545–613.

Levy Leboyer C. (1994), *La crise des motivations*, PUF, Paris

Levy Leboyer C. (2006), *La motivation au travail. Modèles et stratégies*, Edition d'organisation, Paris

Liebenstein H. (1969), « Organizational or frictional equilibria. X-efficiency and the rate of innovation », *The Quarterly Journal of Economics*, Volume 83, Numéro 4, pp. 600-623.

Liebenstein H. (1982), « The prisoners' dilemma in the invisible hand : An analysis of intrafirm productivity », *American Economic Review*, Volume 72, Numéro 2, pp. 92-97.

Lievaut J. (2010), *Le 'maillon faible' de la régulation des dépenses de santé en France : les comportements inattendus des médecins libéraux. Quatre approches micro économétriques longitudinales*, Thèse pour le doctorat de sciences économiques, Université Paris Ouest Nanterre La Défense.

Lievaut, J. (2012), « Les logiques comportementales des médecins. Une approche statistique à partir de leurs pratiques tarifaires », *Journal de gestion et d'économie médicales*, Volume 30, Numéro 3, pp. 175-190.

Lindenauer P., Remus D., Roman S., Rothberg M., Benjamin E., Ma A. et Bratzler D. (2007), « Public reporting and pay for performance in hospital quality improvement », *New England Journal of Medicine*, Volume 356, Numéro 1, pp. 486-496.

Lipsey, R. et Lancaster K. (1956), « The general theory of second best », *Review of Economic Studies*, Volume 24, Numéro 1, pp. 11-32.

Liu T. et Ma C. (2013), « Health insurance, treatment plan, and delegation to altruist physician », *Journal of Economic Behaviour and Organization*, Volume 86, Numéro 1, pp. 79–96.

Logie D., Rowson M. et Ndagije F. (2008), « Innovations in Rwanda's health system: looking to the future », *The Lancet*, Volume 372, Numéro 9634, pp. 256-261.

Loiseau V.P. (2005), *La maitrise des dépenses de santé confrontée à la responsabilité médicale aux États-Unis et en France*, Presses Universitaires d'Aix-Marseille, Marseille.

Lordon F. (2003), *Et la vertu sauvera le monde… Après la débâcle financière*, le salut par l'« éthique », Raisons d'agir: Paris.

Lordon F. (2009), *La crise de trop : reconstruction d'un monde failli*. Fayard, Paris.

Lordon F. (2013), *La société des affects. Pour un structuralisme des passions*, Seuil, Paris.

Löwy I., (1998), « Essais cliniques des thérapies nouvelles : une approche historique », *Médecine/Sciences*, Volume 14, Numéro 1, pp. 122-127.

Lu M. (1999), « Separating the 'true effect' from 'gaming' in incentive-based contracts in health care », *Journal of Economics and Management Strategy*, Volume 8, Numéro 3, pp. 383-432.

Lundback M. (1998), *Quality and efficiency in health care production*, CEFOS Report 10, Goteborg University, Goteborg.

Ma C. (1994), « Health care payment systems : cost and quality incentives », *Journal of Economics & Management Strategy*, Volume 3, Volume 1, pp. 93-112.

Ma C. (2007), « Altruism and incentives in public and private health care », dans Suvanto A. et Vartiainen H. (Eds), *Finance and Incentives of the Health Care System*, Government Institute for Economic Research, Helsinki, pp. 79–104.

Ma C. et McGuire T. (1997), « Optimal health insurrance and provider payment », *American Economic Review*, Volume 87, Numéro 4, pp. 685-704.

Mahieu F-R. (1998), « Altruisme et ingérence », dans *Altruisme. Analyses économiques*, Mahieu F-R. et Rapoport H. (Eds), Economica, pp. 113-138.

Mannion R. et Davies H. (2008), « Payment for performance in health care », *British Medical Journal*, Volume 336, Numéro 7639, pp. 306-308.

Marciano A. (1998), « Éthique et régulation des dépenses de santé. Le cas de la relation médecin-patient », Document de travail GREQAM n°98C16, Université d'Aix-Marseille II et III.

Marciano A. (2005), « Éthique médicale et comportement du patient : les conséquences paradoxales de l'altruisme du patient », dans Batifoulier P. et Gadreau M.. (Eds), *Éthique médicale et politique de santé*, Economica, Paris, pp. 95-108.

Marks H. (1999), *La médecine des preuves. Histoire et anthropologie des essais cliniques*, Institut Edition Synthelabo, Le Plessis Robinson.

Mas A. et Moretti E. (2009), « Peers at work », *American Economic Review*, Volume 99, Numéro 1, pp. 112-145.

Mas B., Pierru F., Smolski N., Torrielli R. (Eds) (2011), *L'hôpital en réanimation*, Editions du croquant, Paris.

Mascret C. (2008), « Les données de la science face à leur normalisation par les autorités sanitaires », *Médecine & Droit*, Volume 93, pp. 161-171.

Maslow A. (1943), « Theory of human motivation», *Psychological Review*, Volume 80, Numéro 4, pp. 370-396.

Mayo E. (1945), *The social problems of an industrial civilization*, Harvard University Press, Cambridge.

McCabe K., Rigdon M. et Smith V. (2003), « Positive reciprocity and intentions in trust games », *Journal of Economic Behavior and Organization*, Volume 52, Numéro 2, pp. 267-275.

McDonald R. et Roland M. (2009), « Pay for performance in primary care in England and California: comparison of unintended consequences », *Annals of Family Medicine*, Volume 7, Numéro 2, pp. 121-127.

McDonald R., Harrison S., Chekland K., Campbell S. et Roland M. (2007), « Impact of financial incentives on clinical autonomy and internal motivation in primary care: ethnographic study », *British Medical Journal*, Volume 334, Numéro 7608, pp. 1357-1359.

McGovern M., Boroujerdi M., Taylor M., Taylor M., Williams D., Hannaford P., Lefevre K. et Simpson C. (2008), « The effect of the UK incentive-based contract on the

management of patients with coronary heart disease in primary care », *Family Practice*, Volume 25, Numéro 1, pp. 33-9.

McGrégor D. (1969), *La dimension humaine de l'entreprise*, Gauthier-Villars, Paris.

McGuire T. (1983), « Patient's trust and the quality of physicians », *Economic Inquiry*, Volume 21, Numéro 2, pp. 203-222.

McGuire T. (2000), « Physician agency », dans A. Culyer and J. Newhouse, (Eds), *Handbook of health economics*, Volume 1A, Chapitre 9, North Holland, pp 461-536.

McGuire T. et Pauly M. (1991), « Physician response to fee changes with multiple payers », *Journal of Health Economics*, Volume 10, Numéro 4, 385-410.

McMaster R. (2007), « On the need for a heterodox health economics », *Post-Autistic Economics Review*, Volume 41, Numéro 1, pp. 9–22.

Mellström C. et Johannesson M. (2008), « Crowding out in blood donation: was Titmuss right? », *Journal of the European Economic Association*, Volume 6, Numéro 4, pp. 845-863.

Milcent C. et Rochut J. (2009), « Tarification hospitalière et pratique médicale », *Revue Économique*, Volume 60, Numéro 2, pp. 489-506.

Milgrom P. et Roberts J. (1997), *Économie, organisation et management*, De Boeck, Bruxelles.

Monod J. (1970), *Le hasard et la nécessité : essai sur la philosophie naturelle de la biologie moderne*, Seuil, Paris.

Mooney G., Ryan M. (1993), « Agency in health care: getting beyond first principles », *Journal of Health Economics*, Volume 12, Numéro 2, pp. 125-135.

Mougeot M., (2000), « La tarification hospitalière : de l'enveloppe globale à la concurrence par comparaison », *Annales d'Économie et de Statistique*, Volume 58, pp. 195-213.

