

# Réflexions sur l'organisation du système de santé

Rapport  
Gilles Saint-Paul

Commentaires  
Brigitte Dormont  
Franck von Lennepe

Compléments  
Grégoire de Lagasnerie, Thomas Weitzenblum  
et Jérôme Wittwer

*Conception et réalisation graphique en PAO  
au Conseil d'Analyse Économique  
par Christine Carl*

© Direction de l'information légale et administrative. Paris, 2012 - ISBN : 978-2-11-009300-4  
« En application de la loi du 11 mars 1957 (article 41) et du Code de la propriété intellectuelle du 1er juillet 1992, toute reproduction partielle ou totale à usage collectif de la présente publication est strictement interdite sans l'autorisation expresse de l'éditeur.

Il est rappelé à cet égard que l'usage abusif de la photocopie met en danger l'équilibre économique des circuits du livre. »

# Sommaire

<b>Introduction</b> .....	5
<i>Christian de Boissieu</i>	

## **RAPPORT**

<b>Réflexions sur l'organisation du système de santé</b> .....	7
<i>Gilles Saint-Paul</i>	

<i>Avant-propos</i> .....	7
---------------------------	---

<i>1. Introduction</i> .....	9
1.1. Un bien pas comme les autres ? .....	9
1.2. L'anti-économique ? .....	11
<i>2. Intervention publique et imperfections de marché</i> .....	13
2.1. Imperfections de marché et assurance publique .....	15
2.2. Le système actuel pallie-t-il les problèmes d'information du marché des soins ? .....	16
2.3. La fourniture publique d'assurance-maladie et de soins est-elle une forme de redistribution en nature ? .....	19
2.4. Conclusion .....	21
<i>3. Problèmes liés à la tarification à l'activité</i> .....	23
3.1. Les principes de base d'une tarification à l'activité .....	23
3.2. Le problème du prix .....	27
3.3. Le problème de l'homogénéité des GHM .....	41
3.4. Le problème de l'absence de mécanisme de réallocation .....	44
<i>4. Perspectives de politique économique</i> .....	48
4.1. La difficile conciliation d'objectifs contradictoires .....	19
4.2. L'arbitrage entre incitation et égalité dans l'accès aux soins .....	50
4.3. De meilleures incitations à l'endroit des professionnels de santé .....	53

## **COMMENTAIRES**

<i>Brigitte Dormont</i> .....	59
<i>Franck von Lennep</i> .....	67

## **COMPLÉMENTS**

<b>A. Régulation de la demande de soins : une analyse de la littérature</b> .....	73
<i>Grégoire de Lagasnerie</i>	
<b>B. Assurance-maladie : évaluation de modifications du système de remboursement des soins</b> .....	99
<i>Grégoire de Lagasnerie</i>	
Annexe. Présentation de la maquette ISIS .....	129
<i>Bérengère Mesqui</i>	
<b>C. Évaluation économique de la vie humaine : fondements théoriques, mesure empirique et implications en matière de choix social</b> .....	137
<i>Thomas Weitzenblum et Jérôme Wittwer</i>	
 <b>RÉSUMÉ</b> .....	 163
<b>SUMMARY</b> .....	171

# Introduction

Le rapport qui suit a été préparé dans le cadre de l'ancien CAE, celui qui a fonctionné jusqu'en octobre 2012. C'est à ce titre que je l'introduis.

Sans chercher à aborder les nombreuses questions au cœur de l'économie de la santé, ce rapport privilégie deux aspects :

- Quelles sont les justifications économiques de l'intervention de l'État dans le secteur de la santé ?
- Que penser des systèmes en vigueur de tarification des actes médicaux et comment les améliorer ?

La première interrogation renvoie aux arguments habituels en faveur d'une intervention publique : incomplétude des marchés, risque et même réalité de sélection défavorable, biais poussant les assurés vers des excès de dépenses ou de risques. Mais d'après l'auteur, il convient de remettre à plat un certain nombre de paramètres et d'idées reçues, et pas seulement du fait du resserrement de la contrainte budgétaire avec la crise. Sous certains angles, des imperfections des marchés devraient conduire à une amélioration des réglementations et des dispositifs incitatifs, pas toujours à la fourniture publique des activités considérées.

Le débat sur la « bonne » tarification des activités de santé est à la fois plus précis et plus documenté. La tarification à l'activité (T2A) pratiquée en France est un progrès en comparaison du système antérieur. Elle soulève cependant un certain nombre de problèmes, qu'il s'agisse de la classification des activités, de la fixation des tarifs, de leurs impacts sur le comportement des parties concernées : médecins, patients, assurance-maladie, etc. Le rapport souligne la difficulté de mettre en œuvre un pseudo-marché dans un contexte où les mécanismes concurrentiels de fixation des prix ont été dès le départ neutralisés. Demeure posé le défi central de toute tarification des services de santé : comment articuler et pondérer les considérations d'efficacité dans l'allocation des ressources et de justice sociale. Comment concilier redistribution et incitations ? La discussion sur les franchises, le ticket modérateur est alors sollicitée.

Ce rapport adopte une approche transversale, appliquant au domaine de la santé des questionnements issus de l'économie publique, de la théorie des incitations, de la théorie de l'équilibre partiel et de l'analyse de la

concurrence Par son côté stimulant et un brin provocateur, il va susciter des controverses et un débat animé chez les spécialistes de l'économie de la santé mais aussi bien au-delà.

Le rapport engage avant tout son auteur comme tous les rapports du CAE. Il a bénéficié du concours de Thomas Weitzenblum, conseiller scientifique au CAE.

*Christian de Boissieu*  
*Ancien Président délégué du CAE*  
*Professeur à l'Université de Paris I-Panthéon-Sorbonne*

# Réflexions sur l'organisation du système de santé

**Gilles Saint-Paul**

*Professeur à l'Université de Toulouse I*

## **Avant-propos**

Les rapports du CAE sont bien souvent coécrits par un membre du Conseil et un spécialiste de la question. Pour des raisons qu'il est inutile d'évoquer ici, il n'a pas été possible de trouver un spécialiste français de l'économie de la santé qui accepte de collaborer au rapport proprement dit. Certains se sont cependant prêtés aux auditions et ont accepté de rédiger des compléments ; qu'ils en soient remerciés.

Du coup, ce rapport, qui ne prétend évidemment pas traiter de tous les aspects économiques de la santé en France, est inévitablement entaché d'un certain amateurisme. L'auteur espère cependant que ce défaut ne soit pas exempt de qualités, et que la fraîcheur de sa perspective enrichisse le débat.

## 1. La genèse de la Sécurité sociale en France

La naissance de la Sécurité sociale peut être précisément située au 4 octobre 1945. Toutefois, des mesures tout à fait significatives, en matière de protection sociale, avaient déjà vu le jour précédemment. En 1898, et à la suite des initiatives allemandes suscitées par le chancelier Bismarck, une loi sur les indemnités d'accidents du travail fut promulguée en France. Puis, à l'issue de la Première Guerre mondiale, le retour de l'Alsace-Moselle en France donna l'occasion de mieux découvrir l'organisation germanique en matière de protection sociale. Cela fut à l'origine de travaux parlementaires, certes longs, laborieux et retravaillés de manière récurrente, qui débouchèrent sur la loi sur l'assurance sociale du 25 avril 1930. Cette loi mettait en place un régime de protection sociale tant pour la maladie, que pour la maternité, l'invalidité et la vieillesse, mais uniquement pour les salariés en dessous d'un plafond de rémunération.

Durant la seconde guerre mondiale, à mesure que la victoire se dessinait, le Conseil national de la Résistance entama des réflexions qui menèrent, le 28 février 1944, au Programme d'action de la Résistance, qui faisait mention pour la première fois du terme « Sécurité sociale ». Puis, Alexandre Parodi, alors ministre du Travail, mandata Pierre Laroque pour proposer la création d'un système de protection sociale refondée. À la sortie de la Seconde Guerre mondiale, le besoin de solidarité, tel qu'évalué par le gouvernement provisoire, était très fort. Il s'agissait tout à la fois de débarrasser les travailleurs de lourdes incertitudes, mais également de créer un nouvel ordre social dans lequel les travailleurs assument davantage de responsabilités.

