

Asymétries d'information et incertitude en santé : les apports de la théorie des contrats

In: *Économie & prévision*. Numéro 129-130, 1997-3-4. Nouvelles approches micro-économiques de la santé. pp. 11-24.

Résumé

Asymétries d'information et incertitude en santé : les apports de la théorie des contrats

par Lise Rochaix

Le caractère imparfait, voire asymétrique, de l'information dont dispose le patient conduit à remettre en cause l'utilisation du modèle néoclassique standard et à appliquer les modèles micro-économiques de type Principal - Agent. L'objet de cet article est de dresser un bilan des applications récentes au secteur de la santé de la théorie des contrats. Si cette dernière offre une grille de lecture pertinente, elle est toutefois demeurée peu opérationnelle, essentiellement en raison de la fragmentation des décisions sur le marché des soins. Il convient donc de s'interroger sur la portée et les limites de ce type d'approche. Cette démarche apparaît d'autant plus importante que de nouvelles familles de modèles théoriques, de type multi-principaux, multi-agents ou avec tiers offrent aujourd'hui des perspectives rénovées d'application au champ de la santé.

Abstract

Information Asymmetry and Health Uncertainty: The Contribution of Contract Theory

by Lise Rochaix

The imperfect and even asymmetric nature of the information at the patient's disposal calls into question the use of the standard neo-classical model and has led to application of microeconomic principal-agent models. This paper sets out to assess recent applications of contract theory to the health sector. Although the latter offers a relevant model of interpretation, it is still of relatively little use, mainly because of the fragmentation of decisions relating to the health-care market. The scope and limitations of such an approach must therefore be examined. This is all the more important in that new families of theoretical models involving multiple principals, multiple agents or third parties are now offering fresh angles of application to the field of health.

Citer ce document / Cite this document :

Rochaix Lise. Asymétries d'information et incertitude en santé : les apports de la théorie des contrats. In: *Économie & prévision*. Numéro 129-130, 1997-3-4. Nouvelles approches micro-économiques de la santé. pp. 11-24.

doi : 10.3406/ecop.1997.5861

http://www.persee.fr/web/revues/home/prescript/article/ecop_0249-4744_1997_num_129_3_5861

telles tentatives de modélisation. À cet effet, la formule célèbre de Joan Robinson (1972) selon laquelle "tout l'art consiste à découvrir les simplifications qui permettent d'utiliser le modèle sans pour autant éliminer l'essentiel de la situation réelle qu'il est destiné à éclairer" sera source constante d'inspiration.

Trois temps caractérisent l'évolution de la modélisation appliquée au champ de la santé, qui formeront les trois parties de cette synthèse : les premières tentatives d'intégration des phénomènes d'incertitude et d'asymétrie d'information par des modifications plus ou moins fondamentales des modèles néo-classiques traditionnels, les applications des modèles *Principal-Agent*, les extensions liées aux développements plus récents de la théorie des contrats.

La souveraineté du consommateur sur la sellette

La mise en cause de la souveraineté du consommateur dans le champ d'application qu'est la santé a été contemporaine aux interrogations suscitées au plan théorique par les problèmes liés à l'information imparfaite⁽²⁾. Pourtant, ces phénomènes d'imperfection, voire d'asymétrie d'information n'ont reçu qu'un traitement bien tardif par rapport à d'autres champs d'application comme l'économie du travail ou l'économie industrielle. Il convient donc de retracer pas à pas ce cheminement vers une approche plus réaliste du secteur de la santé, afin de mieux cerner les réticences mais aussi les difficultés réelles justifiant ce retard.

À l'origine de la première remise en question du bien fondé de l'application du modèle néo-classique au marché des soins de santé se trouvent, de fait, des interrogations de nature empirique, suscitées par l'hypothèse d'induction de demande⁽³⁾. Selon cette hypothèse émise par Evans (1974), le médecin disposerait d'un pouvoir discrétionnaire lui permettant de gonfler artificiellement la demande des services médicaux s'il le désire. Le débat sur la demande induite a rapidement dépassé la querelle empirique car il remet en question le postulat fondamental des modèles néo-classiques, à savoir l'indépendance des fonctions d'offre et de demande de soins médicaux, ou encore, la souveraineté d'un consommateur parfaitement informé. Face aux implications fortes de cette asymétrie d'information entre médecin et patient, deux types de modélisation ont été développés, impliquant une remise en cause plus ou moins importante du modèle néo-classique de départ.

L'élargissement de la fonction objectif du médecin

La première et la plus ancienne offre une réponse que l'on pourrait qualifier d'hétérodoxe. Selon Evans (1984), il existe une différence de nature entre les soins médicaux et les autres biens. Il n'est pas garanti que le patient puisse toujours s'informer afin d'exercer son libre choix et les coûts d'information sont prohibitifs, dans la mesure où ils impliquent parfois un retard de traitement préjudiciable à la santé. En conséquence, la contrainte de marché traditionnellement exercée par le consommateur ne peut pas s'exercer sur le marché des soins médicaux.

Les premières tentatives de représentation du comportement économique du médecin se sont inscrites dans le cadre théorique défini par Evans (1974) autour de la notion de revenu cible. Cette approche s'inspire de la théorie du *satisficing* de Simon (1957) selon laquelle l'individu ne maximise pas son niveau de satisfaction, mais cherche plutôt à obtenir un niveau de satisfaction "acceptable". En conséquence, selon cette hypothèse, les médecins n'utiliseront leur pouvoir de création de demande que s'ils sont en deçà de leur revenu-cible. Bien que réalistes, car permettant d'analyser le phénomène de demande induite comme résultant d'une situation de déséquilibre, les modèles du revenu-cible n'ont toutefois aucune valeur prédictive, n'étant pas fondés sur la notion traditionnelle de maximisation d'une fonction d'utilité.

La plupart des autres modèles s'inscrivent dans une perspective de maximisation du profit ou d'un arbitrage travail/loisirs. Mais la prise en compte de l'absence d'arbitrage par le consommateur requiert, pour atteindre l'équilibre, l'introduction de nouveaux arguments dans la fonction d'utilité du médecin.

C'est en général l'éthique qui est ajoutée comme limite à l'utilisation du pouvoir discrétionnaire. Certains auteurs comme Murray (1974) et Richardson (1981) introduisent directement dans la fonction d'utilité du médecin celle d'un patient représentatif, comme limite au pouvoir de création de demande. D'autres, comme Wolfson (1974) font l'hypothèse que le pouvoir discrétionnaire du médecin est limité par le niveau d'excès de demande que les médecins désirent maintenir de façon à pouvoir sélectionner les cas les plus intéressants professionnellement, ainsi que par le désir de fournir des services de qualité. Dionne et Contandriopoulos (1984) ajoutent le prestige du médecin, par delà le revenu, le loisir et l'éthique professionnelle, cette dernière étant une fonction du degré de précision de l'information diagnostique donnée par le médecin au malade.

La critique générale que l'on peut adresser à ces premiers modèles est l'hypothèse de totale dépendance de la fonction de demande à l'égard des préférences des offreurs de soins : les patients

n'interviennent pas en tant qu'acteurs à part entière, dans la mesure où leurs choix s'alignent toujours sur ceux de leurs médecins.

Le retour du patient comme acteur à part entière

Ce deuxième courant postule une simple différence de degré entre les soins de santé et les autres biens et services. Les coûts d'information sur les soins de santé sont certes élevés, voire plus élevés que sur la plupart des autres marchés. Mais il est toujours possible au patient d'acquérir cette information. En conséquence, le modèle néo-classique centré sur l'arbitrage conventionnel entre revenu et loisir doit simplement être modifié à la marge pour intégrer les coûts d'information et de contrôle supportés par le patient afin de réduire l'asymétrie d'information. Cette remise en cause du modèle néo-classique est moins fondamentale puisque l'hypothèse de maximisation est conservée. Elle s'inscrit dans un mouvement général de relâchement des hypothèses fortes en matière d'information du modèle de concurrence pure et parfaite.

Pauly et Satterthwaite (1981) offrent une première tentative de modélisation centrée sur la notion de monopole croissant. Partant de l'idée que les services de santé sont des biens d'expérience dont la qualité ne peut être appréciée en une seule fois (par opposition aux biens de recherche), ils obtiennent le résultat selon lequel plus le nombre de médecins est élevé, plus le nombre d'expériences est limité et moins précise sera l'estimation de la qualité des services du médecin. Du fait de l'aversion pour le risque du patient, les fonctions de demande sont moins élastiques dans les zones à forte densité médicale. La rente informationnelle du médecin croît donc avec la densité médicale, de telle sorte que les tarifs peuvent être plus élevés dans les zones de plus forte densité. Ce modèle, critiqué pour son manque de réalisme⁽⁴⁾, a cependant ouvert la voie en matière de modélisation du comportement du patient.