Muller P. (2000), « L'analyse cognitive des politiques publiques : vers une sociologie politique de l'action publique », *Revue Française de Sciences Politiques*, Volume 50, Numéro 2, pp. 189-207.

Murdock K. (2002), « Intrinsic motivation and optimal incentive contracts », *Rand Journal of Economics*, Volume 33, Numéro 4, pp. 650-671.

Mushkin S. (1958), « Towards a Definition of Health Economics », *Public Health Reports*, Volume 73, Numéro 9, pp. 785-93.

Nahra T., Reiter K., Hirth R., Shermer J. et Wheeler J. (2006), « Cost-effectiveness of hospital pay-for-performance incentives », *Medical Care Research Review*, Volume 63, Numéro 1, pp. 49-72.

Newhouse J. (1970), Toward a theory of nonprofit institutions : an economic model of a hospital », *American Economic Review*, Volume 60, Numéro 1, pp. 64-74.

Newhouse J. (1996), « Reimbursing health plans and health providers : efficiency in production versus selection », *Journal of Economic Literature*, Volume 34, Numéro 3, pp. 1236-1263.

Newhouse J. (2002), *Pricing the priceless : a health care conundrum*, MIT Press, Boston.

NHS (2012), *2013/14 general medical services (GMS) contract quality and outcomes framework. Guidance for GMS contract 2013/14*, NHS England.

Nix K. (2013), « What Obamacare's pay for performance programs mean for health care quality », *Backgrounder*, Numéro 2856, pp. 1-9.

Noether M. (1986), « The Growing Supply of Physicians : Has the Market Become More Competitive ? », *Journal of Labor Economics*, Volume 4, Numéro 4, pp. 503-537.

Norton E. (1992), « Incentive regulation of nursing homes », *Journal of Health Economics*, Volume 11, Numéro 2, pp. 105-128.

Ogien A. (1995), *L'esprit gestionnaire*, Éditions de l'EHESS, Paris.

Ogien A. (2013), *Désacraliser le chiffre dans l'évaluation du secteur public*,Editions Quae, Paris.

Organisation mondiale de la Santé (1946), Préambule de l'Organisation mondiale de la Santé, adopté par la Conférence internationale sur la Santé, New York, 19-22 juin 1946.

OMS (2003), *Adherence to long-term therapies : evidence for action*, http://www.who.int/chp/knowledge/publications/adherence_report/en/, consulté le 15/10/14.

OCDE (2013), *Panorama de la santé 2013*, http://www.oecd.org/fr/sante/systemes-sante/panorama-de-la-sante.htm, consulté le 15/10/14.

Orléan A. (2003), « Réflexion sur les fondements institutionnels de l'objectivité marchande », *Cahier d'Économie Politique*, Volume 1, Numéro 44, pp. 181-196.

Orléan A. (Ed) (2004), *Analyse économique des conventions*, PUF, Paris.

Osborne D. et Gaebler T. (1992), *Reinventing government: how the entrepreneurial spirit is transforming the public sector?*, Addison-Wesley, Reading.

Palier B. et Bonoli G., (1999), « Phénomènes de path dependence et réformes des systèmes de protection sociale », *Revue Française de Science Politique*, Volume 49, Numéro 3, pp. 399-420.

Palpacuer F., Vercher C., Tessier N. et Bourdon I. (2010), « Sous l'épée de Damoclès : l'évaluation des cadres, entre performance individuelle et esprit d'équipe », Management & Avenir, Volume 5, Numéro 35, p. 15-32.

Patient protection and affordable care act (2010), https://democrats.senate.gov/pdfs/reform/patient-protection-affordable-care-act-as-passed.pdf, consulté le 15/10/14.

Pauly M. (1968), « The economics of moral hazard. Comment », *American Economic Review*, Volume 58, Numéro 3, pp. 531-537.

Pauly M. (1978). « Is medical care different ? », dans Greenberg W., Germantown, MD (Eds), *Competition in the health care sector : past, present, and future*. Aspen Systems, Aspen.

Pauly M. (1980), Doctor*s and their workshops : Economic models of physician behavior*, University of Chicago Press, Chicago.

Pauly M. (2000), « Insurance reimbursement », dans Cluyer A. et Newhouse J. (Eds), *Handbook of Health Economics*, Chapitre 10, Elsevier, Amsterdam.

Pauly M. et Redisch M. (1973), « The not-for-profit hospital as a physicians' cooperative », American Economic Review, Volume 63, Numéro 1, pp. 87-99.

Pauly. M., Satterthwaite M. (1981), « The pricing of primary care physician' services: a test of the role of consumer information », *Bell Journal of Economics*, Volume 12, Numéro 2, pp. 488-506.

Peabody J., Shimkhada R., Quimbo S., Florentino J., Bacate M. McCulloch C. et Solon O. (2011), « Financial incentives and measurement improved physicians' quality of care in the Philippines », *Health Affairs*, Volume 30, Numéro 4, pp. 773-781.

Peckham S. et Wallace A. (2010), « Pay for performance in primary care : what have we learnt ? », *Quality in Primary Care*, Volume 18, Numéro 2, pp. 111-116.

Penneau M. et Penneau J. (1998), « Recommandations professionnelles et responsabilité médicale », *Médecine & Droit*, Volume 28, pp. 4-6.

Performance Publique, (2014), *Rapport sur l'évolution de l'économie nationale et sur les orientations des fiances publique. Tome 2*, http://www.performance-publique.budget.gouv.fr/sites/performance_publique/files/farandole/ressources/medias/documents/ressources/2014/DOFP/DOFP_TOME_2.pdf, consulté le 15/10/14.

Perronnin M., Pierre A., et Rochereau T. (2011), « La complémentaire santé en France en 2008 : une large diffusion mais des inégalités d'accès », *Question d'économie de la santé*, IRDES, Numéro 161.

Petit E. (2012), « L'apport de la psychologie sociale à l'analyse économique », *Revue d'Économie Politique*, Volume 121, Numéro 6, pp. 797-837.

Pierru F. (2007), *Hippocrate malade de ses réformes*, Edition du croquant, savoir/agir, Paris.

Pierru F. (2011), « Budgétiser l'assurance maladie. Heurs et malheurs d'un instrument de maîtrise des dépenses publiques : l'enveloppe globale (1976-2010) », dans Bezes P. et Siné A. (Eds), *Gouverner (par) les finances publiques*, pp. 393-449.

Pollit C. (2001), « Clarifying convergence - striking similarities and durable differences in public management reform », *Public Management Review*, Volume 4, Numéro 1, pp. 471-492.

Polton D. (2010), « Accroître la qualité et l'efficience du système de santé en s'appuyant sur les soins primaires, la démarche de l'assurance maladie en France », *Revue Française des Affaires Sociales*, Volume 3, Numéro 3, pp. 97-101.

Polton D. et Aubert J-M, « Le contrat d'amélioration des pratiques individuelles. Aboutissement ou nouveau départ pour la gestion des soins ? », *Collège des Économistes de la Santé, La Lettre du Collège*, Volume 21 Numéro 3, pp. 5-7.

Postel N. et Sobel R. (2009), « Institutionalism as the way of unification of the heterodox theories », *Journal of Philosophical Economy*, Volume 3, Numéro 1, pp. 47-77.

Pratt J. (1964), « Risk aversion in the small and in the large », *Econometrica*, Volume 32, Numéro 1/2, pp. 122-136.

Prendergast C. (2007), « The motivation and bias of bureaucrats », *American Economic Review*, Volume 97, Numéro 1, pp. 180-196.

Prescrire (2013a), « Rémunération sur résultats : étendue à l'ensemble des médecins libéraux en France », Volume 33, Numéro 351, pp. 63-67.

Prescrire (2013b), « Rémunération sur résultats : objectifs de soins et de coûts (n° 1 à n° 5) », Numéro 33, Volume 352, p. 147-149.