L'ordonnance du 4 octobre 1945 consacra la création de la Sécurité sociale moderne. Elle se caractérisait notamment par les points suivants :

- elle concernait tous les salariés : sans considération de plafond de rémunération, ceux-ci étaient rattachés au système général et obligatoire d'assurance sociale. Cette unique caisse remplaçait alors les très nombreux organismes spécialisés par secteurs d'activité, et par le type de risque pris en charge ;
- les risques maladie/maternité/invalidité étaient financés par une cotisation sociale sur les salaires de 12 %, répartie à part égales en cotisations employeur et cotisations employé ;
- la mise en place du système général et obligatoire englobait des professions particulières qui n'étaient pas couvertes jusqu'alors (chauffeurs de voiture par exemple). Une assurance longue maladie fut créée, le délai de carence pour les indemnités journalières fut abaissé, et l'assurance fut étendue aux travailleurs étrangers ;
- les médecins étaient en position favorable. En l'absence de signature de convention, ils étaient libres de choisir leurs honoraires, et le remboursement s'effectuait sur la base de « tarifs d'autorité » fixés par arrêtés interministériels, à des niveaux très faibles ;
- alors que l'intention initiale était de créer une unique caisse regroupant tous les risques, cela s'est avéré impossible à imposer dans la pratique. Dans l'esprit, l'unique caisse donnait la possibilité aux transferts de fonds entre branches ;
- un ticket modérateur (part de la dépense de santé demeurant à la charge de l'assuré) s'élevant à 20 % était institué, de manière à éviter les consommations excessives de soins.



## 1. Introduction

La santé est l'un des postes les plus importants des dépenses publiques françaises, et les décisions politiques la concernant ont un enjeu considérable. Depuis sa mise en place en 1945 (voir encadré 1), le système de santé a fait l'objet de nombreuses réformes et ajustements, et les rapports publics à son sujet ne manquent pas. Si celui-ci peut prétendre à l'originalité, c'est peut-être parce qu'il essaye d'éclairer les pistes de réforme futures par des fondements de la théorie économique. À cet égard, la première question à poser est : qu'est-ce qui justifie l'intervention publique en matière de santé ? La réponse généralement admise est que la santé n'est pas un bien comme les autres et que le marché de la santé fonctionne mal. Encore faut-il expliciter les raisons pour lesquelles ce serait le cas ; bien comprendre ces raisons est non seulement utile sur le plan de l'analyse mais permet aussi d'entrevoir des pistes de réforme pertinentes.

### 1.1. Un bien pas comme les autres ?

On considère souvent que la santé est un bien « pas comme les autres » ce qui justifie la régulation publique du marché des soins. Il est intéressant d'analyser cette assertion. On peut lui donner un contenu éthique ou économique.

Sur le plan éthique, on peut penser qu'il n'y a pas à transiger avec la vie humaine. Le problème est que ce point de vue engendre des problèmes de cohérence lorsqu'il s'agit de le mettre en application. Si on l'interprète comme signifiant qu'il n'y a pas de limites aux ressources que l'on est prêt à dépenser pour sauver une vie humaine, on se heurte au fait que ce n'est pas sur ce principe que reposent les choix publics dans d'autres domaines que la santé. Par exemple lorsque l'analyse coût-bénéfice s'applique aux infrastructures routières, on prend en compte un « coût de la vie humaine » fini. En principe, le coût de la vie humaine devrait être le même quels que soient les choix publics considérés. Dans le cas contraire ceux-ci seront inefficaces. On pourrait évidemment envisager que le coût de la vie humaine soit infini dans tous les domaines d'intervention publique, mais cela signifierait que tout investissement devrait être réalisé, quel que soit son coût, dès lors qu'il épargne au moins une vie, ce qui conduirait finalement à consacrer l'ensemble des ressources nationales à la préservation de la vie humaine (qui, en l'absence d'autres formes de consommation, notamment l'alimentation, deviendrait impossible). La perspective selon laquelle la vie humaine n'aurait « pas de prix » (au sens propre) semble donc déboucher sur une impasse.

On peut objecter qu'en ce qui concerne la santé, les malades sont « connus » des praticiens et de leurs proches, tandis que les victimes des accidents de la route sont *a priori* anonymes. En d'autres termes les choix d'infrastructures se font sous un voile d'ignorance sur l'identité des victimes, tandis que les soins sont administrés en connaissance de cause.

Selon cette perspective, une vie humaine anonyme traitée statistiquement n'aurait pas la même valeur que celle d'un individu particulier. En fait si l'on se place du point de vue de l'individu lui-même, on ne voit pas où est la différence. Toutes choses égales par ailleurs<sup>(1)</sup>, il n'y a pas de raison de penser que le coût psychologique du fait de mourir d'une maladie soit plus élevé que celui de mourir d'un accident (et si l'on valorise la vie humaine par années restant à vivre, alors le coût d'un accident est généralement supérieur à celui d'une maladie du fait de l'âge de la victime). Ce qui est en jeu lors d'une maladie, ce sont les sentiments de compassion et les dilemmes moraux que subissent les praticiens et les proches du patient. D'un point de vue économique, il s'agit là de « coûts » externes associés à la maladie et en principe on pourrait les prendre en compte dans l'analyse coût-bénéfice des dépenses de santé<sup>(2)</sup>. On déboucherait alors sur une situation où la prise en compte de ces externalités morales et psychologiques impliquerait un coût social de la vie humaine différent selon le type de décès. Ceci se traduirait sans doute par une valeur plus élevée de la vie humaine s'appliquant au secteur de la santé, mais pas par des différences qualitatives profondes dans le traitement de la santé relativement aux autres biens.

Sur le plan économique, la santé serait un bien différent des autres de sorte que les transactions et contrats auxquels elle donnerait lieu seraient eux-mêmes spécifiques.

D'une part il s'agit d'un marché d'assurance qui comme tel est sujet aux imperfections inhérentes aux asymétries d'information présentes sur ces marchés. Ces problèmes sont discutés dans la section suivante. Ils sont amplement discutés dans la littérature économique et communs à tous les marchés d'assurance. Un aspect plus spécifique est qu'il peut exister des conditions préexistantes observables (comme des problèmes congénitaux) associées à un risque élevé. Il est alors impossible pour l'individu de s'assurer contre ces conditions et on peut vouloir l'assurer avant même sa naissance, ce qui justifie le rôle d'une assurance sociale<sup>(3)</sup>. D'autre part, les consommateurs de services de santé ne connaissent pas la valeur d'usage de ces services. Ils sont contraints de s'en remettre à leur médecin pour juger

---

(1) Ce qui n'est évidemment pas le cas, mais, en bonne logique, nous ne mettons aucun *a priori* sur la question de comparer les coûts pour la personne concernée de différents types de décès.

(2) On peut penser qu'il est sordide de mettre un « prix » sur les sentiments moraux et sur la souffrance psychologique, tout comme d'ailleurs sur la vie humaine. Cependant (dans la mesure où certaines décisions sont aux mains de l'État, c'est-à-dire d'une organisation bureaucratique, plutôt que des individus), on voit mal comment faire l'économie d'une telle perspective dans la formulation des choix publics, dans la mesure où ceux-ci doivent obéir à une logique. Par ailleurs même en l'absence d'État les individus sont amenés à aligner (ou pas) leurs actions sur leurs sentiments moraux et ces actions sont coûteuses. Toute décision concrète (telle que don d'organe, don du sang, etc.) est donc implicitement associée à un équivalent monétaire des considérations morales qui ont été mobilisées pour cette décision.

(3) On pourrait cependant envisager que les parents eux-mêmes assurent leurs enfants avant leur naissance, cependant si la condition préexistante est héréditaire cette possibilité disparaît.

de l'efficacité de tel ou tel traitement, ce qui crée un problème d'aléa moral dans la mesure où le revenu du médecin est plus ou moins lié à la durée et au coût du traitement. Ce problème est en fait inhérent à de nombreux biens sur l'utilité desquels le vendeur est mieux informé que l'acheteur, mais il est particulièrement saillant dans le domaine de la santé à cause d'une part de l'importance qu'on lui accorde, et d'autre part du fait que le coût du traitement n'est pas acquitté par le patient.

## 1.2. L'anti-économique ?

Ces caractéristiques des biens de santé expliquent pourquoi de nombreuses personnes auditionnées dans le cadre de ce rapport sont convaincues que le marché laissé à lui-même fonctionnerait moins bien que le système actuel, et ceci indépendamment des considérations d'équité. Ainsi, tandis qu'une hausse de l'offre engendre généralement une baisse des prix, beaucoup pensent qu'une augmentation du nombre de médecins se traduirait par une hausse mécanique du nombre de prescriptions qui serait fatale au financement des dépenses de santé. Ce genre de considérations est d'ailleurs régulièrement invoqué comme une justification au *numerus clausus* des professions de santé. Cependant il est difficile de distinguer les dysfonctionnements inhérents aux caractéristiques du marché de la santé, de ceux induits par le système d'assurance publique et de prix administrés qui prévaut en France. Ce n'est pas la liberté d'installation et de prescription qui est en soi responsable des dérives que nous venons d'évoquer, mais le fait que celle-ci soit couplée avec des prix administrés et une socialisation des dépenses.

Il en va de même pour la pénurie de médecins dans certaines régions<sup>(4)</sup>. On peut penser que cette pénurie est due aux spécificités du « bien santé » et qu'une solution administrée contraignant certains médecins à s'installer dans certaines zones doit être imposée. Cependant, on peut également considérer que ce sont les tarifs administrés qui découragent l'installation des médecins. En l'absence d'une régulation de ces tarifs, ceux-ci s'élèveraient dans les zones considérées, ce qui contribuerait à attirer les médecins. Cette situation peut être considérée comme « injuste » mais elle doit être comparée aux coûts correspondants à la « pénurie » actuelle, ainsi qu'à ceux associés à des mesures contraignantes comme celles évoquées plus haut.