Dranove (1988) introduit de manière explicite la notion d'incertitude dans un modèle où le médecin établit un diagnostic et choisit à la fois le prix et le traitement recommandé, de façon à maximiser son utilité (définie comme une fonction du revenu et du niveau d'activité, eux-mêmes fonction du nombre de patients traités). Le patient effectue la demande de première consultation et dans un deuxième temps, accepte ou refuse le traitement. Cette deuxième décision dépend de sa propre évaluation de la gravité de la maladie et de la recommandation de traitement du médecin. Un élément stratégique est introduit par le fait que les patients interprètent de manière subjective la recommandation de traitement du médecin (comme reflétant son attitude à prescrire juste ou pas). Le fait que cette dernière constitue un signal pour les patients est intégré dans le modèle par le recours à la notion d'équilibre d'anticipations rationnelles. La limite au pouvoir discrétionnaire du

médecin est constituée par le fait que plus les médecins utilisent ce pouvoir (en recommandant des niveaux de traitement de plus en plus élevés), plus ils perdent de la crédibilité aux yeux de leurs patients et donc moins ceux-ci sont prêts à accepter le traitement recommandé. Les prédictions du modèle sont que les médecins prescrivent un niveau de traitement trop élevé, surtout lorsque la densité médicale est forte, ou lorsque les patients sont assurés. Bien que le modèle s'adresse spécifiquement au problème du différentiel d'information entre médecin et patient, l'équilibre d'anticipations rationnelles nécessite malgré tout que le patient connaisse précisément la stratégie de prescription du médecin et puisse adapter sa stratégie d'acceptation ou de refus du traitement en fonction des modifications par le médecin de sa propre stratégie.

Le modèle de contrats implicites de Coyte (1984) offre une explication théorique alternative des phénomènes de discrimination par les prix et de limite au pouvoir de création de demande du médecin à partir d'un simple modèle d'arbitrage travail/loisir. Il postule que le médecin est employé par une organisation de santé maximisant les profits attendus par le choix d'une combinaison d'*inputs* médicaux (incluant le temps passé par le médecin) qui minimise les coûts. Le médecin pour sa part maximise sa fonction d'utilité attendue, composée du revenu et du loisir. Ces organisations se font concurrence en offrant au consommateur un contrat implicite : un prix (prenant en considération à la fois la dépense financière et en temps pour le patient) et une quantité. Ce contrat est la variable de choix des organisations et le contrat optimal est choisi par la solution d'un programme de maximisation des profits. L'équilibre est étudié pour les cas d'information soit parfaite soit asymétrique. Dans le premier cas, le modèle parvient à expliquer l'existence d'un phénomène de discrimination par les prix en relâchant l'hypothèse d'homogénéité des patients. Dans le cas d'information asymétrique, les médecins tirent parti du différentiel d'information en surestimant la réduction par le traitement de la durée attendue de la maladie, ce qui conduit les patients à consommer plus de services médicaux que ce qui serait nécessaire sur un plan médical. Le modèle permet alors d'expliquer le fait que les médecins parviennent à maintenir le niveau des honoraires, malgré une situation d'excès d'offre sur le marché, du fait des coûts élevés d'information et de contrôle du médecin par le patient.

La critique émise par les tenants d'une différence de nature entre soins de santé et autres biens à l'égard de ce dernier modèle, à savoir la mise en cause de la capacité des patients à exercer un arbitrage entre les différents contrats implicites, est généralisable à toute tentative de réintroduction d'une fonction de demande (Mooney et Ryan, 1993).

Le survol sélectif des premiers travaux de modélisation du marché des services médicaux retrace les vives controverses sur la capacité du modèle néo-classique à rendre compte de la spécificité de ce marché. Le recours à des modèles de type *Principal-Agent* s'est alors imposé à l'issue de cette confrontation animée entre économistes orthodoxes et hétérodoxes. Ils offrent en effet une voie plus consensuelle pour tenter d'appréhender les deux phénomènes si prégnants en santé que sont l'asymétrie d'information et l'incertitude. Ceci peut expliquer l'écho favorable qu'ils ont rencontré auprès des économistes de la santé, sans toutefois que l'on ait pu prédire tant l'engouement suscité que les difficultés rencontrées dans leur application.

Apports et limites des modèles *Principal-Agent* appliqués à la santé

Dès 1980, Pauly souligne le rôle d'agent du patient joué par le médecin sur le marché des services médicaux. Arrow (1986) pour sa part ira jusqu'à soutenir que la relation patient - médecin constitue l'illustration quasi parfaite d'une relation *Principal-Agent* avec risque moral⁽⁵⁾. Avant d'apprécier le bien fondé de cette analogie, il convient de présenter dans un premier temps les grandes lignes de ces modèles *Principal-Agent*.

Le cadre général

Les modèles *Principal-Agent* constituent un cas particulier de la théorie des contrats dont l'objet est d'appréhender les relations d'échange entre des parties en tenant compte des contraintes institutionnelles et informationnelles dans lesquelles elles évoluent⁽⁶⁾. Cette théorie trouve ses origines dans les insuffisances du modèle d'équilibre général et constitue, selon Salanié, (1994) le détour de production nécessaire par des modèles d'équilibre partiel pour fournir une description plus réaliste de l'économie. Le but est ainsi de tenir compte "de toute la complexité des comportements stratégiques des agents au sein des liens institutionnels qui définissent les possibilités de leur action".

Les modèles *Principal-Agent* analysent des échanges bilatéraux en mettant en relation une partie informée (l'*Agent*) et une partie non informée (le *Principal*), le but pour ce dernier étant de proposer à l'*Agent* un contrat qui lui soit acceptable et qui aboutisse à l'utilité maximale pour le *Principal*. Plusieurs types de modèles sont définis en théorie des contrats, selon que l'on s'intéresse à des situations d'information cachée sur les caractéristiques (anti-sélection) de l'*Agent* ou à des problèmes d'action cachée de l'*Agent* (aléa de moralité). Baron et Myerson (1982) intègrent ces

deux aspects pour le cas où les coûts de la firme régulée ne sont pas observables *ex post*, Laffont et Tirole (1986) développant pour leur part un modèle dans lequel des observations sur les coûts permettent d'améliorer la performance du mécanisme incitatif.

Au vu de l'importance du modèle d'aléa de moralité⁽⁷⁾ il est utile de rappeler brièvement quelques-uns de ses traits caractéristiques. La notion de risque y est inhérente du fait du caractère aléatoire de la relation qui lie les actions et les résultats (relation appelée "technologie"). Le contrat doit permettre un partage optimal de ce risque entre *Principal* et *Agent*, le premier étant supposé neutre à l'égard du risque alors que l'*Agent* a de l'aversion au risque : étant plus "petit", il est en effet moins à même de diversifier les risques.

Deux cas-limite permettent d'obtenir un optimum de premier rang :

- celui où l'*Agent* et le *Principal* ont les mêmes préférences quant au choix ;
- celui où les préférences diffèrent mais où le *Principal* peut vérifier le comportement de l'agent sur la base de l'observation des coûts. Dans ce cas, le *Principal* peut ordonner à l'*Agent* de choisir l'action la plus efficace et le contrat veillera seulement au partage optimal des risques. Dans un tel contrat de premier rang, le *Principal* assure l'*Agent* en lui offrant une rémunération indépendante de la réalisation de l'*aléa*.

Ces deux situations ne sont cependant guère réalistes. Dans le premier cas, il suffit en effet d'introduire dans la fonction d'utilité de l'*Agent* d'autres arguments, notamment la notion d'effort, ou bien la dépense supportée par l'*Agent* au nom du *Principal* pour que leurs intérêts divergent. La désutilité retirée de cet argument supplémentaire peut alors conduire l'*Agent* à choisir une action qui ira à l'encontre de l'intérêt du *Principal*. Dans le deuxième cas, la probabilité est forte que le *Principal* ne puisse observer parfaitement les décisions de l'*Agent* qui affectent son utilité car cela nécessiterait des coûts de *monitoring* très élevés.

En conséquence, les situations d'information asymétrique conduisent en règle générale à des solutions de second rang liées au caractère conflictuel des intérêts entre *Principal* et *Agent* et à l'imparfaite observabilité de l'action de l'*Agent* : le *Principal* observe en effet une variable corrélée avec l'action prise par l'*Agent*, à savoir le résultat. Le contrat optimal défini par le *Principal* doit donc opérer un arbitrage entre :

- le partage des risques, qui implique que la rémunération offerte dépende peu du résultat (du fait de l'aversion au risque de l'*Agent*) ;
- la recherche des incitations, qui nécessite une rémunération pour partie fonction du résultat.

Ce sont ces incitations (appelées contraintes d'incitation) qui réduisent le profit du *Principal* en situation de second rang. Elles garantissent à l'*Agent* une rente informationnelle en s'assurant que le principal ne s'appropriera pas tout le surplus de l'échange.