Prescrire (2013c), « Rémunération sur résultats : objectifs de soins et de coûts (n° 6 à n° 11) », Volume 33, Numéro 353, p. 222-226.

Prescrire (2013d), « Rémunération sur résultats : objectifs de soins et de coûts (n° 12 et n° 13) », Volume 33, Numéro 355, p. 384-385.

Prescrire (2013e), « Rémunération sur résultats : objectifs de soins et de coûts (n° 14 à n° 24) », Volume 33, Numéro 356, p. 456-463.

Prescrire (2013f), « Rémunération sur résultats : trop peu d'objectifs 2011 en faveur de l'amélioration des soins », Volume 33, Numéro 357, p. 539-542.

Prescrire (2013g), « Rémunération conventionnelle sur résultats : premiers chiffres », Volume 33, Numéro 357, p. 541.

Rabier C. (2010), « Une révolution médicale ? Dynamiques des professions de santé entre révolution et empire », *Annales historiques de la Révolution française*, Volume 359, Numéro 1, p. 141-159.

Ramaux C. (2010), « Note de lecture. Pierre Dardot et Christian Laval, La nouvelle raison du monde. Essai sur la société néolibérale, (2009) et Luc Boltanski, De la critique. Précis de sociologie de l'émancipation, (2009). Aporie de la critique funèbre », *Revue de la Régulation*, Volume 7, http://regulation.revues.org/7722, consulté le 15/10/14.

Ramaux C. (2012), *L'État social. Pour sortir du chaos néolibéral*, Fayard, Mille et une nuits, Paris.

Rauly A. (2013), « Dispositifs de rémunération de la télémédecine : de la diversité des propositions de contrats à la singularité du système de santé français », *Journal d'Économie et de Gestion Médicales*, Volume 31, Numéro 7/9, pp 473-486.

Rawls J. (1971), *Théorie de la justice*, Seuil, Paris. Traduction française 1987.

Reinhardt U. (1989), « Economists in healthcare : saviors or elephants in a porcelain shop ? », *American Economic Review*, Volume 79, Numéro 2, pp. 337-342.

Reis H., Sheldon K., Gable S., Roscoe J. et Ryan R. (2000), « Daily Well-Being : The Role of Autonomy, Competence, and Relatedness », *Personality and Social Psychologie Bulletin*, Volume 26, Numéro 4, pp. 419-435.

Rice T. (1998), *The economics of health reconsidered*, Health administration press, Chicago.

Richardson J. (1981), « The inducement hypothesis: That doctors generate demand for their own services », dans, *Health, Economics, and Health Economics*, Van der Gaag J. et Perlman M. (Eds), North Holland, Amsterdam, pp. 189–214.

Ricoeur P. (1996), « Les trois niveaux du jugement médical », *Esprit*, Volume 12, pp. 21-33.

Rizzo J. et Zeckhauser R. (2003), « Reference incomes, loss aversion and physician behaviour », *Review of Economics and Statistics*, Volume 85, Numéro 4, pp. 909–922.

Robelet R. (1999), « Les médecins placés sous observation. Mobilisations autour du développement de l'évaluation médicale en France », *Politix*, Volume 12, Numéro 46, pp. 71-97.

Rochaix L. (1989), « Information asymmetry and search for the market for physicians' services », *Journal of Health Economics*, Volume 8, Numéro 1, pp. 53-84.

Rochaix L. (1997), « Asymétrie d'information et incertitude en santé : les apports de la théorie des contrats », *Économie et Prévision*, n°129, Numéro 3/4, pp. 11-24.

Rodwin V. (1997), « Managed care US : pysicians and patients react », *Filières et réseaux, Lettre du disease management à la Française*, pp. 431-445.

Rolland C. et Sicot F. (2012), « Les recommandations de bonne pratique en santé. Du savoir médical au pouvoir néo-managérial », *Gouvernement et Action Publique*, Volume 3, Numéro 3, pp. 53-75.

Rose-Ackerman S. (1996), « Altruism, nonprofits, and economic theory », *Journal of Economic Literature*, Volume 34, Numéro 2, pp. 701–728.

Ross S. (1973), « The economic theory of agency: the principal's problem », *American Economic Review*, Volume 63, Numéro 2, pp. 134-139.

Ross L., Greene D. et House P. (1977), « The 'false consensus effect' : An egocentric bias in social perception and attribution processes », *Journal of Experimental Social Psychology*, Volume 13, Numéro 3, pp. 279-301.

Rosenthal M., Frank R., Li Z., Epstein A. (2005), « Early experience with pay for performance : from concept to practice », *Journal of the American Medical Association*, Volume 294, Numéro 14, pp. 1788-1793.

Rosenthal M., Landon B., Normand S., Frank R. et Epstein A. (2006), « Pay for performance in commercial HMOs », *New England Journal of Medicine*, Volume 355, Numéro 18, pp. 1895-1902.

Rosenthal M., Landon B., Howitt K., Ryu Song H. et Epstein A. (2007), Climbing up the pay-for-performance learning curve : where are the early adopters now ? » *Health Affairs*, Volume 26, Numéro 6, pp. 1674-1682.

Roski J., Jeddeloh R., An L., Lando H., Hannan P., Hall C. et Zhu, S. H. (2003), « The impact of financial incentives and a patient registry on preventive care quality: increasing provider adherence to evidence-based smoking cessation practice guidelines », *Preventive Medicine*, Volume 36, Numéro 3, pp. 291-299.

Roussel P. (1996), *Rémunération, motivation et satisfaction au travail*, Economica, Paris.

Rubinstein A., Rubinstein F., Botargues M., Barani M., et Kopitowski K. (2009), « A multimodal strategy based on pay-per-performance to improve quality of care of family practitioners in Argentina », *Journal of Ambulatory Care Management*, Volume 32, Numéro 2, pp. 103-114.

Ruffin R. et Leigh E. (1973), « Charity, competition and the pricing of doctors' services », *Journal of Human Resources*, Volume 8, Numéro 2, pp. 212–222.

Ryan A. (2009), « Effects of the Premier hospital quality incentive demonstration on Medicare patient mortality and cost », *Health Services Research*, Volume 44, Numéro 3, p. 821-842.

Ryan R. et Deci E. (2001), « On Happiness an Human Potentials : A Review of Reasearch on Hedonic and Eudaimonic Well-Being », *Annual Review of Psychology*, Volume 52, Volume 1, pp. 141-166.

Sabatino S., Habarata N., Baron R., Coates R., Rimer B., Kerner J., Coughlin S., Kalra G. et Chattopadhyay S. (2008), « Interventions to incrase recommendation and delivery of screening for breast, cervical, and colorectal cancers by health care provideers. Systematic reviews of provider assessment and feedback and provider incentives », *American Journal of Preventive Care*, Volume 35, Numéro 1, pp. 67-74.

Sackett D., Rosenberg W., Gray M., Haynes B. et Richardson S. (1996), « Evidence based medicine : what it is and what it isn't », *BMJ*, Volume 312, pp. 71-72.

Salais R., (2010), « Usages et mésusages de l'argument statistiques: le pilotage des politiques publiques par la performance », *Revue Française des Affaires Sociales*, Volume 1, Numéro 2/3, pp. 129-147.

Saint-Lary O., Plu I. et Naiditch M. (2011), « Adhérer ou pas au CAPI : de quel clivage des généralistes le P4P est-il le révélateur ? », *Revue française des affaires sociales*, Volume 2/3, Numéro 2/3, p. 180-209.

Santo M.-V. et Verrier P.-E. (2007), *Le management public*, Que sais-je ?, Paris.

Sargos P. (1998), « Références médicales opposables et responsabilité des médecins », *Médecine et Droit*, Volume 28, pp. 9-12.

Satterthwaite M. (1979), « Consumer information, equilibrium price and the number of sellers », *Bell Journal of Economics*, Volume 10, Numéro 2, pp. 483-502.