---

(4) Notons que cette idée n'est pas nécessairement validée empiriquement. Une étude récente de Barlet et Collin (2010) compare la répartition territoriale des professions de santé avec celle d'autres services et ne trouve pas de différence significative. En général, plus un service est spécialisé et couvre une fraction faible de la population à un instant donné, plus il aura tendance à être localisé dans les grands centres urbains. Il n'est donc pas étonnant que certaines spécialités médicales soient peu présentes dans les zones rurales, pour la même raison que bien d'autres services n'y sont pas présents. Les personnes qui pour des raisons personnelles sont particulièrement demandeuses de ces spécialités peuvent en principe choisir elles-mêmes leur lieu de résidence de façon adéquate. Bien entendu, un tel choix est coûteux et ceci d'autant plus qu'il existe des rigidités sur le marché du logement. Par exemple, un foyer titulaire d'un logement social risque de le perdre s'il déménage. Mais ces questions sont inhérentes au marché du logement et pas spécifiques au système de soins.

De fait, les pouvoirs publics semblent bien conscients de cette carence, puisque depuis 2007, un avenant à la convention entre l'assurance-maladie et les médecins libéraux prévoit une hausse de 20 % des honoraires pour ceux qui accepteraient de s'installer dans certaines zones dites « déficitaires ». Rien ne dit, évidemment, que cette majoration constitue le « bon » prix, ni que les zones « déficitaires » soient correctement identifiées. On peut se demander s'il ne serait pas judicieux de laisser le marché déterminer cette majoration. Cela remettrait cependant en question le fait que les prix soient réglementés et pourrait conduire à un système à deux vitesses où les zones « déficitaires » pratiqueraient la liberté des prix tandis que les autres auraient un tarif « sécu ». Par ailleurs d'aucuns objecteraient qu'il n'est pas acceptable que les patients en zone rurale paient plus cher leur consultation<sup>(5)</sup>.

Ces exemples illustrent les tensions auxquelles est soumis notre système de santé. La nécessité d'allouer correctement les ressources et la reconnaissance graduelle du fait que, malgré les imperfections du marché, un système de prix peut jouer un rôle positif (par exemple en éliminant les déséquilibres et en favorisant l'adoption de thérapies les moins coûteuses) se heurtent au souci d'équité, voire d'égalité, extrêmement prégnant dans le débat public. En effet, un système de prix ne fonctionne correctement que si *quelqu'un* acquitte le prix car sinon il n'y aurait pas de raison intrinsèque de préférer un prix plus faible à un prix plus élevé. Mais comme l'incidence sur le bien-être d'un prix donné diffère selon les individus, la mise en place d'un système de prix a des conséquences distributives que l'on doit analyser, voire traiter, selon les préférences sociales en matière de redistribution.

Cette tension est en quelque sorte le thème de ce rapport, qui, rédigé par un « outsider » par rapport à l'économie de la santé, ne prétend pas aborder tous les thèmes. Il comprend deux sections principales. La première (section 2) s'interroge sur le bien fondé de la fourniture publique de soins et d'assurance étant donné les imperfections du marché des soins. Il suggère entre autres que le système actuel ne semble pas être la meilleure réponse aux imperfections de marché. La seconde (section 3) analyse les problèmes inhérents au fonctionnement de la tarification à l'activité (T2A) mise en place dans les années 2000 à l'hôpital. Cette analyse illustre les difficultés à mettre en œuvre un système de prix quand ce dernier n'est pas déterminé par l'équilibre entre l'offre et la demande et n'est pas non plus acquitté par les principaux intéressés.

(5) Il est évidemment tentant d'évacuer le problème des conséquences distributives d'une tarification plus élevée dans les zones de pénurie en transférant le coût supplémentaire vers la collectivité. Ainsi, Barlet et Collin (2010) proposent de « donner aux agences régionales de santé la possibilité de mieux rémunérer les professionnels de santé intervenant dans les territoires où l'offre médicale est insuffisante ». On voit mal cependant pourquoi on mettrait en place un mécanisme compensatoire pour des différences de niveau des prix spécifique aux biens de santé, plutôt qu'à l'ensemble des biens et services. Par exemple les tarifs d'assurance automobile diffèrent selon les départements, et il en va évidemment de même pour l'immobilier. Voir par ailleurs la section 2 pour une discussion des arguments selon lesquels la santé devrait faire l'objet d'une redistribution spécifique en nature.

Enfin, la section 4 évoque quelques pistes de politique économique. Plus que dans d'autres domaines, il est difficile de séparer les considérations éthiques des aspects économiques. Cela signifie par exemple qu'il est difficile d'envisager une réduction substantielle des dépenses de santé sans remettre en question certains postulats éthiques sur l'égalité dans la consommation de soin ; ou encore, que l'on ne peut à la fois maintenir des prix administrés et résorber les pénuries sans remettre en question le principe de la liberté d'installation. Plus généralement, cela illustre le fait déjà évoqué que l'on ne peut concilier principes et gestion publique qu'en mettant un « prix » sur ces principes, c'est-à-dire en les relativisant. Il est souhaitable que ces arbitrages soient explicites et fassent l'objet d'un débat, plutôt que déterminés par des décisions hâtives forcées par une crise budgétaire.

Le message général de la section 4 est dans la continuité de maintes propositions déjà faites par d'autres économistes. L'idée générale est une meilleure internalisation des coûts par les acteurs auxquels les décisions incombent – médecins et patients. La mise en place de dispositifs de type franchise/plafond indexés sur le revenu permettrait de concilier l'implémentation de mécanismes incitatifs pour les patients tout en préservant le caractère redistributif du système. De manière complémentaire, on pourrait envisager de responsabiliser les praticiens en mettant en place des mécanismes d'intéressement dans le cadre de la T2A, et d'affiliation à des organismes d'assurance qui se livreraient à une vraie concurrence par les prix auprès des consommateurs.

## 2. Intervention publique et imperfections de marché

Le système français d'assurance-maladie est bien connu. Il repose sur une prise en charge publique partielle complétée éventuellement par le recours à une assurance privée ou à une mutuelle. Deux exceptions existent cependant pour les bénéficiaires de la couverture maladie universelle (CMU) et les patients en affection de longue durée (ALD) qui sont exonérés du ticket modérateur. Ce système est lui-même couplé avec une fixation par l'État des tarifs, qui fait également l'objet d'exceptions (secteur 2, dépassements d'honoraires). La situation actuelle n'a pas été conçue *a priori* comme « contrat optimal » reflétant les préférences de la « société » et les problèmes d'information inhérents à la santé. Au contraire, elle est le produit d'une évolution sur laquelle ont pesé tant les contraintes budgétaires que les intérêts propres des diverses parties prenantes (médecins, patients, partenaires sociaux entre autres).

Il est cependant légitime pour l'économiste de se demander en quoi un tel système peut être optimal, et quelles sont les réformes qui permettraient de le rapprocher de l'optimalité. La question soulevée est de savoir pourquoi l'État devrait s'impliquer dans l'assurance-maladie, et pourquoi cette implication devrait être partielle et reposer sur la coexistence d'une assurance publique et d'une assurance complémentaire. Une réponse économique cohérente à cette question devrait invoquer les imperfections de marché.

### 1. Part des dépenses de santé en fonction de l'agent financeur en 2008

*En % des dépenses totales de santé*

	Canada	France	Allemagne	Italie	Japon	Suède	Roy.-Uni	États-Unis
Administrations publiques	70,5	77,7	76,6	77,5	80,8	81,5	82,4	46,0
Secteur privé	29,5	22,3	23,5	22,5	19,2	18,5	17,6	54,0
• assurance privée	12,7	13,3	9,4	1,0	2,4	0,2	1,3	33,7
• versements directs des ménages privés	14,6	7,4	13,3	19,7	15,8	16,4	11,2	12,7
• institutions sans but lucratif au service des ménages (hors assurance sociale)	—	0,1	0,4	1,9	—	0,2	4,0	3,5
• sociétés (hors assurance-maladie)	2,2	1,5	0,4	—	1,0	1,7	—	4,1

Source : Base de données de l'OCDE sur la santé (2011), mai 2012.

## 2.1. Imperfections de marché et assurance publique

Ces imperfections de marché sont bien connues.

Il s'agit tout d'abord des problèmes d'*aléa moral*, c'est-à-dire que l'assuré exerce des actions non observables par l'assureur, susceptibles de réduire ou d'accroître les risques, mais que l'assureur ne peut pas contrôler.