Les parallèles établis avec le marché des services médicaux

L'analyse approfondie de l'interaction patient - médecin conduit à nuancer le parallèle établi entre relation patient - médecin et *Principal-Agent*. Elle souligne ainsi l'existence de certaines caractéristiques fondamentales qui rendent peu adaptée la solution traditionnelle des modèles de *Principal-Agent*, à savoir la définition de contrats optimaux. En particulier ont été mis en évidence les éléments suivants⁽⁸⁾ :

– le marché se caractérise par un éclatement des responsabilités entre ceux qui passent les contrats (les financeurs), les agents (les offreurs) et les bénéficiaires (les patients). Ceci a pour double conséquence que ce n'est pas le patient qui définit les termes du contrat avec le médecin mais le tiers-payeur et que ce dernier ne peut pas vérifier *ex post* la quantité et/ou qualité du service. Cette configuration du marché rend délicate une démarche dont l'objet serait fondamentalement la définition de contrats optimaux entre médecin et patient seulement ;

– outre cette fragmentation, se pose la question de produits joints : le médecin - *Agent* est sollicité à titre d'expert pour révéler un message sur la gravité de la maladie (diagnostic). Il est aussi sollicité pour l'effort qu'il va fournir (le temps et le nombre de consultations, par exemple). Message et effort sont ici des produits joints ;

– enfin et surtout, le même acteur peut entrer dans plusieurs relations d'agence. C'est le cas du médecin par exemple, agent du patient et de plus en plus, agent de la collectivité (représenté selon les systèmes par la tutelle, l'assureur public voire même l'assureur privé qui représente alors les intérêts du groupe d'assurés) qui lui confère la responsabilité de l'allocation de ressources rares. Dans la mesure où intérêt individuel et intérêts collectifs ne peuvent coïncider (du fait de la solvabilisation de la demande), le médecin est amené à effectuer des arbitrages qui sont en grande partie influencés à la fois par le mode de rémunération et par le degré de solvabilisation des patients.

Au vu de ces caractéristiques, la modélisation de l'interaction patient - médecin, si elle apparaît bien relever du cadre théorique du *Principal-Agent*, ne semble pas toutefois pouvoir reposer sur les contrats optimaux comme solution.

De fait, l'observation de la relation patient - médecin indique que le seul pouvoir dont dispose réellement le patient face à son médecin est le "vote avec les pieds", selon l'expression anglo-saxonne consacrée.

En conséquence, une solution alternative est constituée par la résolution d'un programme de recherche de marché en avenir incertain. Rochaix (1989) tente ainsi d'intégrer deux caractéristiques centrales du champ de la santé, à savoir l'incapacité du patient à observer la réalisation de la variable aléatoire "sévérité" et le rôle dominant du tiers-payeur. Dans le modèle, l'asymétrie informationnelle est définie comme une inégalité face à l'incertitude quant à la sévérité de la maladie. En effet, seul le médecin dispose d'une information grâce à ses connaissances et aux tests diagnostiques et ce différentiel d'information est la source même de son pouvoir discrétionnaire. La réduction de l'incertitude passe pour le patient par une recherche de marché et la règle d'arrêt optimale est définie comme l'obtention d'une valeur d'intensité de traitement acceptable (comprise entre une borne minimale et une borne maximale). La contrainte de marché est donc formalisée comme la "menace" *ex ante* exercée par le patient de refuser le traitement et de rechercher une deuxième opinion au cas où l'écart entre sa distribution subjective de sévérité et le diagnostic révélé par le médecin serait trop important. Le modèle tente aussi d'intégrer la variété des cadres institutionnels dans lesquels exercent les médecins, en particulier les deux systèmes-type de rémunération que sont le paiement à l'acte et le paiement forfaitaire⁽⁹⁾ et de montrer l'impact de l'assurance du patient sur les coûts de la recherche. Les résultats indiquent l'importance des phénomènes d'incertitude et d'asymétrie d'information qui conduisent le médecin à s'écarter d'une solution d'agence parfaite (définie comme la solution que retiendrait le patient s'il avait la même information que le médecin). Ils montrent aussi l'intérêt de la menace qui permet de se rapprocher de cette situation d'agence parfaite, ainsi que l'importance du cadre institutionnel de l'interaction médecin - patient. Un exercice de simulation permet d'étendre les prédictions du modèle au cas d'un marché avec plusieurs médecins et plusieurs patients, les résultats indiquant l'intérêt d'une information sélective des patients (il suffit en effet qu'un petit nombre de patients soit informé pour que la contrainte de marché joue), avec cependant un intérêt supérieur de la régulation de l'offre de soins comparé à celui de campagnes d'information des patients.

À l'évidence, le modèle n'épuise pas les multiples dimensions de l'incertitude médicale⁽¹⁰⁾. De plus, dans la mesure où il repose sur l'hypothèse de continuité de la variable traitement, il s'applique mieux à des décisions thérapeutiques concernant des intensités différenciées de traitement qu'à des décisions dichotomiques impliquant un choix entre, par exemple, une intervention chirurgicale et un traitement médicamenteux. Si les hypothèses en matière d'information du patient ont été relâchées de façon à prendre en compte la double incertitude du patient (sur la sévérité et sur la qualité du médecin

sollicité), la critique précédente de Mooney et Ryan (1993) - selon laquelle on part du postulat que les patients reconnaissent le besoin d'exercer cette contrainte et qu'ils en sont capables - demeure valide.

Une tentative de réinterprétation de ce modèle dans un cadre *Principal-Agent* standard a été offerte par Fermon (1991a). La menace de vote avec les pieds est réintégrée dans la fonction d'utilité du patient comme une utilité minimale tolérable en deçà de laquelle le patient consulte un autre médecin. L'auteur reconnaît toutefois explicitement que cette "application de la théorie *Principal-Agent* ne cherche pas à déterminer le mode de rémunération optimal de l'*Agent* par le *Principal*, puisque dans la relation médecin - patient, le *Principal* ne peut déterminer directement le montant du prix à payer". Il s'agit plutôt d'étudier l'influence de deux variables : le niveau de la demande et l'attitude du médecin face au risque et de mettre en évidence les comportements possibles du médecin compte tenu de ces deux paramètres. La pression forte de la demande y est perçue comme un facteur incitatif pour le médecin à se comporter de manière optimale. La recommandation principale est alors de définir la régulation de l'offre de soins sur une base géographique de façon à renforcer cette pression de la demande.

Béjean (1994) suggère pour sa part d'endogénéiser cette menace dans les modèles *Principal-Agent* par l'introduction d'une contrainte de participation du *Principal* à la relation⁽¹¹⁾. L'auteur note qu'à la différence des modèles *Principal-Agent* standard, ce n'est pas le *Principal* qui maximise son utilité sous contrainte de participation de l'*Agent* mais l'inverse. Mais on peut alors légitimement questionner la logique d'un modèle de délégation décisionnelle de la part du *Principal* dans lequel ce dernier s'interrogerait sur le fait même de participer au contrat qu'il est censé définir.

Ces deux tentatives de réintégration de l'interaction patient - médecin dans le cadre standard de la théorie de l'agence soulignent bien toute la difficulté de l'exercice. Quand bien même elles ont permis de réécrire sous forme de modèle *Principal-Agent* la relation patient - médecin, elles n'ont pas pour autant conduit à la définition de contrats optimaux de rémunération de l'*Agent*. De fait, l'endogénéisation de la menace conduit à faire dépendre la décision du médecin et de lui seul, alors que l'apport du modèle de recherche était de modéliser l'interaction indépendante de l'offre et de la demande et d'engendrer par là même des prédictions.

Une solution peut être trouvée dans le passage à la modélisation de la relation État (*Principal*) - offreur (*Agent*). Telle est l'approche adoptée par Fermon (1993) dans une analyse de l'enveloppe globale en ambulatoire, qui permet à l'auteur de définir un

contrat optimal. Les résultats indiquent en particulier que la rémunération individuelle doit être variable et non strictement proportionnelle d'une part et que l'enveloppe globale doit être appliquée à une collectivité de médecins (comme c'est le cas en Allemagne par exemple). Restent toutefois non résolus les problèmes de détermination des seuils de progressivité de la rémunération du médecin et les problèmes de détermination de la taille optimale du groupe auquel l'enveloppe doit être appliquée. Surtout, le modèle ne permet pas de montrer comment s'effectuent les arbitrages entre intérêts individuels et intérêt collectif.

Ainsi, la difficulté majeure rencontrée dans ces applications est l'éclatement en trois pôles de ce qui est traditionnellement analysé sous un échange bilatéral entre un *Agent* et un *Principal*. Toute tentative de modélisation de type *Principal-Agent* du marché des soins médicaux se doit donc de reconnaître cette spécificité et d'intégrer à la fois la relation médecin - patient et médecin - tiers-payeur. Il importe, lorsque l'on tente de définir des modes de rémunération optimaux, de prendre en compte, d'une part, la double relation d'agence du médecin qui le conduit à effectuer un arbitrage entre intérêt individuel de son patient et intérêt collectif et, d'autre part, les phénomènes de double risque moral (ou collusion implicite) qui existent sur le marché des soins médicaux entre patient et assureur et entre médecin et assureur⁽¹²⁾.

Par la suite, l'application des modèles *Principal-Agent* a été généralisée à d'autres relations et a fourni une grille de lecture étendue du marché des services médicaux⁽¹³⁾. Cette application s'est surtout inspirée de ce que Charreaux (1987) appelle la théorie "positive" de l'agence, dérivée de la théorie des organisations (par opposition à la théorie "normative" de l'agence, issue de la théorie des contrats⁽¹⁴⁾). La raison en est sûrement la plus grande flexibilité de ce type d'analyse dans laquelle, comme le note l'auteur, la notion d'agence est étendue à toute sorte de coopération sans qu'il y ait nécessairement un mandant (*Principal*) et un mandataire (*Agent*).