Sautter K., Bokhour B., White B., Min D., Young G., Burgess J., Berlowitz D. et Wheeler J. (2007), « The early experience of a hospital-based pay for performance program », *Journal of Healthcare Management*, Volume 52, Numéro 2, pp.95-107.

Schoen C., Osborn R., Doty M., Squires D., Peugh J. et Applebaum S. (2009), « A Survey Of Primary Care Physicians In Eleven Countries, 2009: Perspectives On Care, Costs, And Experiences », *Health Affairs*, Volume 28, Numéro 6, pp. 1171-1183.

Schroeder S. (1992), « Physician supply and the US medical marketplace », *Health Affairs*, Volume 11, Numéro 1, pp. 235-243.

Scott I. (2007), « Pay for performance in health care: Strategic issues for Australian experiments », *Medical Journal of Australia*, Volume 187, Numéro 1, pp. 31-35.

Seignour A., Palpacuer F. et Vercher C. (2007), « Ouvrir la boîte noire du licenciement pour motif personnel », *Gérer et Comprendre*, Numéro 90, p. 20-29.

Sen A. (1987), *On ethics and economics*, Basil Blackwell, Oxford.

Setbon M. (2000), « La qualité des soins, nouveau paradigme de l'action collective? ». *Sociologie du travail*, Volume 42, Numéro 1, pp. 51-68.

Shapiro C. et Stiglitz J. (1984), « Equilibrium unemployment as a worker discipline device », *American Economic Review*, Volume 74, Numéro 3, pp. 433-444.

Sheldon K. et Kasser T. (1998), « Pursuing personal goals : Skills enable progress, but not all progress is beneficial », *Personality and Social Psychologie Bulletin*, Volume 24, Numéro 12, pp. 1319-1331.

Shen Y. (2003), « Selection incentives in a performance-based contracting system », *Health Services Research*, Volume 38, Numéro 2, pp. 535-552.

Siciliani L. (2009), « Paying for performance and motivation crowding out », E*conomic Letters*, Volume 103, Numéro 2, pp. 68-71.

Sicsic J., Le Vaillant M. et Franc, C. (2012), « Intrinsic and extrinsic motivations in primary care: An explanatory study among French general practitioners » *Health policy*, Volume 108, Numéro 2, pp. 140-148.

Simon S., Pan R., Sullivan A., Clark-Chiarelli N., Connelly M., Peters A., Singer J., Inui T. et Block S. (1999), « Views of managed care », *New England Journal of Medicine*, Volume 276, Numéro 1, pp. 928-936.

Simpson C., Hannaford P., Lefevre K., Williams D. (2006), « Effect of the UK incentive-based contract on the management of patients with stroke in primary care », *Stroke*, Volume 37, Numéro 1, pp. 2354-2360.

Skinner B. (1938), *The behavior of organisms*, Prentice-Hall, Englewood Cliffs, New Jersey.

Sliwka D. (2007), « Trust as a signal of a social norm and the hidden costs of incentive schemes », *American Economic Review*, Volume 97, Numéro 3, pp. 999-1012.

Sloan F. (1975), « Physician Supply Behavior in the Short Run », *Industrial and Labor Relations Reviews*, Volume 28, Numéro 4, pp. 549-569.

Smallwood O. (2006), « La normalisation des règles de l'art médical : une nouvelle source de responsabilité pour les professionnels de santé ? », *Médecine & Droit*, Volume 79/80, pp. 121-126.

Solow R. (1971), « Blood and Thunder », *Yale Law Journal*, Volume 80, Numéro 8, pp. 1696-1711.

Spire A. (2012), *Faibles et puissants face à l'impôt*, Raisons d'agir, Paris.

Spooner A., Chapple A. et Roland M. (2001), « What makes British general practitioners take part in a quality improvement scheme ? », *Journal of Health Service Research & Policy*, Volume 6, Numéro 3, pp. 145-150.

Stano M. (1987), « A further analysis of the physician inducement controversy », *Journal of Health Economics*, Volume 6, Numéro 3, pp. 229-238.

Steiner P. (2001), « Don de sang et don d'organes : le marché et les marchandises fictives », *Revue française de sociologie*, Volume 42, Numéro 2, pp. 357-374.

Stigler G. (1961), « The Economics of Information », *Journal of Political Economy*, Volume 69, Numéro 3, pp. 213-225.

Srirangalingam U., Sahathevan S., Lasker S., Chowdhury T. (2006), « Changing pattern of referral to a diabetes clinic following implementation of the new UK GP contract », *British Journal of General Practice*, Volume 56, Numéro 1, pp. 624-626.

Sweeney G. (1982), « The market for physicians' services: Theoretical implications an empirical test of the target income hypothesis », *Southern Economic Journal*, Volume 48, Numéro 3, pp. 594–614.

Tabuteau D. (2013), *Démocratie sanitaire : les nouveaux défis de la politique de santé*, Odile Jacob, Paris.

Tanti-Hardouin N. (2011), *Aux origines de l'économie de la santé*, Ellipses, Paris

Taylor L. (2007), « Optimal wages in the market for nurses: An analysis based on Heyes' model », *Journal of Health Economics*, Volume 26, Numéro 5, pp. 1027-1030.

Théret B. (1990), « La place de l'État dans les théories économiques françaises de la régulation : éléments critiques et repositionnement à la lumière de l'histoire », *Économie appliquée*, Volume 63, Numéro 2, pp. 43-81.

Thévenot L. (1990), « L'action qui convient », dans Pharo P. et Quéré L. (Eds), *Les formes de l'action*, Edition de l'EHESS, Paris, Raison Pratique 1, pp. 39-69.

Thévenot L. (2006), *L'action au pluriel: sociologie des régimes d'engagement*, La Découverte, Paris.

Thirlby R., Quigley T. et Anderson R. (1996), « The shift toward a managed care environment in a multispecialty group practice model: Looking for reciprocal benefits », *Archives of Surgery*, Volume 131, Numéro 10, pp. 1027-1031.

Tirole J. (2002), « Rational irrationality: some economics of self-management », *European Economic Review*, Volume 46, Numéro 4-5, pp. 633-655.

Titmuss R. (1970), *The Gift Relationship*, Allen and Unwin, London.

Thévenot L. (1989), « Equilibre et rationalité dans un univers complexe », *Revue Économique*, Volume 40, Numéro 2, pp. 147-198.

Touraine M. (2014), Conférence de presse – Annonce du Programme national de réduction du tabagisme, Intervention de Marisol Touraine, jeudi 25 septembre.

Town R., Kane R., Johnson P. et Butler M. (2005), « Economic incentives and physician's delivery of preventive care », *American Journal of Preventive Care*, Volume 28, Numéro 2, pp. 234-240.

Town R., Wholey D., Kralewski J. et Dowd B. (2004), « Assessing the influence of incentives on physicians and medical groups », *Medical Care Research Review*, Volume 61, Numéro 3, pp. 80-118.

Upton W. (1973), « Altruism, Attribution and Intrinsic Motivation in the Recruitment of Blood Donors », dans American Red Cross Volume III, *Selected Readings on Donor Motivation and Recruitment*.

Valat B. (2001), *Histoire de la sécurité sociale (1945-1967) : L'État, l'institution et la santé*, Economica, Paris.

Vallerand, R. et Miquelon P. (2008), « Le modèle hiérarchique : une analyse intégrative des déterminants et conséquences de la motivation intrinsèque et extrinsèque », dans R. Joule

et P. Huguet (Eds), *Bilan et Perspective en Psychologie Sociale*, Presses Universitaires de France, pp. 163-203.

Vallerand R. et Thill E. (1993), *Introduction à la psychologie de la motivation*, Etudes vivantes, Laval.

Van Bosch P. (1997), *La Philosophie et le Bonheur*, Flammarion, Essai, Paris.