En présence d'aléa moral, il est optimal pour l'assureur de laisser une partie des dommages à la charge de l'assuré. En effet, c'est la seule façon pour inciter ce dernier à implémenter les actions qui permettraient de réduire le risque. S'il était assuré à 100 %, il n'aurait aucun intérêt à les mettre en œuvre. Dans le cadre de l'assurance-maladie, cela signifierait un recours excessif au système de soins, et donc une part trop importante de la richesse nationale allouée à ce système, au détriment des autres postes de consommation. Mais ceci n'est plus le cas si le contrat d'assurance stipule qu'une partie des dommages reste à la charge de l'assuré ; dans ce cas il est incité à réduire ses risques. C'est ce qui justifie les franchises ou tickets modérateurs. Un patient qui doit payer une partie de sa consultation n'aura ainsi recours au médecin que lorsqu'il a des raisons de penser que les symptômes sont suffisamment graves (voir le complément A pour une mesure de l'élasticité-prix de la demande de soins). Notons que de telles franchises sont dans l'intérêt collectif des patients eux-mêmes, car en les responsabilisant elles réduisent les dépenses et donc les primes d'assurance. Cependant, si le patient trouve une autre compagnie d'assurance prête à prendre en charge son ticket modérateur, il aura intérêt à cumuler ce contrat complémentaire avec son contrat initial. Le cumul des deux contrats réduit évidemment ses incitations à réduire ses risques. Bien entendu, l'assurance complémentaire prend en compte ce fait et le fait payer à l'assuré à travers des primes plus élevées ; mais cette hausse des primes ne reflète que partiellement les pertes dues à la réduction de l'effort de l'assuré, car une partie en est payée par l'assureur initial.

C'est pour éviter ce problème et préserver les propriétés incitatives des contrats que ceux-ci comportent en général des clauses d'exclusivité.

La deuxième imperfection est l'*incomplétude des marchés*, qui ici est associée au fait que les personnes dont la morbidité est plus élevée en raison de certaines de leurs caractéristiques observables, notamment leurs conditions héréditaires, ne peuvent pas s'assurer contre le risque d'hériter de ces conditions, parce que cela impliquerait qu'elles puissent signer un contrat avant leur naissance.

La troisième est la *sélection adverse* : le niveau de risque varie de façon non observable entre les individus. Dans de telles circonstances, les individus de risque faible peuvent être incités à ne pas s'assurer, et dans ce cas les primes augmentent pour le reste des assurés, ce qui incite les meilleurs risques parmi ceux-ci à quitter à leur tour le marché de l'assurance, et ainsi de suite. Cette spirale peut en principe conduire à la disparition pure et simple du marché d'assurance. Pour pallier cette difficulté, on préconise en général de rendre l'assurance obligatoire, comme c'est le cas pour l'automobile.

bile. On peut cependant se demander si cette justification est empiriquement pertinente. En effet, dans le cas de la santé, l'ampleur des dommages encourus devrait encourager les gens à s'assurer, même si leur risque est faible. Bien entendu, dans le cas où les agents à risque faible s'assureraient librement, ils auraient alors intérêt à révéler de manière crédible leur faible risque à leur assureur, de façon à réduire leur prime<sup>(6)</sup>. Tout comme une sortie du marché des meilleurs risques, cet « écrémage » conduit également à une hausse des primes pour les autres assurés. On peut également le pallier en régulant le secteur de l'assurance de façon à l'interdire<sup>(7)</sup>. La réglementation peut aussi traiter la question de l'incomplétude des marchés, par exemple en n'autorisant pas de discriminer selon les conditions préexistantes.

## 2.2. Le système actuel pallie-t-il les problèmes d'information du marché des soins ?

On voit donc qu'il existe des raisons économiques pertinentes justifiant une intervention publique dans le domaine de l'assurance-maladie. La question qui subsiste est celle de la forme de cette intervention. Les arguments qui précèdent ne plaident pas plus en faveur d'une *fourniture* publique de l'assurance-maladie, que pour une gestion par un marché concurrentiel réglementé. Les clauses d'exclusivité qui permettent de rendre opérationnels les contrats assortis de franchises ou de ticket modérateur peuvent fort bien être imposées par des assureurs privés. Une réglementation visant à éviter l'écrémage ou la discrimination selon les conditions préexistantes suffit à traiter les problèmes d'incomplétude de marché ou d'asymétrie d'information, comme cela est pratiqué notamment en Suisse et aux Pays-Bas<sup>(8)</sup>

(6) C'est exactement ce que fait le système du bonus-malus dans le cas de l'automobile ; on peut se demander pourquoi ceci est, dans le contexte français, éthiquement acceptable et non le même phénomène dans le domaine de la santé. On pourrait penser qu'il est à la portée de chacun d'être un bon conducteur, tandis qu'il n'est pas à la portée de chacun d'être en bonne santé, ce qui justifierait cette différence d'un point de vue moral. Cependant, on sait que les actions préventives (nutrition, exercice) ont un impact important sur l'état de santé. Inversement, on peut imaginer un conducteur ayant des problèmes de concentration, pour des raisons innées. Il serait alors tout aussi « injuste » de lui imposer un malus que d'imposer des primes plus élevées à une personne en mauvaise santé.

(7) Précisons cependant que l'écrémage ne conduit pas nécessairement à des situations inefficaces sur le plan économique. Les travaux de Prescott (notamment Boyd et Prescott, 1986) montrent qu'il existe des équilibres concurrentiels en présence d'information incomplète où différents risques payent des prix différents. Ces équilibres sont Pareto-optimaux malgré les incitations des hauts risques à se faire passer pour des bas risques. Ce sont avant tout leurs conséquences distributives qui les distinguent des équilibres sans écrémage où tous les risques payent la même prime. Le bien-être des bas risques est inférieur dans ces derniers équilibres, tandis que les hauts risques y gagnent.

(8) Au sein de l'OCDE, cinq pays permettent aux assurés de choisir leur caisse d'affiliation : l'Allemagne, les Pays-Bas, la République tchèque, la Slovaquie et la Suisse. Les caisses en concurrence dans ces cinq pays sont tenues d'accepter toutes les demandes d'affiliation (première inscription comme renouvellement), quelles que soient les caractéristiques des assurés. En Allemagne, République tchèque et Slovaquie, les cotisations sont proportionnelles au revenu, ce qui garantit qu'elles ne dépendront pas de l'état de santé des assurés. Aux Pays-Bas et en Suisse, les cotisations ne sont pas uniquement liées au revenu, mais des règles ont été édictées pour plafonner les augmentations de cotisation à l'issue du renouvellement du contrat d'assurance.



(voir Paris et *al.*, 2010). On peut cependant avancer qu'un assureur public, dont le motif n'est pas le profit, aurait moins d'incitations à contourner la réglementation qu'un assureur privé<sup>(9)</sup>. Quoi qu'il en soit, on voit mal en quoi le système actuel peut être justifié par les imperfections de marché que nous venons de discuter. Le ticket modérateur implique évidemment que l'assurance publique n'est que partielle. L'existence d'assurances complémentaires annihile le bénéfice du ticket modérateur en termes de gestion de l'aléa moral (notons cependant que cette critique ne s'applique pas à la contribution d'un euro, récemment introduite, qui n'est pas remboursable par une mutuelle. Reste à savoir si une somme aussi modeste est susceptible de créer les bonnes incitations). Et de plus, ces complémentaires peuvent également tenter de contourner la réglementation pour se livrer à l'écémage, ce qui réduit la pertinence de l'argument précédent<sup>(10)</sup>. En fait, sur la base des arguments qui précèdent, on pourrait s'attendre à ce que les interventions publiques soient structurées de façon à alléger les problèmes d'aléa moral et de sélection adverse. Pourtant, le système des complémentaires entérine une situation de multi-assurance dont on vient précisément de voir qu'elle est inefficace. Et les inefficacités engendrées par le défaut de coordination entre la Caisse nationale d'assurance-maladie (CNAM) et les complémentaires vont au-delà de ce problème d'aléa moral. Par exemple, les différentiels de coût entre hôpital public et cliniques privées ne sont pas répercutés dans les remboursements aux patients. Dans son rapport de juillet 2011, la CNAM relève que de nombreuses thérapies sont moins coûteuses, dépassements d'honoraires compris, dans le secteur privé que dans le secteur public : « Ainsi en 2010, pour une cholécystectomie sans exploration de la voie biliaire pour affections aiguës (...), une clinique facturait en moyenne 2 549 euros, dépassements d'honoraires inclus, soit 746 euros de moins que l'hôpital public (23 %) » (p. 36).

Le rapport de la CNAM contient d'autres exemples de ce problème, et relève qu'il est également générateur d'inéquités entre patients, puisque ceux qui se font soigner dans le privé acquittent un prix plus élevé, en dépit du fait qu'ils contribuent à une réduction des dépenses totales en choisissant l'option la moins coûteuse.