Plutôt que de présenter ici l'éventail des relations d'agence que l'on peut alors définir dans le domaine de la santé, on en établira plutôt une typologie à partir des travaux de Jones et Zanola (1995) qui repèrent les principaux acteurs entre lesquels vont s'établir des relations d'agence dans les systèmes de santé. Cinq groupes d'acteurs sont retenus :

- les représentants de l'autorité de tutelle (politiques, décideurs) ;
- le public dans son double rôle de citoyen et de patient ;
- les assureurs, tant publics que privés, obligatoires ou volontaires, au premier franc ou complémentaires ;

- les acheteurs, et tout particulièrement en Grande-Bretagne : les autorités de district et les médecins gestionnaires de budget ;
- les producteurs de soins, incluant des organisations comme l'hôpital, mais aussi les médecins et autres professionnels de santé comme l'industrie pharmaceutique.

De nombreuses relations d'agence peuvent être définies à partir de ces cinq grandes catégories d'acteurs. On peut de surcroît ajouter quelques relations supplémentaires à l'intérieur de certaines d'entre elles. Ainsi :

- afin de prendre en compte les particularités du cadre français, on pourrait analyser, comme le suggère Béjean (1992), à l'intérieur de la première catégorie, la prise en compte d'un intermédiaire : l'assurance maladie, entre médecin et tutelle. Ceci permettrait d'analyser les relations complexes que nouent ces deux institutions dans leur tentative de régulation du secteur de la santé ;
- en ce qui concerne la cinquième catégorie, de nombreuses voies pourraient être explorées : Scott (1996) propose d'analyser les relations entre syndicats de professionnels de santé, responsables de la négociation collective et tiers payeur ; Fermon (1991a, 1991b) analyse entre autres les relations entre corps médical et administration à l'intérieur de l'hôpital ; Gaynor (1994) évoque quant à lui les relations d'agence à l'intérieur d'associations de médecins de type *HMO* aux États-Unis.

L'application des modèles *Principal-Agent* au domaine de la santé a donc connu son heure de gloire et le nombre de relations d'agence entre ces grandes catégories d'acteurs et à l'intérieur de chaque catégorie apparaît virtuellement sans limites. Ce modèle connaît toutefois une remise en cause qu'il importe de retracer.

Les remises en cause de la validité de l'application

L'argumentation, tant du côté anglo-saxon (Mooney et Ryan, 1993) que français (Béjean, 1994) s'établit en deux temps. Elle tient, d'une part, au fait que ces tentatives, dans la plupart des cas, sont restées dans le domaine de la description des comportements plutôt que de la définition de contrats optimaux. Elle tient, d'autre part, à ce que, pour les quelques cas où un contrat optimal a été défini, il s'est avéré beaucoup plus complexe que ceux que la réalité offre.

En ce qui concerne le premier temps, des éléments de réponse sont dans les traits distinctifs du champ d'application mentionnés précédemment. En particulier, devant la complexité des relations nouées entre ces nombreux acteurs, les auteurs ont, dans la plupart des cas, privilégié la démarche extensive, conduisant à recenser toutes les relations d'agence possibles, plutôt que la démarche intensive qui aurait permis d'approfondir l'analyse des

relations d'agence les plus importantes et surtout leur interaction.

Pour ce qui est de la trop grande complexité des contrats définis, il convient ici de souligner que ce fait stylisé n'est pas particulier à la santé. Arrow (1986) suggère en effet que les développements futurs de la théorie des contrats devraient mieux rendre compte de ce que les contrats effectivement passés sont plus simples que ceux que la théorie suggère. Cette simplicité s'explique, selon lui, par l'importance des coûts de la contractualisation, de même que ceux de suivi (parfois mal définis) et par le fait que les incitations positives et négatives sont souvent autres que pécuniaires. Les développements récents en matière de contractualisation intègrent cette critique⁽¹⁵⁾.

C'est ce tout dernier point que Mooney et Ryan retiennent comme explication possible de ce phénomène de complexité. Ils argumentent de ce que, comme le notait Evans dès 1974, dans le domaine de la santé, les fonctions d'utilité sont interdépendantes entre *Principal* et *Agent*, à la différence du modèle standard, en particulier, que la fonction d'utilité du patient constitue un argument supplémentaire dans celle du médecin. En conséquence, la contrainte de compatibilité du mécanisme incitatif peut être saturée par l'introduction de cet argument supplémentaire dans la fonction d'utilité du médecin. Une bonne connaissance des préférences du médecin et du patient⁽¹⁶⁾ apparaît alors aux deux auteurs comme un préalable indispensable à l'application des modèles *Principal-Agent*.

Scott (1996) reprend et précise l'analyse de Mooney et Ryan en soulignant quelques traits distinctifs supplémentaires :

- l'effort dans la fonction d'utilité du médecin devrait selon lui entrer positivement et non négativement comme c'est en général le cas, de façon à appréhender l'idée selon laquelle le médecin tente de faire de son mieux pour le patient. La dérivée seconde serait par contre négative comme indication de rendements décroissants ;
- le résultat n'est pas de nature financière mais s'interprète en terme d'amélioration de l'état de santé ;
- enfin, la contrainte de participation peut être saturée par le simple fait que dans beaucoup de pays, les médecins n'ont d'autre alternative que d'honorer le contrat global signé entre le tiers-payeur et les représentants de la profession.

Ces analyses des difficultés rencontrées dans les premières applications des modèles *Principal-Agent* au domaine de la santé sont utiles car elles permettent d'identifier les points sur lesquels pourrait éventuellement porter l'adaptation du modèle *Principal-Agent* standard⁽¹⁷⁾. En particulier, elles nous incitent à dépasser le cadre trop simplifié

des modèles bilatéraux de la théorie des contrats. Si ceux-ci ont permis une bonne compréhension des relations contractuelles entre deux individus, cette forme d'organisation demeure néanmoins peu réaliste. L'étape suivante devra en conséquence porter sur l'étude des organisations à trois individus, voire plusieurs individus. Ceci permettra de parer la critique des modèles *Principal-Agent* formulée par Sappington (1991) et qui s'applique particulièrement bien au champ de la santé, à savoir, une tendance à analyser chaque relation d'agence de manière trop indépendante et isolée.

Vers les modèles plus complexes de la théorie des contrats

Les travaux récents d'application à la santé de la théorie des contrats se tournent donc vers des familles de modèles certes plus complexes, mais mieux adaptées aux traits distinctifs évoqués précédemment. Chacune d'entre elles permet de prendre en compte un trait spécifique du champ d'application, sans toutefois qu'une présentation unifiée ne soit véritablement envisageable, du fait de la complexité impliquée. Quatre extensions sont offertes, la première constituant une simple modification du modèle bilatéral, les trois suivantes constituant par contre l'ouverture à de nouvelles familles de modèles de la théorie des contrats.

Les modèles "*multi-task*"

Ces modèles appartiennent toujours à la famille des modèles bilatéraux (un *Principal*, un *Agent*) mais représentent une extension par la prise en compte dans le contrat de la possibilité de tâches multiples demandées à l'*Agent*.

Holmstrom et Milgrom (1991) identifient les cas où le producteur prend des décisions concernant à la fois le coût et la qualité, sous le terme d'agence à fonctions multiples (*multitask agency*). Les modèles se différencient ici en fonction des hypothèses sur la capacité du *Principal* à observer la qualité et donc à exercer une contrainte sur le producteur. Deux situations sont traditionnellement envisagées : le cas de tâches complémentaires et le cas où elles sont substitués. Les résultats indiquent que dans le premier cas (par exemple, un investissement important dans la première tâche permet de rendre la deuxième moins coûteuse) il convient de faire porter la rémunération sur la première tâche qui donne lieu à un signal observable par le *Principal*. Dans le cas de biens substitués parfaits, inciter l'*Agent* à accomplir une des deux tâches aboutit à le décourager d'entreprendre l'autre. Le *Principal* doit alors rendre le mécanisme moins incitatif.

Le dilemme prix/qualité lié à l'hétérogénéité des biens est central dans le cadre de la santé. Ainsi que le note Mougeot (1994), "dans une situation de risque de moralité, il est immédiat que toute incitation à la réduction des dépenses est aussi une incitation à la réduction de la qualité. Les incitations jouent donc en sens opposé sur les deux variables essentielles : coût des soins et qualité des soins".