Van Herck P., De Smedt D., Annemans L., Remmen R., Rosenthal M.. et Sermeus W. (2010), « Systematic review : Effects, design choices, and context of pay-for-performance in health care », *BMC Health Services Research*, Numéro 1, Volume 10, pp. 1-13.

Van Horn I., Burns I. et Wholey D. (1997), « The impact of physician involvement in managed care on efficient use of hospital resources », *Medical Care*, Volume 35, Numéro 9 pp. 873-889.

Vatan S. (2014), *La tarification des services d'aide à domicile. Une analyse institutionnaliste par le rôle paramétrique des prix*, Thèse de doctorat en sciences économiques, Université Lille 1.

Ventelou B. et Videau Y. (2011), « Des motivations des généralistes français » dans Batifoulier P., Buttard A. et Domin J-P., S*anté et Politiques Sociales : entre Efficacité et Justice. Autour des Travaux de Maryse Gadreau*, Paris, Editions ESKA, p. 93-105.

Verlhiac J-F. (2000), « L'effet de faux consensus : Une revue empirique et théorique », *L'Année Psychologique*, Volume 100, Numéro 1, pp. 141-182.

Videau Y. (2010), *Essais sur l'offre de travail en médecine générale. Du rôle des incitations et des motivations*, Thèse de Doctorat de Sciences Économiques, Université de la Méditerranée – Aix Marseille.

Videau Y., Clerc I. et Ventelou B. (2011), « L'effet d'éviction des motivations intrinsèques par les motivations extrinsèques? Une expérimentation par assignation aléatoire pour les médecins généralistes libéraux français », Séminaire d'économie politique de la santé.

Videau Y., Batifoulier P., Arrighi Y., Gadreau M. et Ventelou, B. (2010), « Le cycle de vie des motivations professionnelles en médecine générale: une étude dans le champ de la prévention », *Revue d'Epidémiologie et de Santé Publique*, Volume 58, Numéro 5, pp. 301-311.

Vroom V. (1964), *Work and motivation*, Wiley, New York.

Weber M. (1971), *Économie et société*, Plon, Paris.

Weissman J. et Bigby J. (2009), « Massachusetts health care reform – near-universal coverage at what cost ? », *New England Medical Journal*, Volume 361, Numéro 361, pp. 2012-2015.

Wiggins D. (2005), « An idea we cannot do without », dans S. Reader (Ed), *The philosophy of need*, Cambridge University Press, Cambridge, p. 25-50.

Wilensky G. et Rossiter L. (1981), « The magnitude and determinants of physician-initiated visits in the United States," dans *Health, economics, and health economics*, Van der Gaag J. et Perlman M. (Eds), North-Holland, Amsterdam, pp. 215-43.

Wilensky G. et Rossiter L. (1984), « Identification of Physician-induced Demand », *Journal of Human Resources*, Volume 19, Numéro 2, pp. 231-244.

Wilkinson R. et Pickett K. (2010), *The spirit level: why equality is better for everyone*, Penguin, London.

Winslow C-E. (1951), *The cost of sickness and the price of health*, World Health Organization, Genève.

White E. (1987), « Projected changes in breast cancer incidence due to the trend toward delayed childbearing », *American Journal of Public Health*, Volume 77, Numéro 4, pp. 495-497.

Wennberg J. (1985), « On patient need, equity, supplier-induced demand, and the need to assess the outcome of common medical practices », *Medical Care*, Volume 23, Numéro 5, pp. 512-520.

Woodward R., Schnitzler M. et Kvols L. (1998), « Reduced uncertainty as a diagnostic benefit », *Health Economics*, Volume 7, Numéro 2, pp. 675-700.

Woodward R. et Warren-Boulton F. (1984), « Considering the Effects of Financial Incentives and Professional Ethics on Appropriate Medical Care », *Journal of Health Economics*, Volume 3, Numéro 3, pp. 223–237.

Wynia M. (2009), « The risk of rewards in health care: How pay-for-performance could threaten, or bolster, medical professionalism », *Journal of General Internal Medicine*, Volume 24, Numéro 7, pp. 884-887.

Zweifel P. (1981), « Demande médicale induite par l'offre : chimère ou réalité », *Consommation, revue de socio-économie*, Volume 4, Numéro 1, pp. 39-62.

Zweifel P. et Breyer F. (1997), *Health Economics*, Oxford University Press, New York.

Table des matières

Remerciements **5**

Sommaire **9**

Introduction générale **11**

L'universalité de la performance et ses résistances 12

Le médecin comme « exemple exemplaire » des apories de la performance 14

1 PROBLÉMATIQUE ET THÈSE 15

2 PERSPECTIVE THÉORIQUE : DE LA CRITIQUE DU *MAINSTREAM* À L'ANALYSE INSTITUTIONNALISTE 17

Une cartographie du mainstream 17

Une perspective institutionnaliste 20

3 MÉTHODOLOGIE 22

4 PLAN DE LA THÈSE 24

Première partie : Inciter le médecin à la performance dans le paradigme mainstream 24

Deuxième partie : Les déboires empiriques de la performance médicale 25

Troisième partie : La politique de la performance dans le néolibéralisme sanitaire 26

Première partie : Inciter les individus à la performance dans le paradigme *mainstream*. L'introuvable originalité du travail médical **29**

Introduction de la première partie **31**

Chapitre 1 Le paiement à la performance médicale : héritage de la norme concurrentielle **33**

INTRODUCTION 33

1. EN DÉPIT DE TOUT : LA NORME CONCURRENTIELLE COMME NIRVANA 35

1.1. Le marché des soins médicaux : l'ajustement marchand *36*

1.1.1. Représenter le médecin et l'activité médicale 37

1.1.2. Régulation médicale et ajustement marchand 39

1.1.3. Les propriétés d'optimalité de l'allocation marchande 40

1.2. L'originalité de la santé chez les économistes pionniers *41*

1.2.1. Reuben Kessel (1958), le marché entravé par les syndicats de médecins 42

1.2.2. Kenneth Arrow (1963), le médecin n'est pas un « barbier » 43
1.2.3. Mark Pauly (1978), l'introuvable consommateur 45
1.2.4. Le marché des services médicaux n'existe pas 47
1.3. Les défaillances du marché des soins médicaux 48
1.3.1. Première faille : une concurrence monopolistique 48
1.3.2. Une synthèse des défaillances de marché des services médicaux 51
1.4. La stratégie du Nirvana en santé : le retour du marché ? 54
1.4.1. Le marché n'existe pas mais les économistes parlent de marché 54
1.4.2. Le programme d'Arrow et l'avènement de la théorie standard étendue 54
1.4.3. La stratégie du Nirvana : vers le marché des soins médicaux ? 55
2. LE PAIEMENT À LA PERFORMANCE MÉDICALE POUR COMBLER L'ÉCART ENTRE LE NIRVANA ET LE MONDE RÉEL 58
2.1. L'insupportable asymétrie d'information 60
2.1.1. L'incertitude médicale 60
2.1.2. Les asymétries d'information 62
2.2. La fin du marché : l'hypothèse de demande induite 66
2.3. Les modes de rémunération traditionnels du médecin marqués d'inefficacité 70
2.3.1. Le salariat 71
2.3.2. La capitation 73
2.3.3. Le paiement à l'acte 74
2.4. Un peu plus près des étoiles : le P4P médicale 75
CONCLUSION 79

Chapitre 2 La performance en univers altruiste. Paradoxes de théorie économique et problèmes de politique économique 81
INTRODUCTION 81
1. LA « PRÉFÉRENCE ÉTHIQUE » DU MÉDECIN ET LE CALCUL RATIONNEL 86
1.1 De la demande induite aux deux formes de l'altruisme médical 86
1.2. Un altruisme médical sans épaisseur propre 91
2 ALTRUISME MÉDICAL INSTRUMENTAL : PARADOXES DE THÉORIE ÉCONOMIQUE ET PROBLÈMES DE POLITIQUE PUBLIQUE 93
2.1 Altruisme médical égocentrique 94
2.1.1. L'altruisme instrumental : la dissolution du jugement dans le calcul 94