Il est donc dans l'intérêt général que ces actes aient lieu dans le secteur privé, mais ceci n'est pas dans l'intérêt des assurances complémentaires, puisqu'elles supportent un coût plus élevé si le patient est traité en clinique qu'à l'hôpital. Le rapport de la CNAM préconise une meilleure prise en charge par la CNAM des actes effectués par le privé, dont la contrepartie serait une réduction des dépassements d'honoraires. En principe, dans les

---

(9) Cependant, un assureur public structurellement déficitaire, comme c'est le cas de la CNAM, sera amené à faire des économies et donc à se comporter à la marge comme un assureur privé qui minimise les pertes...

(10) Une technique consisterait par exemple à réduire les primes pour les assurés qui utilisent une plate-forme internet pour communiquer avec leur mutuelle, ce qui attirerait mécaniquement les plus jeunes. On peut noter cependant qu'il est moins coûteux pour l'assureur que ses clients passent par internet plutôt que par un guichet, ce qui justifie une telle réduction indépendamment de toute considération d'âge.

cas mentionnés ci-dessus, on pourrait même aller plus loin et réaliser des économies en sous-traitant l'ensemble de ces actes au privé tout en prenant en charge les coûts correspondants, dépassements d'honoraires compris, jusqu'à concurrence d'une fraction (par exemple 90 %) du coût du même traitement effectué à l'hôpital. Les choses ne sont cependant pas si simples car il existe un effet d'aubaine pour les patients qui se font soigner dans le secteur privé dans la situation actuelle ; la prise en charge des dépassements d'honoraires pour ces patients ne représente pas des économies mais un transfert à leur profit ou au profit de leur complémentaire. On voit là qu'il n'y a pas de solution simple à ce type de problème tant que l'on reste dans un cadre de multi-assurance. Par contre, le problème disparaîtrait entièrement si, pour un acte donné, n'existait qu'un seul assureur qui remboursait le même montant, voire la même proportion des coûts totaux, quelle que soit l'option choisie par le patient.

Le problème qui est soulevé ici illustre celui, plus général, de la multiplicité des décideurs : patient, médecin traitant, CNAM, mutuelles. Cette multiplicité est inévitable compte tenu du fait que le bénéficiaire des soins (le patient) n'est pas le mieux informé sur leur valeur (c'est le médecin) ni le payeur (c'est l'assurance). Cependant on peut la réduire en faisant en sorte que l'assureur soit unique pour chaque catégorie de soins, ce qui élimine au moins les problèmes de coordination entre assureurs et les inefficacités dues à la multi-assurance. Nous offrons quelques pistes de réforme à la section 4<sup>(11)</sup>. Les problèmes de coordination au sein du triangle patient-médecin-assureur subsisteront néanmoins, mais ceux-ci sont « ontologiques » et ne peuvent être éliminés complètement. Bien entendu, un système purement privé de type assurantiel, impliquerait que les primes seraient indépendantes du revenu. S'il était avéré que l'état de bonne santé fût positivement corrélé avec le revenu, on pourrait même imaginer des primes décroissantes avec ce dernier, mais le lien empirique entre état de santé et revenu n'est pas clairement établi<sup>(12)</sup>. Dans notre système, les contributions sont assises sur le salaire, et donc plus élevées pour les ménages plus aisés. Cette dimension redistributive justifie-t-elle que la fourniture d'assurance-maladie soit publique ? C'est la question vers laquelle nous nous tournons.

---

(11) Pour Grignon (2010), le régime d'adhésion volontaire aux mutuelles permet d'éviter une uniformisation totale des prestations de santé, qui serait inefficace. On voit mal cependant comment une telle diversité ne pourrait pas être obtenue à l'aide d'un assureur privé unique pour chaque patient, qui offrirait un menu de contrats sous une contrainte de cahier des charges imposé par la réglementation.

(12) L'idée naïve selon laquelle les « pauvres sont en moins bonne santé que les riches » n'est que partiellement validée par les données. Lachaud et Rochaix (1995), cités dans le rapport Mougeot (1999), regardent la distribution par quintile de revenu de cinq indicateurs de morbidité et un seul d'entre eux est significativement et négativement corrélé avec le revenu. Notons cependant qu'il s'agit d'un indicateur critique, à savoir le fait d'être atteint d'une maladie de « gravité sévère » (menaçant la vie du patient). Ce scepticisme est partagé par nombre d'études économétriques qui s'attachent à isoler un effet causal du statut socio-économique sur l'état de santé. Voir, par exemple, Adams et al. (2003) et Frijters et al. (2005).

### 2.3. La fourniture publique d'assurance-maladie et de soins est-elle une forme de redistribution en nature ?

Si les préoccupations redistributives constituaient le fondement de l'assurance-maladie publique, celle-ci ne serait pas universelle mais simplement réservée aux plus pauvres. On peut cependant se demander si, *nolens volens*, la société française n'évolue pas vers ce type de modèle. En effet, d'un côté la part des coûts laissée à la charge du patient (ou de son assurance complémentaire) ne cesse d'augmenter sous l'effet des pressions budgétaires ; de l'autre sont mis en place des mécanismes d'exonération de ce reste à charge (notamment la CMU), en partie motivés par le fait que la hausse des tickets modérateurs est plus difficile à supporter pour les ménages les plus modestes. Il serait souhaitable qu'une telle évolution (qui concerne l'ensemble du système de protection sociale et pas seulement l'assurance-maladie), fasse l'objet d'un large débat plutôt que de s'imposer insidieusement à la faveur des « trous » répétés du budget de la sécurité sociale.

Supposons donc que ce soit pour des raisons de redistribution que l'on mette en place un système public d'assurance-maladie<sup>(13)</sup>. En quoi ceci constitue-t-il un bon outil redistributif ?

La réponse de la théorie économique est que ce n'est pas le cas. En effet, le bien-être des bénéficiaires de la redistribution augmenterait, à coût égal, si au lieu de les assurer d'office, on leur versait l'équivalent monétaire de leur assurance. Ils pourraient alors l'utiliser pour s'assurer eux-mêmes, et s'ils ne le font pas (ou s'ils optent pour un contrat d'assurance moins coûteux), c'est nécessairement parce que l'usage alternatif de ces fonds accroît leur bien-être. Les considérations purement redistributives plaident donc pour une redistribution monétaire et écartent *a priori* toute forme de redistribution en nature<sup>(14)</sup>.

On peut cependant raffiner l'analyse et avancer des arguments en faveur de la redistribution en nature. Ainsi Henri et Rochet (1999) montrent qu'un système de redistribution en nature peut avoir des meilleures propriétés incitatives qu'un système purement monétaire si la demande pour le bien redistribué décroît avec le revenu. L'idée est simple : lorsque la redistribution des revenus est sous forme monétaire, elle est généralement associée à des conditions de ressources afin de limiter son coût budgétaire. Mais ces conditions de ressources ont de forts effets désincitatifs, car les bénéficiaires de l'aide savent qu'ils la perdraient s'ils travaillaient plus pour augmenter leur revenu. Cet effet est en revanche atténué lorsqu'on redistribue en nature un bien dont la demande décroît avec le revenu. Certes, il est exact qu'une

(13) Même dans un pareil cas, on pourrait envisager de déléguer la gestion du système à une assurance privée, sélectionnée par exemple sur appels d'offres. Ceci soulève la question de l'efficacité de l'État en tant que producteur de biens et services relativement au secteur privé, que l'on peut évidemment poser à propos de l'ensemble des activités économiques et pas seulement l'assurance-maladie.

(14) On peut évidemment arguer qu'il n'est pas moralement possible à la collectivité de ne pas secourir une personne dans le besoin même si sa situation courante résulte de ses libres choix passés (comme le fait de n'avoir pas souscrit à une assurance-maladie), mais ces considérations ne sont pas tant de nature redistributives que paternalistes, ce dernier aspect étant discuté plus bas.

hausse du revenu est associée à une perte d'éligibilité, mais comme le bien redistribué est moins valorisé par le consommateur lorsque son revenu augmente, cette perte est moins douloureuse que si la redistribution avait lieu sous forme monétaire.

Cet argument est certainement pertinent en ce qui concerne certains programmes comme les « *food stamps* » américains<sup>(15)</sup>. Il est moins clair qu'il s'applique à l'assurance-maladie. Certes, comme le montrent Henriet et Rochet, la morbidité est une fonction décroissante du revenu. Mais c'est également le cas pour l'utilité marginale de la consommation, ce qui signifie que l'on valorise moins celle-ci (en termes hédoniques) lorsque l'on est plus riche. Si l'on prend en compte cette observation, il n'est pas évident que la demande d'assurance-maladie soit elle-même décroissante avec le revenu. De fait, on observe que les classes aisées n'hésitent pas à contracter des assurances complémentaires généreuses et coûteuses. Enfin, en tout état de cause, si l'argument de Henriet et Rochet justifie une redistribution en nature plutôt que purement monétaire, il n'en reste pas moins que ce système ne devrait s'adresser en principe qu'aux plus démunis et non à l'ensemble de la population comme c'est le cas en France. La logique de cet argument ne semble donc pas pertinente au regard du caractère universel de l'assurance-maladie.