Certains travaux explorent les modes de rémunération entre acheteurs (que ce soit l'autorité de tutelle, les compagnies d'assurance ou les *HMO*) et producteurs de services, permettant de réduire les coûts sans réduire la qualité. La difficulté réside bien ici dans le fait que l'effort du producteur n'est pas observable et que la qualité est difficile à mesurer. Parmi ceux-ci, Chalkley et Malcomson (1995a) offrent dans un premier temps une généralisation de résultats antérieurs précisant sous quelles conditions minimales un système de type *fixed price* pur (paiement prospectif ou contrat au coût par cas pour la Grande-Bretagne) permet de traiter le nombre optimal de patients à un niveau de qualité efficient, avec le niveau efficient d'effort de minimisation des coûts de la part du producteur. Ils soulignent surtout le fait que ce résultat ne tient que si la fonction de demande agit effectivement comme contrainte de marché sur le producteur. Or cette hypothèse peut se révéler non vérifiée dans le domaine de la santé pour deux raisons :

- dans certains systèmes de santé, le côté court du marché est l'offre plus que la demande, de par l'insuffisance des moyens à la disposition du producteur (c'est le cas en Grande-Bretagne pour certains services hospitaliers qui doivent pratiquer une forme de rationnement par file d'attente) ;
- les patients sont solvabilisés, de telle sorte que la contrainte de demande ne s'exerce pas pleinement. On peut s'attendre à ce que la consommation s'étende au-delà du niveau d'égalité entre bénéfices et coûts, jusqu'au point où les bénéfices sont nuls.

L'efficacité des systèmes de prix fixes purs peut néanmoins être garantie, sans introduction de mécanisme de surveillance, à condition que le nombre de patients désireux d'être traités et effectivement traités soit observable et que le mécanisme de paiement tienne compte de la différence éventuelle entre ces deux grandeurs. Par contre, dès que l'on introduit le caractère multidimensionnel de la qualité, cette efficacité est remise en cause, sauf si l'on peut considérer que le patient et le producteur ont les mêmes ordres de préférences sur les dimensions de la qualité. Si ceci n'est pas vérifié (ce qui paraît plus que probable), il faut alors définir un contrat plus complexe que celui qui prévaut avec une seule dimension de qualité.

Dans un deuxième article (1995b), ces mêmes auteurs reprennent l'analyse en postulant différentes fonctions objectifs selon que le producteur est à but lucratif ou non. Ceci leur permet de prendre en

compte une autre caractéristique du domaine de la santé soulignée à maintes reprises⁽¹⁸⁾, à savoir l'importance accordée par le producteur au bien-être du patient. Ainsi, les organisations à but non lucratif retireront une utilité directe de la quantité et la qualité des services produits. Trois cas de figure sont envisagés : un établissement altruiste, partiellement altruiste et égoïste (simple maximisation du profit). Les résultats varient selon quatre dimensions : selon la nature de l'établissement, selon l'importance respective accordée par la tutelle à l'objectif de qualité *versus* réduction de coût, selon la capacité pour l'établissement d'emprunter pour couvrir ses déficits et selon que le nombre de patients à traiter est connu au moment de signer le contrat.

Parmi les résultats les plus importants, on notera que le système de paiement à la pathologie (*cost per case*) est optimal si le producteur est égoïste et si la tutelle privilégie l'objectif de réduction des coûts sur celui de l'augmentation de la qualité au-delà du minimum requis légalement. Si l'on introduit l'incertitude au moment de la signature du contrat sur le nombre de patients à traiter et sur la capacité de l'établissement à emprunter pour couvrir un déficit temporaire, des systèmes de type *cost and volume* sont préférables. Les résultats antérieurs d'Ellis et McGuire (1990) selon lesquels un paiement mixte offrant un remboursement partiel des coûts observés et un paiement forfaitaire *per capita* est optimal dans le cas d'un producteur partiellement altruiste sont corroborés. Enfin, le remboursement à 100 % des coûts observés n'est jamais optimal.

Ces premiers travaux d'application de modèles de type "multi-tâches" permettent d'intégrer deux éléments fondamentaux, à savoir le dilemme prix/qualité et la nature altruiste de la fonction-objectif de l'établissement. Les applications au domaine de la médecine ambulatoire restent à définir.

Les modèles multi-principaux

Les modèles multi-principaux constituent pour leur part une nouvelle famille de modèles puisqu'il s'agit de la modélisation d'un *Agent* face à deux ou plusieurs *Principaux*. Cette famille semble en effet particulièrement bien répondre aux conditions d'une modélisation adéquate du marché des services médicaux posées précédemment. Pourtant, peu de travaux ont tenté d'appliquer ces modèles, encore appelés modèles d'agence commune.

Bien que le modèle de Blomqvist (1991) ne s'inscrive pas explicitement dans un modèle d'agence commune, il offre une première piste de réponse en matière de prise en compte des deux relations d'agence entre médecin et patient et entre médecin et assureur (ou tutelle). Il a pour principal objet de clarifier les avantages d'un système de réseaux intégrés de soins (*HMO*) avec salariat des

médecins par rapport à des systèmes d'assurance plus conventionnels avec paiement à l'acte des médecins. L'auteur considère trois acteurs : le médecin, le patient et l'assureur, avec des niveaux d'information différents. Les médecins peuvent établir un diagnostic exact sur la pathologie ; les patients quant à eux se savent malades mais ne connaissent pas la nature exacte de la maladie ; les assureurs sont les moins bien dotés puisqu'ils ignorent même si le patient est ou non malade. Deux équilibres sont tour à tour analysés selon le système de rémunération envisagé pour les médecins : le paiement à l'acte et le salariat en *HMO*. Dans le premier cas, le médecin aura tendance à surfacturer ses propres services (risque moral côté offre), tendance que le patient ne contrôlera pas, voire renforcera si le prix des soins reçus est réduit par la présence de l'assurance (risque moral côté patient). Les résultats indiquent que le fonctionnement en *HMO* résout les problèmes d'asymétrie d'information entre médecin et assureur. Par contre, les problèmes d'asymétrie d'information entre médecin et patient ne sont pas pour autant résolus et un mécanisme de régulation supplémentaire doit être introduit. L'auteur envisage dans le cas des *HMO* des systèmes de garanties de type légal contraignant ces dernières à prendre en compte les préférences de leurs patients. Un tel mécanisme devrait permettre de garantir une relation d'agence parfaite et donc de s'assurer que le niveau d'utilisation des soins correspond bien à ce que le patient aurait consommé s'il avait eu la même information que le médecin. Dans le cas du paiement à l'acte, le mécanisme supplémentaire est destiné quant à lui à résoudre les problèmes de risque moral dans la relation médecin - assureur. Il est par contre plus difficile à mettre en œuvre puisqu'il s'agit d'imposer une amende au médecin lorsque l'état de santé du patient est supérieur à celui auquel on pourrait s'attendre, étant donné l'information et les soins reçus.

D'après M. Gaynor (1994), la préoccupation principale de Blomqvist est surtout une prise en compte explicite des modalités de rémunération, sans véritable modélisation en terme d'agence de la relation entre assureur et médecin ou entre patient et médecin (en particulier, il n'est pas démontré que les deux stratégies du médecin sont les meilleures réponses possibles et la question de l'équilibre de marché n'est pas abordée). Surtout, le modèle n'offre pas de solution opérationnelle pour le cas du paiement à l'acte. Dans le cas *HMO*, il permet tout au plus de "forcer" les intérêts des deux principaux que sont la *HMO* et le patient à être identiques par l'introduction d'une contrainte supplémentaire. Or les intérêts de la *HMO* entrent nécessairement en conflit avec ceux du patient puisque celui-ci est assuré et ne paye qu'une partie des frais. De fait, supposer, comme le fait l'auteur, que la *HMO* peut être contrainte à rechercher la satisfaction de ses patients revient à contourner le problème du

rationnement nécessaire au niveau collectif. Nonobstant ces limites, ce modèle constitue une première tentative de modélisation en agence commune et a l'avantage de souligner l'existence d'un phénomène de double risque moral qui nécessite dans ce modèle l'adoption de deux contraintes indépendantes.

Les modèles multi-agents

Cette famille de modèles implique pour sa part un *Principal* face à plusieurs *Agents*. Ils s'inspirent fortement du cadre théorique défini par Laffont et Tirole (1986) pour la prise en compte des effets de l'asymétrie de l'information dans la réglementation des monopoles naturels. Ce modèle combine l'information cachée sur les coûts et l'action cachée en postulant que la firme peut modifier les coûts par son niveau d'effort, mais l'hypothèse est faite que le régulateur observe les coûts *ex post* sous certaines conditions. La réglementation optimale conduit à isoler les problèmes de tarification des problèmes d'incitations (*incentive pricing dichotomy*). Ces dernières sont optimisées par des règles de transferts, fonction des coûts observés et la tarification est de type Ramsey et Boîteux dans laquelle le coût social des fonds publics est explicitement pris en compte. Il faut alors arbitrer entre des transferts forfaitaires qui induisent de hauts niveaux d'effort mais abandonnent des niveaux de rente élevés aux firmes et des transferts incluant le remboursement des coûts qui n'induisent aucune incitation à l'effort mais permettent de contrôler les rentes des firmes.

La difficulté réside dans le fait que la mise en place d'incitations fortes pour encourager l'effort de la part de la firme moins efficace risque d'être coûteuse s'il s'avère que la firme est en fait du type efficace (le régulateur ne connaissant pas ce paramètre au moment de la signature du contrat). En conséquence, il convient de limiter les incitations, ce qui réduit d'autant la motivation de la firme efficace. En général, des contrats incitatifs intermédiaires sont donc adoptés qui partagent entre le régulateur et la firme les écarts entre coûts annoncés et coûts réalisés⁽¹⁹⁾. L'apport principal de ce type de modèle⁽²⁰⁾ est donc de souligner le fait que "les asymétries d'information et les comportements stratégiques des agents impliquent de nouvelles contraintes sur l'allocation des ressources, ici un lien inévitable entre les rentes informationnelles abandonnées aux entreprises et les niveaux de production réalisables" (Laffont, 1993).