2.1.2. Quand l'altruisme instrumental rencontre la médecine libérale : un chèque en blanc ? 96
2.1.3. Préférences individuelles contre éthique collective 97
2.2. Altruisme médical sadique 98
2.2.1. De l'« autrisme » au sadisme médical 98
2.2.2. Protéger le patient de son médecin 100
2.3. Altruisme médical masochiste 101
2.3.1. Du biais égocentrique au masochisme médical 101
2.3.2. Protéger le médecin de son patient 103
2.4. Prime à l'égoïsme et punition de l'altruisme 104
2.4.1. Quel contrat pour le médecin altruiste ? 104
2.4.2. Les effets délétères de la prime à l'égoïsme 107
3. BESOINS, ÉTHIQUE EN SANTÉ ET ÉTHIQUE MÉDICALE 108
3.1 Le besoin vital 110
3.3. Le besoin objectif 112
3.2.1. Les besoins de santé sont inégalement distribués et involontaires 112
3.2.2. Les besoins de santé sont variés et idiosyncratiques 113
3.2.3. Les besoins de santé ont une considérable propension à évoluer 114
3.4. Le besoin social 115
3.5. Besoin de santé et éthique médicale 117
CONCLUSION 119

Chapitre 3 : Payer plus pour motiver moins ? Performance médicale et *crowding out effect* 123
INTRODUCTION 123
1. AUX RACINES DU *CROWDING OUT EFFECT* : LA THÉORIE DE L'AUTODÉTERMINATION .. 126
1.1. La TAD comme théorie du bien-être 127
1.2. Besoins psychologiques fondamentaux et motivation intrinsèque 129
1.3. La théorie de l'évaluation cognitive 131
2. AU-DELÀ DE L'IMPÉRIALISME TRADITIONNEL DE LA SCIENCE ÉCONOMIQUE : LE TRAVAIL DE PASSEUR DE BRUNO FREY 135
2.1. De l'hypothèse de crowding out effect à la Motivation crowding theory 136
2.2. Acquérir une légitimité à partir de l'accumulation de preuves empiriques 140
2.3. Une théorie des ressources humaines et de la politique publique 143

3 COMMENT LIRE LES MOTIVATIONS INTRINSÈQUES AVEC LA THÉORIE ÉCONOMIQUE DE L'INFORMATION ? 146
3.1. MI, COE et TSE2 : un problème d'information 146
3.1.1. Les ME signalent la difficulté de l'activité médicale 148
3.1.2. Un problème de sélection adverse : recruter les médecins les plus motivés 149
3.1.3. Les médecins conformistes 150
3.1.4. Le médecin et l'estime de soi 152
3.1.5. Un problème de réputation médicale 153
3.2. La dissolution des MI dans le calcul rationnel 155
3.2.1. MI et désutilité de l'effort 155
3.2.2. MI et désintéressement 157
CONCLUSION 160

Deuxième partie : Une évaluation du paiement à la performance médicalE. Le difficile passage de la théorie à la pratique 163

Introduction de la deuxième partie 165

Chapitre 4 La performance du paiement à la performance médicale. Une revue de littérature 167
INTRODUCTION 167
1. LE DISPOSITIF DE PAIEMENT À LA PERFORMANCE MÉDICALE DANS LES FAITS : JUSTIFICATIONS ET INTERNATIONALISATION 168
1.1. Les États-Unis : le berceau du paiement à la performance 169
1.1.1 Le paradoxe américain : une qualité des soins médiocre pour le système de santé le plus cher du monde 169
1.1.2. Les (nombreux) dispositifs américains de paiement à la performance médicale 174
1.2. Le paiement à la performance au Royaume-Uni : le Quality Outcome Framework 177
1.2.1. L'organisation des soins au Royaume-Uni 178
1.2.2. Le dispositif de paiement à la performance britannique 181
1.3. Une synthèse des différents dispositifs de paiement à la performance 185
2. QUELLE PERFORMANCE DU PAIEMENT À LA PERFORMANCE ? 187
2.1. Une évaluation du P4P : efficacité, conception et contexte 188
2.1.1 L'efficacité et ses différents critères : des résultats mitigés 188

2.1.2. Conception et contexte : expliquer les mauvais résultats du P4P médicale ... 194
2.2. Les effets pervers du paiement à la performance médicale ... 198
2.2.1. La sélection des patients ... 199
2.2.2. La sélection des tâches ... 200
2.2.3. Le crowding out effect médical ... 201
2.2.4. Gaming et effets additionnels ... 202
2.3. Une synthèse des évaluations du P4P médicale ... 204
2.3.1. Le P4P médicale : efficacité limitée, efficience douteuse et effets pervers ... 204
2.3.2. Des limites méthodologiques et un programme de recherche aveugle ... 208
3. QUELLE MOTIVATION PROFESSIONNELLE POUR LE PAIEMENT À LA PERFORMANCE ? ... 209
CONCLUSION ... 214

Chapitre 5 Le paiement à la performance médicale en France : les performants ne sont pas forcément les gagnants ... 217
INTRODUCTION ... 217
1 LE PAIEMENT À LA PERFORMANCE MÉDICALE EN FRANCE ... 221
1.1. La justification de la Rémunération sur objectif de santé publique ... 221
1.2. La formule de rémunération du ROSP ... 223
1.3. Les indicateurs de performance du ROSP : entre réalisme et constructivisme . 227
1.3.1. Les indicateurs de la ROSP ... 228
1.3.2. Des indicateurs de performance ? ... 234
2 UNE ANALYSE STATISTIQUE DU ROSP EN ILE DE FRANCE : LES PERFORMANTS NE SONT PAS FORCÉMENT LES GAGNANTS ... 239
2.1. Les données : les bases du Service médical de l'assurance maladie d'Ile-de-France 239
2.2. Le profil moyen du médecin traitant généraliste et de sa patientèle ... 242
2.3. Les résultats ROSP : taux de réalisation et rémunération à la performance ... 244
2.4. Des gagnants et des performants ... 247
2.5. Les performants ne sont pas toujours gagnants ... 254
3 LA VARIABLE CACHÉE DE LA PERFORMANCE : L'EFFET DE COMPOSITION DE LA PATIENTÈLE ... 259
3.1. Méthode ... 259
3.2. Résultats ... 261
3.3. Discussion ... 265

3.3.1. Le rôle de la patientèle ALD dans la performance 266
3.3.2. La part des patients CMUC négativement corrélé à la performance 267
3.3.3. La part de patientes âgées entre 50 et 74 ans négativement corrélée à la performance 268
3.3.4. La prescription d'antibiotiques : une corrélation négative avec la performance mais pas avec sa rémunération 268
3.3.5. La part de patients sous antiagrégants plaquettaires : une corrélation positive avec la performance 269
3.3.6. La performance et les honoraires : le prix ne signale pas la qualité 269
CONCLUSION 270

Troisième partie : Du paiement à la performance médicale à la régulation par les pratiques. Vers le néolibéralisme sanitaire ? 275

Introduction de la troisième partie 277

Chapitre 6 La médecine libérale française contre l'État social.