Une autre justification de l'assurance médicale est de nature *paternaliste*. L'idée est que les individus sont insuffisamment informés sur leurs risques ou sujets à des biais comportementaux tels que la procrastination<sup>(16)</sup>, qui les conduisent à ne pas s'assurer de façon optimale du point de vue *de leur propre intérêt*. Ces considérations sont dangereuses au sens où, après tout, on pourrait les appliquer à l'ensemble des choix individuels et donc mettre en place un système généralisé de contrôle des personnes<sup>(17)</sup>. Supposons cependant que l'on accepte l'existence de ces biais et que l'on considère que leurs conséquences dans certains domaines sont suffisamment sérieuses pour que le gouvernement juge utile d'« orienter » les choix individuels. Beaucoup d'économistes jugent que l'on peut remédier à ces biais au moyen d'interventions plutôt légères (voir, en particulier, Thaler et Sunstein, 2003 et O'Donoghue et Rabin, 2003). Ces interventions incluent l'accès du public à une information objective, des interventions fiscales (par exemple subventions à l'assurance-maladie dans le cas qui nous préoccupe), ou encore le fait de jouer sur la façon dont les choix sont présentés (cela pourrait signifier que les individus seraient enrôlés dans une assurance-maladie « par défaut », même s'ils conservent leur pleine liberté de choix). Même si l'on pense que de telles mesures sont insuffisantes et si l'on envisage des politiques plus

(15) Notons que la redistribution en nature perd tout intérêt dès lors que le bien distribué peut être revendu sur un marché secondaire.

(16) Tendance à retarder des prises de décision.

(17) Par exemple, la recherche sur l'économie du bonheur indique que la consommation est génératrice d'effets d'habituation, et donc qu'elle n'aurait pas d'effets de long terme sur le bonheur individuel. Si l'on prend de tels résultats au sérieux pour la mise en œuvre d'une politique économique, on est conduit à pénaliser toute forme de consommation. De même, l'État paternaliste prétend s'attaquer aux addictions. Mais la liste des biens potentiellement addictifs ne cesse de croître (sodas, pornographie, Internet, travail...) et plus celle-ci s'allonge plus le champ de l'intervention de l'État s'élargit. Voir Saint-Paul (2011).

coercitives, une obligation d'assurance auprès d'un assureur privé semble convenir. L'argument paternaliste, bien qu'il puisse justifier une assurance universelle contrairement à l'argument redistributif, ne semble pas justifier une fourniture publique d'assurance-maladie. Au mieux, il plaide en faveur d'une réglementation imposant aux personnes de s'assurer *a minima*, comme c'est le cas pour l'assurance automobile.

Quoi qu'il en soit, il serait souhaitable d'ouvrir un débat qui permettrait de dégager un consensus sur la philosophie du système. Celle-ci joue en effet un rôle crucial dans le choix des mécanismes d'intervention mais elle semble s'effacer derrière les problèmes de financement et les réformes purement organisationnelles, et il n'est pas du tout clair, par exemple, que la motivation paternaliste (au sens où l'on impose aux citoyens des choix dans le but d'améliorer leur bien-être individuel) soit plus importante que l'idée floue que la « société » serait « responsable » de la santé des citoyens (au sens où elle imposerait une norme de santé à l'ensemble de ceux-ci indépendamment de toute mesure de leur bien-être individuel et de leurs biais comportementaux).

## 2.4. Conclusion

La discussion qui précède suggère qu'il n'y a pas de rationalité économique forte sous-tendant le système actuel. Les justifications reposant sur des imperfections de marché plaident en général pour une réglementation de celui-ci plutôt que pour une fourniture publique des biens considérés. Les considérations redistributives peuvent peut-être justifier une redistribution en nature mais on voit alors mal pourquoi l'ensemble de la population devrait en bénéficier. Les considérations paternalistes devraient être mieux explicitées et l'on devrait alors prouver la supériorité de l'intervention directe sur des mesures plus libérales discutées plus haut.

Le système des assurances complémentaires est générateur d'inefficacités bien connues des économistes. Il apparaît comme une aberration au regard des problèmes d'aléa moral discutés plus haut.

Si l'on admet *a priori* l'existence d'un assureur public, et si l'on admet également que sa contribution ne peut être que partielle – en raison par exemple de la rareté des fonds publics et des conséquences distorsives de l'imposition – alors il est naturel de préconiser une évolution vers un système d'assurances *supplémentaires*. Selon cette logique, la CNAM rembourserait certains actes et pas d'autres, qui seraient laissés aux mutuelles et assureurs privés. Il ne serait pas possible de cumuler les assurances sur le même acte. On note d'ailleurs que ce système est déjà pratiquement en vigueur pour les soins très peu couverts, comme les prothèses dentaires, l'optique, ou les médicaments remboursés à 15 %. La définition du panier de soins ouvrant droit au remboursement, partiel ou total, donne lieu à des différences non négligeables suivant les pays considérés (*cf.* tableau 2). En particulier, on remarque que le critère retenu pour la définition du panier en France est l'efficacité clinique, sans faire référence au rapport coût/efficacité, ni à l'accessibilité en termes financiers du bien de santé.

## 2. Modalités de définition du panier de biens de santé

	France	Allemagne	Royaume-Uni	Canada	Pays-Bas
Définition positive du panier de biens (par inclusion) centralisée	Oui	Non	Non	Non	Oui
Définition négative du panier (exclusion) centralisée	Non	Oui	Oui	Non	Non
Liste établies par les assurances privées	Non	Non	Non	Oui	Non
Critères retenus	Efficacité clinique	Efficacité clinique Rapport coût/efficacité Accessibilité financière	Efficacité clinique Rapport coût/efficacité	Efficacité clinique Rapport coût/efficacité Accessibilité financière	Efficacité clinique Rapport coût/efficacité Accessibilité financière

Source : Paris, Devaux et Wei (2010).

### 3. Problèmes liés à la tarification à l'activité

Sous l'influence de réformes populaires à l'étranger, on a vu graduellement se mettre en place des mécanismes inspirés par la logique économique. Il est donc particulièrement pertinent de s'y pencher dans un rapport du CAE. La plus importante de ces réformes est la mise en place de la tarification à l'activité (T2A) en 2005. Compte tenu du fait que l'hôpital est le plus gros poste de dépenses de santé (*cf.* tableau 3), celle-ci pourrait représenter un enjeu considérable.

#### 3.1. Les principes de base d'une tarification à l'activité

La tarification à l'activité prétend améliorer l'efficacité économique de l'allocation des soins en introduisant un pseudo-marché, c'est-à-dire un système de prix qui rémunère l'offre de soins sur la base du service réel rendu plutôt que sur la simple constatation des frais encourus. Dans l'idéal, les acteurs du système prendraient les « bonnes » décisions économiques en fonction du prix offert par le régulateur pour chaque thérapie donnée.

Pour bien comprendre les mécanismes en jeu ainsi que les difficultés inhérentes à la mise en place d'un tel système, il est utile de commencer par analyser comment fonctionnerait une T2A dans un monde où les frictions économiques seraient quasiment absentes.

Dans une telle situation, le secteur de la santé produit des biens clairement identifiés. À chaque pathologie correspond une thérapie bien définie et c'est cette thérapie qu'achète le patient. Pour chacun de ces biens existent différents fournisseurs, et ceux-ci peuvent être plus ou moins efficaces. Si le coût de la thérapie était à la charge du patient, la concurrence par les prix jouerait ; le patient choisirait le fournisseur le plus économique (notons à ce stade que nous supposons pour l'instant qu'il n'y a aucun problème lié à la qualité des soins). Les fournisseurs les moins performants seraient contraints de s'aligner sur le prix du marché. Certains seraient contraints de fermer boutique, tandis que les autres seraient moins rentables que leurs concurrents qui finiraient par les éliminer graduellement soit en croissant plus vite qu'eux, soit par des mécanismes de fusion.

Cependant, le coût de la thérapie n'est pas à la charge du patient<sup>(18)</sup>, mais au contraire à la charge de l'assureur public ou mutualiste. On sait qu'en France les mutuelles imposent peu de discipline sur les prix facturés par les professionnels. C'est donc à l'assureur public qu'a incombé cette tâche jusqu'ici. Par le passé, c'était sur la base du coût réel des soins que les hôpitaux et cliniques étaient remboursés par la CNAM. Les différences de productivité entre établissements n'étaient donc prises en compte ni par les patients ni par l'assureur : un établissement moins productif percevait

(18) L'effet du remboursement des complémentaires – abaissant considérablement, voire annulant, le reste à charge – sur la demande en biens de santé, a donné lieu à de nombreuses études (*cf.* complément A). Au final, il semblerait que le degré de prise en charge par les complémentaires joue davantage sur la probabilité d'accès aux soins, que sur le montant des dépenses de santé (HCAAM, 2005).