Outre ces solutions pour réduire les rentes informationnelles que sont les transferts et les mécanismes de prix (solutions issues de la théorie de la réglementation des monopoles naturels), s'ouvre l'option de mise en concurrence des agents.

Mougeot (1994) offre une revue exhaustive des possibilités d'adaptation au marché de la santé des

solutions offertes par la mise en concurrence. Comme le note l'auteur, "la fixation de prix optimaux pour un agent central se heurte aux difficultés informationnelles d'évaluation du coût". La concurrence fictive et la concurrence organisée permettent de surmonter ce handicap pour obtenir la meilleure allocation possible des ressources.

La première "consiste à mettre en œuvre un schéma de fixation des prix unitaires plaçant chaque offre en situation de concurrence fictive de marché, sans qu'effectivement un marché ne soit instauré". Deux exemples d'application sont offerts : la globalisation de l'enveloppe (qui conduit bien à fixer le prix unitaire de chaque acte en fonction de l'ensemble des actes fournis pour une période et une zone géographique donnée - il s'agit alors d'une concurrence par les quantités) et la tarification à la pathologie qui constitue une forme de concurrence par comparaison (*yardstick competition*).

La deuxième conduit à mettre en place un quasi-marché que l'État peut organiser pour que les gains en efficacité productive ne soient pas minorés par des phénomènes comme l'écrémage des risques, par exemple. La mise en concurrence peut s'appliquer aux offreurs de soins (cas des réformes britanniques) ou aux assureurs (cas de la réforme hollandaise).

Mougeot et Naegelen (1997) analysent les relations tutelle-hôpitaux et offrent une comparaison des avantages relatifs de chacune de ces deux formes de concurrence, à savoir la tarification à la pathologie (concurrence fictive) et l'achat de soins par la tutelle (concurrence organisée). L'hypothèse est faite que "le but de la réglementation est le contrôle du prix unitaire : la tutelle considère des quantités (qualités) fixes et connaît la valeur sociale des soins". Le raisonnement s'établit pathologie par pathologie, sans envisager les arbitrages prix/qualité ou prix/quantité. Les résultats indiquent que les deux formes de concurrence permettent d'atteindre un niveau de bien-être collectif supérieur au contrat bilatéral optimal, la seconde l'emportant sur la première sous certaines conditions. Ces dernières portent sur le caractère effectif de la concurrence et sur le fait que les croyances du régulateur sont identiques dans les deux cas.

Si la mise en concurrence des assureurs aux Pays-Bas a encore peu fait l'objet de modélisations, une application récente du modèle de Laffont et Tirole a été réalisée par Pellisé (1994) sur la mise en place d'un système de santé expérimental en Espagne pour les fonctionnaires et leurs ayants droit. Les assureurs sont mis en concurrence sur un marché interne et reçoivent un paiement forfaitaire *per capita*. Les résultats permettent de retrouver les conclusions du modèle général concernant l'écrémage du risque.

La structure des relations nouvellement créées par la mise en place de quasi-marchés en

Grande-Bretagne a quant à elle fait l'objet d'analyses récentes. Sont plus particulièrement analysés les trois types de contrats développés entre acheteurs et fournisseurs depuis la mise en place des quasi-marchés : les contrats globaux (*block contracts*), les paiements à la pathologie (*cost per case*), tous deux de type *fixed price* et les contrats coût-volume (*cost and volume contracts*) qui sont de type *cost plus*.

Propper (1995) analyse la structure des incitations créées par la mise en place des réformes de 1991 et montre que le marché des services médicaux peut être caractérisé par un ensemble de relations *Principal-Agent* qui se superposent. L'analyse approfondie de ces relations montre que l'efficacité attendue de la mise en place de mécanismes de concurrence est limitée :

– côté acheteurs, les autorités locales de district (*DHA-District Health Authorities*-) et les médecins gestionnaires de budget (*GP Fundholders*) sont dans une double relation d'agence : envers le gouvernement, qui leur transfère ses compétences en matière d'allocation de ressources et envers le patient, auprès duquel ils ont la responsabilité de l'achat de services de santé. L'auteur note que ces relations *Principal-Agent* sont de nature hiérarchique, de telle sorte que des opportunités de coopération et de collusion existent entre acheteurs (*Agents*) à l'encontre du gouvernement (*Principal*). La mise en place de mécanismes de surveillance réduit les gains attendus de l'introduction de cette concurrence entre acheteurs en matière d'efficacité productive et allocative ;

– côté producteurs, la relation d'agence entre gouvernement et producteurs (ici les *trusts*, qui sont des établissements hospitaliers de statut proche des *Epic*) est affectée par le manque de précision sur les droits de propriété, en particulier en ce qui concerne les économies réalisées.

Parmi les modèles de contrats optimaux développés autour des réformes britanniques, celui de Fenn, Rickman et McGuire (1994) est centré sur la relation entre le *DHA* et les offreurs de soins (les *trusts*). La question est de savoir si les acheteurs de soins font effectivement jouer la concurrence ou bien s'ils s'enferment dans des contrats passés systématiquement avec les mêmes offreurs de soins (*lock in*). Les auteurs montrent que cette tendance dépendra des conditions de régulation mises en place. La nature du contrat sera ainsi fonction d'un certain nombre de paramètres dont, entre autres le niveau du coût de transaction lié au recours à un hôpital extérieur au District et l'importance des phénomènes de capacité excédentaire dans les hôpitaux locaux.

Les modèles avec tiers (ou hiérarchiques)

La dernière famille de modèles considère de manière explicite l'introduction d'un tiers entre tutelle et agent, ce qui conduit à une structure hiérarchique de

type : tutelle/superviseur/producteur. Cette introduction est liée à l'asymétrie d'information entre gouvernement et producteurs de services (par exemple sur le ratio coût/qualité), le superviseur étant supposé combler ce déficit informationnel. Ce dernier dispose toutefois d'un pouvoir discrétionnaire important et peut en conséquence faire l'objet d'une "capture" par le producteur de services (*regulatory capture*). La perspective de collusion entre superviseur et producteur est au cœur des préoccupations de ces modèles.

Barrow (1996) ouvre des perspectives en matière d'application à la production et la distribution de services d'éducation et de santé de ces modèles hiérarchiques qui impliquent le gouvernement, le superviseur et le producteur. Dans ce cadre plus complexe, le gouvernement doit définir une charte ("constitution") qui s'adresse au superviseur qui doit, à son tour, réguler le producteur. Ce type de structure hiérarchique conduit à explorer des situations de collusion entre firme et superviseur à l'encontre du gouvernement. Dans de tels modèles, il convient d'ajouter aux contraintes d'incitation individuelles une contrainte de collusion. Ceci permet par exemple d'analyser le favoritisme dans les enchères, ou encore le problème de la décentralisation des contrats incitatifs dans les hiérarchies (Laffont et Martimort, 1995).

Certaines des perspectives d'application au système de soins de ces modèles ont été évoquées par Propper (1995) dans le cadre des réformes du *NHS*. Elles pourraient toutefois faire l'objet d'applications beaucoup plus larges. En France par exemple, les relations entre tutelle, assurance maladie et producteurs de soins pourraient être fructueusement analysées à travers ce prisme.

Se pose aussi la question de savoir s'il est dans l'intérêt de la tutelle de déléguer à plusieurs superviseurs la responsabilité de la régulation, lorsque l'asymétrie d'information porte sur plus d'une variable (efficacité et qualité, par exemple). À cette interrogation, Tirole (1994) répond par l'affirmative, quand bien même ceci risque d'entraîner des défauts de coordination entre ces deux agences de régulation. Tel est l'objet du modèle de Rizzo et Sindelar (1996) qui offre une nouvelle voie d'explication des phénomènes de surconsommation dans le secteur de la santé en analysant l'impact, en terme de bien-être, des défauts de coordination entre deux agences indépendantes de régulation. Le contexte d'application est celui des services médicaux du programme *Medicare* régulés par deux instances : l'une déterminant les prix, l'autre fixant les normes de qualité. Les auteurs mettent en évidence d'une part des niveaux excessifs d'utilisation et d'autre part des niveaux de prix et de qualité (en termes de caractéristiques désirables de l'offre de soins comme la proximité, l'accueil) inférieurs à ceux qui prévaudraient sans défauts de coordination.

Conclusion

À travers l'application des modèles issus de la théorie des contrats, les économistes de la santé ont tenté de mieux comprendre les stratégies des acteurs du système de santé et l'impact de la rente informationnelle dont certains d'entre eux disposent. Ils ont tenté, ce faisant, d'établir un diagnostic affiné des dysfonctionnements des systèmes de santé. En matière de détermination de modes de rémunération optimaux, en particulier en ce qui concerne les avantages et inconvénients des systèmes de type *cost-plus* versus *fixed price* les avancées sont probantes et plus marquées pour les établissements que pour la médecine ambulatoire ou l'assurance.