De la régulation par les prix à la régulation par les pratiques 279

INTRODUCTION 279
1 MÉDECINE LIBÉRALE ET SOLIDARITÉ NATIONALE. UN APPRENTISSAGE DIFFICILE 282
1.1. L'émergence de la solidarité nationale freinée par la profession médicale 283
1.1.1 La profession se construit contre la concurrence des officiers de santé… 283
1.1.2 … tout en refusant la remise en cause de l'entente directe 288
1.2. L'affirmation de la solidarité nationale par la mise en place progressive du conventionnement 292
1.2.1. Des textes qui reflètent un choix de société audacieux… 292
1.2.2. … mais qui ne se concrétisent que tardivement 296
2. LA RÉGULATION PAR LES PRATIQUES : UN EFFET DE LA REMISE EN CAUSE DE LA SOLIDARITÉ NATIONALE 300
2.1. Une politique de prix au détriment des patients 301
2.1.1. La revanche de l'entente directe… 301
2.1.2. … et une politique délibérée de reste à charge 306
2.2. Le second grand tournant de la médecine libérale : le contrôle de la pratique médicale 311
2.2.1. Du premier au second grand tournant de la médecine libérale 312
2.2.2. Le P4P médicale comme instrument du contrôle des pratiques 317

CONCLUSION 322

Chapitre 7 La qualité des soins est-elle soluble dans la quantification ? 327

INTRODUCTION 327

1. LA NOUVELLE RATIONALITÉ : PRODUIRE DE LA QUALITÉ PAR LES QUANTITÉS 331

1.1. Du New Public Management à la reprise en main nécessaire de la médecine libérale 331

1.1.1. Le New Public Management : il n'existe pas mais il est partout 332

1.1.2. Rompre avec le modèle bureaucratique de l'action publique 333

1.1.3. Les quatre formes du NPM 335

1.1.4. Le NPM dans le contexte Français : l'exemple de la LOLF 338

1.1.5. NPM et médecine libérale : l'exigence du contrôle 340

1.2. Produire et certifier les normes médicales de contrôle : Evidence Based Medicine et agences indépendantes 342

1.2.1. Produire les normes quantifiées : l'Evidence Based Medicine 342

1.2.2. Certifier les normes quantifiées : les Agences 345

1.3. L'évolution de la responsabilité médicale : la Soft Law 348

1.3.1. La mobilisation des normes par les acteurs 349

1.3.2. Les principes de la responsabilité du médecin à partir des données « acquises » de la science 350

1.3.3. L'émergence de la soft law à partir des normes de bonne pratique 351

2. CRITIQUE DE LA NORMALISATION DU TRAVAIL MÉDICAL 354

2.1. Le gouvernement par le chiffre 355

2.1.1. Qualité totale et performance totale 355

2.1.2. Réalisme et constructivisme des objets statistiques 357

2.2. Une épistémologie réductrice de la maladie 358

2.2.1. Deux conceptions de la maladie 358

2.2.2. Quelle neutralité méthodologique ? 360

2.2.3 Evidence based medicine et industrie pharmaceutique 362

2.2.4. L'incertitude sur la qualité persiste et le P4P pose question 364

2.3. Une épistémologie réductrice des services de soins 365

2.3.1. Le soin n'est pas un produit 365

2.3.2 .L'impossible nomenclature des maladies et des actes 366

2.3.3. Le patient est au cœur du service : la « coproduction » 368

CONCLUSION 372

Chapitre 8 Une analyse conventionnaliste de la régulation industrielle des pratiques médicales 377

INTRODUCTION 377

1. LA CONVENTION DE QUALITÉ DU « SOIN » : CONVENTION INSPIRÉE/DOMESTIQUE VERSUS CONVENTION INDUSTRIELLE 379

1.1. Le paradigme de l'Économie des conventions et les conventions de qualité 380

1.2. La convention industrielle de qualité des soins 383

1.3. De la sociologie des professions à la convention inspirée/domestique de qualité des soins 385

1.4. Convention inspirée/domestique versus convention industrielle 388

2. BOULEVERSEMENT DE L'ÉTHIQUE MÉDICALE ET MARCHANDISATION DE LA SANTÉ 392

2.1. L'éthique médicale sous influence de la convention industrielle 393

2.1.1. Les « valeurs propres » et l'éthique médicale 393

2.1.2. Ce que fait la convention industrielle à l'éthique médicale 396

2.1.3. Un réagencement des valeurs 400

2.2. L'industrialisation de la médecine, levier à la marchandisation des soins 401

2.2.1. Le paiement à la performance médicale : entre logique industrielle et logique marchande 401

2.2.2. Un interprétation du crowding ou effect médical 403

2.2.3. La statistique : l'instrument de gouvernement derrière l'outil de preuve 405

2.2.4. De l'État néolibéral à la gouvernementalité néolibérale 410

2.2.5. La convention industrielle propagatrice du néolibéralisme 411

2.2.6. Le rôle de la convention industrielle dans la privatisation de la santé 413

CONCLUSION 416

Conclusion Générale 419

La « défaite du patient » 422

Perspectives pour la médecine ambulatoire 424

Les deux logiques de la performance 427

Références bibliographiques 429

Table des matières 473

Table des encadrés 483

Table des tableaux **485**

Table des graphiques **487**

Table des encadrés

Encadré 1.1. La concurrence monopolistique (A partir de McGuire, 2000) **50**

Encadré 1.2. Biens publics et biens privés **52**

Encadré 1.3. Présentation canonique de la relation d'agence. Le modèle de Ross (1973) **63**

Encadré 1.4. L'hypothèse de demande induite. Le modèle de McGuire et Pauly (1991) **69**

Encadré 1.5. Case mix, capitation et P4P. Le modèle de Baker (1992) **78**

Encadré 2.1. L'éthique médicale comme la désutilité du pouvoir d'induction. Le modèle de Rossiter et Wilenski (1984) **88**

Encadré 2.2. L'éthique médicale discriminatoire. Le modèle de Ruffin et Leigh (1973) **90**

Encadré 2.3. Le sadisme médical dans la configuration du dilemme du prisonnier **100**

Encadré 2.4. Le dilemme du Samaritain **103**

Encadré 2.5. Inciter des médecins à révéler leur altruisme. Le modèle de Jack (2005) **107**

Encadré 2.6. Les différents types de besoin chez Wiggins (2005) **111**

Encadré 2.7. Les préférences adaptatives. Elster (1985) : le renard et les raisins **116**

Encadré 3.1. La théorie de l'intégration organismique. Deci et Ryan (2002) **132**

Encadré 3.2. La Motivation crowding out theory. Le modèle de Frey (1997) **141**

Encadré 4.1. La fragmentation du système de santé américain **170**

Encadré 4.2. Trois programmes de P4P médicale emblématiques **177**

Encadré 4.3. La recherche de l'efficience : une pluralité de réformes **179**

Encadré 5.1. La performance selon l'assurance maladie **226**

Encadré 6.1. Les régimes d'accumulation dans la théorie de la régulation Boyer (2004) **287**

Encadré 6.2. La théorie de l'aléa moral « machine de guerre contre le patient ». Batifoulier (2014) **308**

Encadré 6.3. AcBUS, CBP et CSP : D'après la Loi de Financement de la Sécurité Sociale (JO, 2000 et 2002) **317**

Encadré 6.4. Les conditions de rémunération dans le cadre du CAPI D'après Journal officiel (2009) **319**

Encadré 6.5. Le ROSP comme troisième pilier de la rémunération des médecins libéraux français Extraits de l'article 25 de la Convention médicale (2011, p. 33-34) **322**

Encadré 7.1. Les dix commandements de la nouvelle gestion publique Osborne et Gaebler (1992) **338**

Encadré 7.2. L'émergence du gold standard de l'essai clinique randomisé. Keel (2011) **344**

Encadré 7.3. Quelques exemples de références médicales opposables (1993) Le Pape et Sermet (1998) **353**

Encadré 7.4. La dimension politique des essais cliniques : le cas de l'étude Diet Heart. Marks (1999) **362**

Encadré 8.1. Les Économies de la grandeur : une pluralité de valeur **383**

Encadré 8.2. Singularité et complexité : deux caractéristiques majeures de l'activité permettent de distinguer les métiers des professions **386**