### 3. Répartition des dépenses de santé et modalités de financement hospitalier dans plusieurs pays de l'OCDE en 2009

	France	Allemagne	Roy.-Uni	États-Unis	Italie	Japon	Suède	Canada
Dépenses totales de santé (en % du PIB)	11,8	11,6	9,8	17,4	9,5	8,5 <sup>(*)</sup>	10	11,4
Dépenses publiques de santé (en % dépenses totales)	77,9	76,9	84,1	47,7	77,9	80,8 <sup>(*)</sup>	81,5	70,6
Dépenses en soins hospitaliers (en % dépenses totales)	29,9	27,7	—	18,9	—	31,2 <sup>(*)</sup>	26,3	16,4
Dépenses pharmaceutiques (en % dépenses totales)	16,1	14,9	11,6 <sup>(*)</sup>	12	18,3 <sup>(*)</sup>	19,4 <sup>(*)</sup>	12,5	17
Modalités de financement à l'hôpital	GHM	GHM	GHM (70 %) Budget global (30 %)	—	GHM	Paiement par procédure + <i>per diem</i> ajusté du diagnostic	GHM (55 %) budget global	Budget global prospectif (79 %) par cas (9 %) <i>per diem</i> (9 %)

Legende : GMH = groupe homogène de malades ; la dénomination anglo-saxonne des GHM est DRG (*diagnosis related group*).

Note : (\*) 2008.

Sources : Paris, Devaux et Wei (2010) et base de données de l'OCDE sur la santé (2011).



simplement des sommes plus élevées de la part de l'assurance-maladie. Les inconvénients de ce système sont nombreux et bien connus. Il n'y a aucune incitation à s'adresser aux établissements les plus efficaces. Les moins efficaces n'ont aucun intérêt à se réorganiser pour réduire leurs coûts. L'ensemble des établissements gagne à gonfler artificiellement les dépenses pour chaque prise en charge. Inversement, ils ne gagnent rien à implémenter les nouvelles technologies, ce qui tend à réduire le progrès technique dans le secteur de la santé (en effet, la taille du marché pour ce type d'innovation est diminuée, voir Acemoglu et Linn, 2004).

L'idée de la T2A est que l'assureur public paye un prix unique pour chaque thérapie, par exemple 100 pour une opération de l'appendicite. Un hôpital incapable d'effectuer une telle opération pour un coût inférieur à 100 aura tendance à abandonner cette activité. Un hôpital capable de la faire pour 80 gagnera moins d'argent par opération qu'un établissement où celle-ci ne coûte que 50. Le système tend donc à réallouer chaque type de soin vers les établissements les plus performants, ce qui améliore l'efficacité productive du système de santé. De plus, à prix donné, la mise en place d'une technologie qui permet de réduire les coûts augmente la rentabilité de l'établissement ; ce système de tarification favorise donc le progrès technique. Bien entendu, le progrès technique devrait également se traduire par une baisse des tarifs, puisque les coûts ont diminué. Notons cependant que cela ne signifie pas nécessairement que les dépenses diminuent, en effet le progrès technique peut également impliquer une hausse de la consommation et/ou de la qualité des soins<sup>(19)</sup>. Il incombe donc au régulateur de prendre en compte le changement technologique dans l'évolution des prix, par un mécanisme adéquat. Il n'en reste pas moins qu'au niveau de l'établissement individuel, les incitations au progrès technique et à la baisse de coûts demeurent, puisque les tarifs auxquels il est rémunéré dépendent de l'évolution de l'ensemble du secteur et non de ses propres décisions.

(19) Afin d'analyser avec précaution les effets du progrès technique dans le domaine de la santé, il convient de distinguer la substitution de traitement de la diffusion de traitement. La substitution de traitement permet de remplacer un traitement ancien par un traitement plus récent, plus efficace et donc moins onéreux, à qualité donnée. La diffusion, elle, engendre nécessairement un surcroît de coût global. Cutler et McClellan (2001) se penchent sur plusieurs pathologies spécifiques, dont l'infarctus du myocarde aigu. Ils montrent, entre autres, que l'apparition de l'angioplastie (pose de stents) comme alternative aux interventions chirurgicales (pontages) a engendré les deux effets (substitution notable, mais également diffusion) en sorte que l'effet sur la dépense totale est globalement neutre.

D'un point de vue de l'analyse du bien-être, il est fort possible que la hausse des dépenses induite par le progrès technique (à supposer que l'effet net soit effectivement une hausse) soit optimale. Cela signifierait simplement :

- soit que les services de santé sont suffisamment substituables aux autres services pour que l'on préfère accroître leur consommation (ou leur qualité, ce qui revient au même si celle-ci est exprimée en équivalent-consommation) plus que proportionnellement à la baisse de leur coût lorsque ceux-ci deviennent moins onéreux ;
- soit que les services de santé sont un bien supérieur au sens où leur part dans les dépenses augmente avec le revenu, ce qui implique que le supplément de pouvoir d'achat permis par le progrès technique dans le secteur de la santé – mais aussi dans tout autre secteur – est plus que proportionnellement consacré aux soins. Notons cependant qu'empiriquement, il n'est pas clair que les dépenses de santé aient une telle propriété, voir Acemoglu et al. (2009).

Il est donc évident que la T2A représente un progrès considérable par rapport à la logique du système précédent. On peut cependant se demander si, dans sa forme actuelle, elle est conçue de façon optimale. En particulier, la T2A est sujette à trois failles potentielles susceptibles de remettre en cause les gains qu'on peut en attendre :

- Les prix sont-ils fixés correctement ?
- Le service rendu correspondant à une catégorie de prix donnée est-il réellement homogène ?
- Les surplus et déficits comptables se traduisent-ils par une réallocation des ressources susceptible d'accroître l'efficacité ?

## 2. La tarification à l'activité

La tarification à l'activité (T2A) est un mode de financement des établissements de santé fondé sur le paiement à l'activité réalisée. Mise en place en 2004, elle fut programmée pour monter en puissance progressivement jusqu'en 2008.

Une classification des différentes prestations hospitalières est nécessaire pour déterminer quel sera le montant du paiement à l'établissement et repose sur la distinction par groupe homogène de malades (GHM). Il existe plus de 2 000 GHM, répartis en vingt-huit catégories majeures de diagnostic (CMD). Chaque GHM est lui-même traduit en groupe homogène de séjour (GHS) de manière à prendre en compte les spécificités de certains séjours hospitaliers. À chaque GHS est associé un tarif. Cette tarification est prospective, et fixée à une périodicité annuelle par le ministère de la Santé dans le cadre de la loi de financement de la Sécurité sociale.

Parallèlement à ce financement à l'acte, certaines activités donnent lieu :

- à un financement forfaitaire : forfait annuel « urgences », « coordination des prélèvements d'organes », « greffes » ;
- à un financement spécifique pour les médicaments et dispositifs médicaux implantables particulièrement onéreux ;
- au financement d'activités qui ne peuvent pas donner lieu à des tarifs : ce sont les missions d'intérêt général et d'aide à la contractualisation (MIGAC) : soins pour les détenus, missions d'enseignement, mise en œuvre de plans de santé publique.

La T2A représente environ les trois quarts du budget des établissements, les MIGAC s'élevant à un peu moins de 20 % de ce même budget.

### 3.2. Le problème du prix

Dans un marché « parfait », le prix d'équilibre reflète à la fois la valeur du bien pour les consommateurs et son coût de production<sup>(20)</sup>. C'est la concurrence par les prix entre les deux côtés du marché qui conduit à cette situation. D'un côté, le prix est pour les producteurs un signal de la valeur du bien pour les consommateurs ; au cours du processus de décision qui leur permet de fixer la quantité offerte, ils comparent la valeur du bien à leur coût. Ainsi, seuls les producteurs dont le coût est inférieur à la valeur du bien pour les consommateurs participeront au marché. Inversement, le prix constitue pour les consommateurs un signal de coût. En décidant s'il est souhaitable d'acheter le bien, les consommateurs comparent implicitement le coût de production du bien considéré à la quantité de bien-être, exprimée en termes monétaires, que celui-ci leur procure.

#### 3.2.1. La difficile prise en compte de la demande dans un mécanisme de T2A

Dans un système tel que la T2A, ce sont les producteurs qui sont rémunérés par la puissance publique pour les soins qu'ils délivrent. Le raisonnement qui précède suggère donc que les prix devraient être déterminés par la « demande », ce qui signifie que pour chaque type de soin, l'État devrait calculer la « valeur sociale » de cette activité et fixer le prix au niveau de cette valeur sociale<sup>(21)</sup>.

Ces principes soulèvent quelques difficultés, et ces difficultés ne sont pas étrangères au fait qu'en pratique la tarification n'est pas fixée de cette manière. Cependant, comme on le relève plus bas, des considérations de demande publique pour certains types de soins participent à la formation des tarifs, au sens où ceux-ci sont également ajustés pour refléter des priorités gouvernementales comme celles formulées dans les divers plans santé. Le problème de fond est qu'il n'existe pas de mécanisme permettant aux consommateurs de révéler au planificateur leur propension à payer pour les soins, du fait que ces consommateurs sont assurés. Cela rend difficile, voire impossible, l'établissement d'une « valeur sociale » pour chaque thérapie.