Cependant, aucun modèle ne semble à même de prendre en compte des phénomènes de durée, alors que les effets de réputation sont centraux sur ce marché des services de santé. De fait, des modèles dynamiques (Rubenstein and Yaari, 1983) permettent aujourd'hui d'analyser les phénomènes d'aléa de moralité ou d'antisélection dans le cas d'interaction répétée⁽²¹⁾. La perspective de dynamiser certaines des relations mises en évidence sur le marché des services médicaux apparaît alors particulièrement riche.

En dernier ressort, il convient toutefois de s'interroger sur le bien fondé de l'application au secteur de la santé, d'instruments développés dans un premier temps soit pour l'analyse de la régulation des monopoles naturels soit pour des relations bilatérales privées. Barrow (1996) note à cet égard que les différences qui existaient au préalable dans le fonctionnement des secteurs public et privé, en particulier en ce qui concerne la fonction - objectif (l'objectif de maximisation du profit n'étant pas nécessairement partagé par les services publics) tendent à s'estomper. En effet, dans la plupart des cas, des budgets prospectifs sont mis en place dans le secteur public, avec des sanctions en cas de dépassement. Mais l'auteur s'interroge sur la présence d'un sens de l'altruisme plus marqué dans le secteur public, qui atténuerait la nécessité de définir des mécanismes incitatifs. Ceci rejoint la notion d'atmosphère de Williamson (1975) ainsi que la relation de don (*gift relationship*) de Titmuss (1970). Si de telles considérations sont marquées dans le secteur public, il importe de les prendre en compte à la fois au plan théorique (ainsi que le font Chalkley et Malcomson) et dans la définition de réformes institutionnelles de telle sorte que les changements d'incitations visant à améliorer la relation *Principal-Agent* ne modifient pas de manière significative cette "atmosphère". C'est bien en ce sens qu'une certaine complémentarité peut émerger entre théorie des contrats et l'approche par les coûts de transaction et par les conventions.

Notes

- (1) Ainsi, selon Gaynor (1994), Adam Smith serait le précurseur de l'analyse du fonctionnement imparfait du marché des soins de santé. Il aurait en effet souligné l'importance des effets de réputation et des barrières à l'entrée.
- (2) Cf. Akerlof (1971) et Pauly (1974).
- (3) Cf. Rochaix et Jacobzone (1997), dans ce même numéro pour une revue empirique de la littérature sur ce thème.
- (4) Hay et Leahy (1982) prennent le contre-exemple suivant afin de mettre en lumière le fait qu'une augmentation de la densité médicale n'entraîne pas toujours une perte d'information sur la réputation du médecin : lorsqu'un marché local ne comporte qu'un seul médecin qui est un charlatan, l'arrivée sur ce marché d'un autre médecin compétent augmentera fort probablement l'information des patients sur la qualité des services du premier médecin au lieu de la réduire.
- (5) "*The physician - patient exemplifies the principal - agent relation almost perfectly. The principal (patient) is certainly unable to monitor the efforts of the agent (the physician). The relation between effort and outcome is random, but presumably, there is some connection*".
- (6) Cf. Salanié B. (1995).
- (7) De fait, les exemples d'aléa moral sont si nombreux dans la réalité que le modèle d'aléa moral est parfois identifié au modèle principal - Agent, comme le fait D. Sappington (1991) dans sa revue de la littérature sur ce sujet.
- (8) Cf. Rochaix L. (1986).
- (9) Le premier s'apparente à des systèmes de remboursement *ex post* des dépenses (*cost-plus*), le deuxième étant basé sur une rémunération forfaitaire définie *ex ante* (*fixed price*).
- (10) Une autre dimension de l'incertitude a été analysée dans le cadre d'un modèle d'équilibre partiel du médecin par Eeckhoudt L., Lebrun T. et Saily J.C (1985). L'incertitude porte dans ce cas sur le diagnostic et la décision thérapeutique, l'aléa pour le médecin étant l'état de santé du patient.
- (11) Cf. S. Béjean, 1994, p. 199 et 204.
- (12) De fait, Gaynor (1994) dans ses recommandations de recherches futures indique en tout premier la modélisation jointe des problèmes d'agence entre le patient et son médecin et l'assureur avec le médecin. Selon les termes mêmes de l'auteur, *this is a fascinating and difficult problem, and will undoubtedly take considerable ingenuity*, p. 43.
- (13) Cf. Dranove D. et White W.D., (1987).
- (14) B. Fermon (1991a) puis S. Béjean et M. Gadreau (1992) ont appliqué cette distinction aux travaux sur la théorie de l'agence en santé. Cependant, G. Charreaux reconnaît lui-même que la distinction peut paraître arbitraire : selon lui en effet, la seconde, formalisée et pour l'essentiel due à des économistes, n'est pas dépourvue d'aspects positifs (tentatives d'explication de l'existence de franchises dans des contrats d'assurance, par exemple). La première, issue de l'analyse financière, "utilise le cadre de raisonnement de la théorie de l'agence pour expliquer le comportement réel des organisations et plus particulièrement des sociétés privées". Elle n'est pas dépourvue pour autant d'implications idéologiques : "les conclusions de la théorie positive de l'agence en matière de théorie de la firme sont d'essence libérale" et elle "conduit à des conclusions normatives". Il apparaît donc plus précis de se référer à l'axiomatique différente qui caractérise ces deux approches (théorie des organisations dans le premier cas, théorie des contrats dans le deuxième), plutôt que de les opposer sur la dimension positif *versus* normatif. La terminologie de "*Principal-Agent*" sera donc utilisée ici pour signifier l'application au domaine de la santé de ce "jeu de clefs" spécifique de la boîte à outils des

économistes (au sens où l'entendait Robinson J.). Cette clarification sémantique aura de surcroît l'intérêt de souligner que l'application des modèles *Principal-Agent* doit nécessairement recouvrir ces deux dimensions intrinsèquement liées que sont le positif et le normatif. Quel serait en effet l'intérêt d'une théorie dont le seul pouvoir serait normatif, sans capacité d'explication de la réalité et vice versa.

(15) À titre d'illustration, les travaux de Gasmi F., Laffont J.J. et Sharkey W. (1996) définissent des contrats optimaux basés sur des règles simples de remboursement linéaire des coûts.

(16) Tel est l'objectif des recherches empiriques menées par S. Vick et A. Scott (1995).

(17) Mais elles ne sont pas sans rappeler certaines tentatives qui ont conduit par le passé à déclarer que la santé est spécifique et ne saurait être comprise à travers le prisme de l'analyse économique. À titre d'illustration, attendre d'identifier précisément les arguments de la fonction d'utilité du médecin pour pouvoir définir un modèle *Principal-Agent* adapté risque en effet de retarder, voire de mettre en cause une telle application.

(18) Cf. Ellis et McGuire, (1990) et Glazer et McGuire (1994).

(19) Soit T le transfert reçu par la firme : $T = a + (1 - b)(C_r - C_a)$, où a est une somme fixe, C_r est le coût réalisé, C_a le coût annoncé et $b \in (0, 1)$. Si $b = 0$, on retrouve le système de *cost-plus* et inversement si $b = 1$. Le paramètre b mesure la proportion du coût non remboursée et est identifié au pouvoir incitatif du mécanisme. Par delà les solutions offertes par les mécanismes avec transferts, on peut aussi évoquer les mécanismes sans transferts où le prix est fonction du coût moyen, par exemple. Dans ce cas toutefois, les problèmes de tarification et d'incitation sont liés.

(20) Cf. Demsky J., Sappington D. (1991).

(21) Cf. Baron, Besanko, (1987).

Bibliographie

Akerlof G.A. (1971). "The Market for Lemons: Quality Uncertainty and the Market Mechanism", *Quarterly Journal of Economics*, vol. 74, pp. 488-500, août.

Arrow K. (1963). "Uncertainty and the Welfare Economics of Medical Care", *AER*, n° 53.

Arrow K.J. (1986). "Agency and the Market, in: K.J. Arrow and M.D. Intriligator, eds.", *Handbook of Mathematical Economics*, vol. III (Elsevier, Amsterdam)

Barrow M. (1996). "Public Services and the Theory of Regulation", *Policy and Politics*, vol. 24, n° 3, pp. 263-276.

Baron D., Besanko D. (1987). "Commitment and Fairness in Dynamic Regulatory Relationship", *The Review of Economic Studies*, vol. 54, pp. 413-436.

Baron D., Myerson R. (1982). "Regulating a Monopolist with Unknown Costs", *Econometrica*, vol. 50, pp. 911-930.

Béjean S. (1992). *Fonctionnement et régulation du système de santé*, thèse de doctorat en sciences économiques, décembre.

Béjean S. (1994). "Économie de la santé", *Economica*.

Béjean S., Gadreau M. (1992). "Asymétries d'information et régulation en médecine ambulatoire", *Revue d'Économie Politique*, mars/avril.

Bernheim B., Douglas M., Whinston M. (1986). "Common Agency", *Econometrica*, pp. 923-942, juillet.