Encadré 8.3. La coloration marchande de l'éthique médicale **396**

Table des tableaux

Tableau 1.1. La théorie standard et les problèmes du comportement médical **39**

Tableau 1.2. Les textes fondateurs d'économie de la santé et l'idéal type marchand..... **48**

Tableau 1.3. L'évaluation des modes de rémunération traditionnels par la TSE1..... **71**

Tableau 2.1. Divers termes pour qualifier l'altruisme médical..... **92**

Tableau 2.2. Des paradoxes de théorie économique aux problèmes de politique économique **94**

Tableau 2.3. Le dilemme du prisonnier avec un médecin sadique..... **100**

Tableau 2.4. Le dilemme du samaritain **103**

Tableau 2.5. Comparaison des besoins de santé, d'éducation et d'alimentation **112**

Tableau 2.6. Besoins de santé, éthique en santé et éthique médicale..... **117**

Tableau 3.1. L'étude empirique du COE dans différents champs de la littérature : un échantillon **143**

Tableau 3.2. Le crowding out effect dans la TSE2 ..... **147**

Tableau 4.1.Quelques indicateurs d'état de santé et de qualité des soins, 2011 ..... **173**

Tableau 4.2. Les programmes de P4P médicale dans le secteur privé aux États-Unis **176**

Tableau 4.3. Synthèse des indicateurs du QOF..... **181**

Tableau 4.4. Le traitement de l'asthme dans le QOF..... **182**

Tableau 4.5. Les indicateurs de qualité et de productivité dans le QOF..... **184**

Tableau 4.6. Incitation et P4P dans les grands systèmes de santé internationaux **185**

Tableau 4.7. Les principaux programmes de P4P médicale **186**

Tableau 4.8. L'évaluation du P4P médicale..... **205**

Tableau 4.9. Les effets pervers du P4P médicale..... **207**

Tableau 4.10. Ce que révèle le paiement à la performance sur la motivation médicale **210**

Tableau 4.11. Ce que révèlent les effets d'autres types d'incitations sur la motivation médicale **211**

Tableau 4.12. Les tests sur le niveau de motivation intrinsèque des médecins et sur le crowding out effect..... **212**

Tableau 5.1. L'indicateur de dépistage du cancer du sein..... **224**

Tableau 5.2. Indicateurs portant sur l'organisation du cabinet et de qualité du service **229**

Tableau 5.3. Les indicateurs de qualité de la pratique médicale. Thème suivi des pathologies chroniques **231**

Tableau 5.4. Les indicateurs de qualité de la pratique médicale. Thème prévention..... **232**

Tableau 5.5. Les indicateurs de qualité de la pratique médicale. Thème efficience **233**

Tableau 5.6. Synthèse des points ROSP **234**

Tableau 5.7. Type d'indentification et critiques des indicateurs de performance..... **235**

Tableau 5.8. Profil des médecins traitants généralistes en 2012 ..... **242**

Tableau 5.9. Résultats ROSP des médecins traitants généralistes par thème en 2012..... **245**

Tableau 5.10. Résultats ROSP des médecins traitants généralistes par indicateur en 2012 .. **247**

Tableau 5.11. Caractéristiques des médecins performants..... **249**

Tableau 5.12. Profil du médecin gagnant..... **251**

Tableau 5.13. Profil de la patientèle du médecin gagnant..... **253**

Tableau 5.14. Les performants ne sont pas toujours gagnants..... **254**

Tableau 5.15. Les variables retenues..... **261**

Tableau 5.16. Régression par quantiles et MCO sur taux de réalisation..... **262**

Tableau 5.17. Régression par quantiles et MCO sur la rémunération à la performance..... **264**

Tableau 6.1. Les quatre principaux régimes d'accumulation..... **287**

Tableau 7.1. Adoption du NPM en fonction de la couleur politique dans les pays de l'OCDE..... **333**

Tableau 7.2. La mission santé de l'État français..... **340**

Tableau 7.3. Les régimes de performance (à partir de Jany-Catrice, 2012)..... **357**

Tableau 8.1. Convention inspirée/domestique versus convention industrielle **389**

Tableau 8.2. L'État, le marché et les statistiques (Desrosières, 2009, p. 12)..... **408**

Tableau 8.3. Les deux conceptions du soin..... **416**

Table des graphiques

Graphique 0.1. Table d'orientation (Favereau, 2010) **18**

Graphique 0.2. Table d'orientation des TSE **20**

Graphique 4.1. Dépenses de santé en dollars par habitant selon le financeur (2011) **171**

Graphique 5.1. Les performants ne sont pas toujours gagnants **255**

Graphique 5.2. Nuage de points Honoraires/Taux de réalisation **256**

Graphique 5.3. Nuage de points Nombre de patients MT RG/taux de réalisation **256**

Graphique 5.4. Nuage de points Nombre de patients MT RG/Rémunération ROSP **257**

Graphique 5.5. Nuage de points Honoraires/Rémunération ROSP **258**

Graphique 5.6. Nuage de points Taux de réalisation/Rémunération ROSP **258**

Graphique 8.1. Une pluralité de privatisations (Batifoulier, 2014) **414**

Instituer la performance : une application au travail du médecin

Résumé

L'émergence de la logique de performance marque un changement majeur dans les stratégies publiques ou privées de management des ressources humaines. La relation médicale est exemplaire de cette évolution. Alors qu'historiquement les négociations entre l'Etat et les médecins portaient exclusivement sur des problématiques de prix, depuis le début des années 1990, la régulation publique se fait par les pratiques. L'objectif du contrôle est alors de promouvoir la qualité des soins – notamment sur les enjeux de santé publique – et la réduction des dépenses – en évitant le développement des maladies chroniques et en favorisant la prescription de médicaments génériques. L'introduction d'un dispositif de paiement à la performance médicale, en 2011, est l'étape ultime de ce tournant métrologique de la profession qui conduit à multiplier les dispositifs d'évaluation chiffrée de la pratique médicale et à mettre en indicateurs le travail du médecin. Dans une perspective institutionnaliste, notre thèse propose d'interroger la pertinence de cette réforme visant à instituer la performance.

Nous montrons que cette institution de la performance n'est ni efficace ni efficiente. En conduisant à de nombreux effets pervers, elle se fait au détriment des intérêts des patients et des médecins de première ligne. L'injonction à la performance ne conduit pas à améliorer la qualité des soins et à renforcer la maîtrise des dépenses de santé, contrairement aux objectifs annoncés. Par contre, dans l'esprit du néolibéralisme contemporain, la santé est assimilée à un bien comme un autre autour duquel il est possible de mettre en concurrence les producteurs et les consommateurs.

Mots clés : Médecin, performance, nouvelle gestion publique, médecine fondée sur les preuves, néolibéralisme.

Instituting performance: applied to physician labour

Abstract

The emergence of the logic of performance illustrates a major change in public or private human resources management strategies. The health care relation is exemplary of these evolutions. Historically, negotiations between the State and physicians exclusively focused on prices. However, since the beginning of the 90's, public regulation is carried out through professional norms. The goal is to improve the quality of care and to control public spending (avoiding chronic diseases and encouraging the use of generic drugs). The introduction of pay-for-performance in France in 2011 is the ultimate stage in this metrological turning point defined by a numerically-based assessment of medical work. In an institutionalist viewpoint, our thesis seeks to question the relevance of this reform which aims at instituting performance.

We show that this project is neither effective nor efficient. Leading to numerous perverse effects, it is implemented regardless of the patients' interests and of the general practitioners. Contrary to what was proclaimed, this injunction for performance did not achieve the improvement of quality in medical care and the control of health expenditures. Yet, in the contemporary neoliberalism spirit, health is associated with a good as any other, surrounding which it is possible to generate competition among producers and consumers.

Keyword: Physician, performance, new public management, evidence based medicine, neoliberalism.