Blomqvist A. (1991). "The Doctor as Double Agent : Information Asymmetry, Health Insurance, and Medical Care", *Journal of Health Economics*, vol. 10, pp. 411-432.

Chalkley M., Malcomson J.M. (1995a). "Contracting for Health Services When Patient Demand Does not Reflect Quality", *Discussion Paper*, n° 9514, Department of Economics, University of Southampton.

Chalkley M., Malcomson J.M. (1995b). "Contracting for Health Services with Unmonitored Quality", *Discussion Paper*, n° 9510, Department of Economics, University of Southampton.

Charreaux G. (1987). "La théorie positive de l'agence : une synthèse de la littérature". In Charreaux G. *et alii*, De nouvelles théories pour gérer l'entreprise, *Economica*.

Coyte P.C. (1984). "The Market for Medical Services and Physicians : an Application of Hedonic Price Theory", *Research Paper*, Department of Economics, University of Alberta.

Demsky J., Sappington D. (1991). "On Optimal Incentive Contracts with Multiple Agents", *Journal of Economic Theory*, vol. 17, pp. 152-171.

Dionne G., Contandriopoulos A.P. (1984). "Doctors and their Workshops : a Review Article", Département d'administration de la santé, Université de Montréal.

Dranove D. (1988). "Demand Inducement and the Physician/Patient Relationship", *Economic Inquiry*, vol. 24, pp. 281-298.

Dranove D., White W.D. (1987). "Agency and the Organisation of Health Care Delivery", *Inquiry*, vol. 24, pp. 405-415.

Eeckhoudt L., Lebrun T., Saily J.C. (1985). "Risk Aversion and Physicians Medical Decision-Making", *Journal of Health Economics*, vol. 4, pp. 273-281.

Ellis R.P., McGuire Th. (1990). "Optimal Payment Systems for Health Services", *Journal of Health Economics*, vol. 9, n°4, pp. 375-396.

Evans R. (1974). "Supplier-Induced Demand : Some Empirical Evidence and Implications", pp. 163-173, in

- Perlman M. the Economics of Health and Medical Care, Wiley and Sons.
- Evans R. (1984).** "Strained Mercy : the Canadian Health Care System", Brotherwell, Toronto.
- Fenn P., Rickman N., McGuire A. (1994).** "Contracts and Supply Assurance in the UK Health Care Market", *Journal of Health Economics*, vol. 13, pp. 125-144.
- Fermon B. (1991a).** "Application de la théorie de l'agence", thèse de doctorat, université de Paris-Dauphine.
- Fermon B. (1991b).** "Évaluation théorique du mode de financement hospitalier : apports de la théorie de l'agence à l'analyse du budget global", XI^{èmes} journées d'économie sociale sur le thème : "le financement de l'économie sociale", septembre.
- Fermon B. (1993).** "Enveloppe globale, régulation de la médecine ambulatoire et relation d'agence", *Cahier du CERESA*, n° 3, université Paris-Dauphine.
- Gasmi, F., Ivaldi, Laffont J.J. (1996).** "Rent Extraction and Incentives for Efficiency in Recent Regulatory Proposals", *Journal of Regulatory Economics*, vol. 6, pp. 151-176.
- Gaynor M. (1994).** "Issues in the Industrial Organization of the Market for Physician Services", *NBER Working Paper Series*, n° 4695, avril.
- Glazer J., McGuire Th. (1994).** "Payer Competition and Cost Shifting in Health Care", *Journal of Economics and Management Strategy*, vol. 3, pp. 231-244.
- Harris M., Raviv A. (1978).** "Some Results on Incentive Contracts with Applications to Education and Employment, Health Insurance and Law Enforcement", *American Economic Review*, vol. 68, pp. 20-30.
- Hay J., Leahy M. (1982).** "Physician-Induced Demand : an Empirical Analysis of the Consumer Information Gap", *Journal of Health Economics*, n° 1, pp. 231-244.
- Holmstrom B., Milgrom P. (1991).** "Multitask Principal-Agent Analyses : Incentive Contracts, Asset Ownership and Job Design", *Journal of Law Economics and Organization*, vol. 7, pp. 24-51.
- Jones A., Zanola R. (1995).** "Agency and Health Care : a Suitable Case for Public Policy", Paper Presented at the Catania Conference in Public Policy, septembre.
- Laffont J.J. (1993).** "Nouvelles formes de réglementation", *Actualité économique*, vol. 69, n° 2, pp. 3-15.
- Laffont J.J., Tirole J. (1986).** "Using Cost Information to Regulate Firms", *Journal of Political Economy*, vol. 94, n° 3, pp. 614-641.
- Laffont J.J., Martimort D. (1995).** "Collusion and Delegation", *document de travail de l'IDEI*, n° 54, décembre.
- McDonald G.M. (1984).** "New Directions in the Economic Theory of Agency", *Canadian Journal of Economics*, vol. 17, pp. 415-440.
- Mooney G., Ryan M. (1993).** "Agency in Health Care: Getting Beyond First Principles", *Journal of Health Economics*, vol. 12, pp. 125-238.
- Mougeot M. (1986).** "Le système de santé français", *Economica*.
- Mougeot M. (1994).** "Systèmes de santé et concurrence", *Economica*.
- Mougeot M., Naegelen F. (1997).** "La réglementation hospitalière : tarification par pathologie ou achat de soins ?" *Économie et Prévision*, numéro spécial santé, n° 129-130.
- Murray J.A. (1974).** "The Behaviour of the Physician, Key to the Health Care Sector", *Journal of Behavioral Economics*, vol. III, n° 1, pp. 245-272.
- Pauly M.V. (1980).** "Doctors and their Workshops: Economic Models of Physician Behavior", *NBER Monograph*, University of Chicago Press, n° 132.
- Pauly M.V. (1974).** "Overinsurance and Public Provision of Insurance: the Roles of Moral Hazard and Adverse Selection", *Quarterly Journal of Economics*, vol. 88, pp. 44-62
- Pauly M., Satterthwaite M.A. (1981).** "The Pricing of Primary Care Physician Services : a Test of the Role of Consumer Information", *Bell Journal of Economics*, vol. 12, pp. 488-506.
- Pellisé L. (1994).** "Reimbursing Insurance Carriers: The Case of MUFACE in the Spanish Health Care System", *Health Economics*, vol. 3, pp. 243-253.
- Propper C. (1995).** "Agency and Incentives in the NHS Internal Market", *Social Science and Medicine*, vol. 40, pp. 1683-1690.
- Richardson J. (1981).** "The Inducement Hypothesis : that Doctors Generate Demand for their Own Services", in Health, Economics and Health Economics, J. van der Gaag, M. Perlman, *North Holland*, pp. 189-215.
- Rizzo J., Sindelar J. (1996).** "Optimal Regulation of Multiply-Regulated Industries: the Case of Physician Services", *Rand Journal of Economics*, pp. 966-978.
- Robinson J. (1971).** "Hérésies économiques", Calmann-Lévy, Édition française, 1972.
- Rochaix L. (1986).** "Asymétries informationnelles et Comportement Médical", thèse de doctorat en sciences économiques, université de Rennes 1, décembre.
- Rochaix L. (1989).** "Information Asymmetry and Search in the Market for Physicians Services", *Journal of Health Economics*, vol. 8, pp. 53-84.
- Rochaix L., Jacobzone S. (1997).** "L'hypothèse de demande induite : un bilan économique", *Économie et Prévision*, numéro spécial santé, n° 129-130.
- Rubinstein A., Yaari M. (1983).** "Repeated Insurance Contracts and Moral Hazard", *Journal of Economic Theory*, vol. 30, pp.74-97.
- Salanie B. (1994).** "Incitation et théorie des contrats, dans les nouvelles théories économique", *Les cahiers français*, n° 272, juillet-septembre 1995.
- Salanie B. (1995).** "Théorie des contrats", *Economica*, collection *Économie et Statistiques Avancées*.
- Sappington D.E.M. (1991).** "Incentives in Principal-Agent Relationships", *Journal of Economic Perspectives*, vol. 5, n° 2, Spring, pp. 45-66.
- Scott A. (1996).** "Agency, Incentives and Doctor Behaviour : What Do We Want Doctors to Do and How Can We Get Them To Do It ?", Paper Presented to the Inaugural Conference of the iHEA, Vancouver, mai.
- Simon H. A. (1957).** *Models of Man*, New York, John Wiley.
- Tirole J. (1986).** "Hierarchies and Bureaucracies : on the Role of Collusion in Organizations", *Journal of Law Economics and Organization*, vol. 2, n° 2, Fall, pp. 181-214.
- Titmuss R.M. (1970).** "The Gift Relationship: from Human Blood to Social Policy", London: Allen and Unwin.
- Vick S., Scott A. (1995).** "What Makes a Perfect Agent ? A Pilot Study of Patients Preferences in the Doctor-Patient Relationship", *mimeo*, HERU, University of Aberdeen .
- Williamson O.E. (1975).** "Markets and Hierarchies", New York: Free Press.
- Wolfson A. (1974).** "The Supply of Physicians' Services", in Fraser (ed), *Health Economics Symposium*, Proceedings of the First Canadian Conference.