---

(20) Il s'agit en fait de la valeur et du coût *marginaux*, c'est-à-dire correspondant à une unité supplémentaire de bien produite ou consommée. Dans ce marché « parfait », ces quantités sont les mêmes quels que soient les consommateurs (ou les producteurs) qui consomment (ou produisent) le bien. Pour cette raison, le prix représente également le coût *social*, ainsi que la valeur sociale, d'une unité de bien supplémentaire. Cela signifie qu'à la marge, un planificateur social voudrait précisément envoyer aux consommateurs et aux producteurs un signal-prix égal à ce prix d'équilibre. Ce raisonnement ne tient évidemment plus lorsque le marché n'est pas parfait.

(21) Il est intéressant de noter la similarité entre ces questions et celle de la fixation des prix dans une économie planifiée, discutée par Lange (1936). Cette similarité s'applique également à la fonction objective de l'hôpital. On peut noter que Lange proposait l'imposition bureaucratique aux entreprises publiques de règles de gestion qui auraient reproduit le comportement d'une entreprise privée, afin d'obtenir un input mix et un niveau de production optimaux. Lange ignorait totalement les problèmes d'incitation et de collusion inhérents aux mécanismes qu'il proposait. Ces problèmes n'ont été traités que bien plus tard, voir, par exemple, Laffont (2000).

De plus, même s'il était possible de mesurer la valeur des soins pour les consommateurs, par exemple par des méthodes d'évaluation contingente telles que celles utilisées en économie de l'environnement (voir le complément C sur les différentes méthodes d'évaluation de la valeur de la vie humaine), fixer les tarifs sur la base de telles études reviendrait au mieux à s'assurer que le coût des ressources consacrées à la santé reste en deçà de leur valeur *moyenne* pour la société. D'une part cela n'empêche pas que cette valeur moyenne soit supérieure à la valeur *marginale*, qui est celle qui devrait être prise en compte. Cette différence crée un biais vers une quantité de soins trop élevée. D'autre part la valeur estimée des soins correspond à un niveau de production qui n'est pas le niveau optimal, mais celui qui prévaut dans la réalité, alors que le signal-prix que l'on devrait envoyer aux producteurs est celui qui correspond à un tel optimum. (Ceci est compliqué par le fait que l'offre et la demande ne coïncident pas nécessairement, voir encadré 3.) Pour atteindre la situation efficace, il faudrait donc mettre en place un système de tâtonnement où le système de prix est régulièrement mis à jour sur la base de nouvelles évaluations après que les fournisseurs de soins ont eu le temps d'ajuster leur offre à la nouvelle tarification. Il n'est par ailleurs même pas certain qu'un tel processus de tâtonnement converge vers une situation stable (voir, par exemple, Scarf, 1960). On voit là la difficulté de répliquer des mécanismes de marché à l'aide d'outils centralisés, même dans le cas où ceux-ci empruntent au marché certains de ses instruments tels que le système de prix.

Ces difficultés inhérentes à la mesure de l'utilité peuvent en principe être surmontées en utilisant, à la place des évaluations contingentes, des mesures empiriques des QALYs<sup>(22)</sup>, qui permettent de valoriser des incréments de santé et pas uniquement la bonne santé en tant que telle. De cette façon, une valorisation, certes approximative, de la valeur marginale de la bonne santé, est rendue possible. De fait, la vaste littérature sur les QALYs tente de mesurer des paramètres relativement fins de l'utilité individuelle, tels que l'aversion pour le risque, l'aversion à la perte ou encore la non-constance des taux d'actualisation (voir, par exemple, Nyman, 2011 ou encore Sambuc, 2011 pour des discussions récentes). Mais il n'en reste pas moins qu'elle repose en général, tout comme l'évaluation contingente, sur des questionnaires, c'est-à-dire sur la réponse des individus à des situations imaginaires et non réelles.

Dans la pratique, les tarifs de la T2A ne sont pas censés représenter la valeur hédonique des biens fournis, mais plutôt leur coût moyen de production, puisque c'est sur la base d'une estimation de ces coûts qu'ils sont officiellement fixés. En quelque sorte, on abandonne l'idée de transmettre aux producteurs un signal sur la valeur sociale de leur activité. À la place, le système de tarification est utilisé comme procédure de « *benchmarking* »

(22) QALY est l'acronyme de *Quality Adjusted Life Years*. Le QALY est utilisé pour fournir une mesure cardinale de l'état de santé d'un individu. Le QALY est compris entre 0 (niveau le plus bas, correspondant au décès) et 1 (niveau le plus haut, correspondant à la parfaite santé). Pour plus de détails, voir le complément C au présent rapport.

qui permet de mettre en concurrence les établissements sanitaires sur le plan de l'efficacité. Les objectifs qui sous-tendent cette méthode de tarification sont sensiblement moins ambitieux que l'approche qui voudrait rétablir un pseudo-système de prix au sein d'un mécanisme de planification centralisée, dans le but de répliquer les mécanismes d'équilibre de marché pour parvenir à l'allocation optimale des soins. Il ne s'agit ici que de rémunérer chaque thérapie sur la base d'une estimation raisonnable de son coût, plutôt que d'entériner mécaniquement les dépenses encourues pour chaque maladie. Il est alors naturel de fonder cette estimation de coût sur la pratique des autres établissements, ce qui permet de réduire considérablement les problèmes d'aléa moral en désindexant le tarif versé du coût encouru par l'établissement (contrairement aux pratiques antérieures à la T2A), tout en s'assurant qu'il existe un lien entre les tarifs pratiqués et le coût effectif de traitement de la maladie considérée. *A priori*, on est en droit d'attendre d'un tel système que les établissements les plus efficaces soient les plus rentables, puisqu'ils perçoivent la même rémunération que leurs concurrents pour un traitement donné alors que leurs coûts sont plus faibles. Il reste à savoir dans quelle mesure de telles différences de rentabilité vont se traduire par une réallocation des ressources, et si cette réallocation va dans le sens de l'efficacité. Cette question complexe sera traitée plus bas, et elle est cruciale pour déterminer dans quelle mesure la T2A peut contribuer à améliorer l'efficacité productive de l'offre de soins si les établissements sont hétérogènes.

### 3.2.2. La T2A et la structure de coûts des producteurs de soins

Si l'on fait abstraction de ce problème, quels sont les effets attendus de la tarification par les coûts sur l'offre de soins ? La réponse à cette question dépend crucialement de la structure de coûts de la production hospitalière ainsi que de la nature de l'hétérogénéité entre fournisseurs de soins. On peut considérer différents cas de figure :

- supposons tout d'abord que les établissements soient identiques, et que leur coût marginal soit constant, c'est-à-dire que le coût unitaire de chaque établissement soit le même quelle que soit la quantité de biens qu'il produit. Il s'agit de la situation dite de rendements constants. Dans ce cas, le coût unitaire d'un soin donné est le même dans tous les établissements et il correspond aussi au coût marginal de production, quel que soit l'établissement considéré. Dans ce cas, la T2A ne fait que rembourser ce coût et son impact est le même pour tous les établissements. Elle n'est alors rien d'autre qu'une comptabilité précise des dépenses afférentes à chaque maladie et tous les établissements sont incités à effectuer les actes et dissuadés d'accroître artificiellement leurs coûts ;
- supposons maintenant que les coûts marginaux d'un établissement donné soient constants, mais que ceux-ci diffèrent entre établissements, ce qui signifie que (pour la thérapie considérée) certains établissements sont plus efficaces que d'autres. Le tarif T2A est calculé

### 3. Les pièges d'une tarification fondée sur l'évaluation de la demande

Au-delà des difficultés techniques inhérentes au fait de vouloir mesurer la vraie valeur sociale de la santé (difficultés qui incluent le fait que même si l'on obtenait le consentement à payer des consommateurs, celui-ci ne prendrait pas en compte les effets externes)<sup>(\*)</sup>, l'idée de fonder les tarifs sur une mesure directe de la valeur hédonique des soins se heurte à deux problèmes fondamentaux :

- l'écart entre la propension marginale à payer et la propension moyenne (sauf à utiliser exclusivement les QALYs) ;
- le fait que le niveau initial des soins ne coïncide pas nécessairement avec l'optimum que l'on cherche à atteindre, ce qui implique que la propension à payer que l'on mesure sera différente de celle que l'on veut utiliser pour déterminer les tarifs, qui est la propension marginale à payer à l'optimum.

Du fait que le coût des soins est pris en charge par la collectivité, le prix acquitté par le patient (dit encore reste à charge) est inférieur au coût des soins et, *a fortiori*, au tarif T2A remboursé à l'offreur. Pour fixer les idées et simplifier la discussion, nous nous placerons dans le cas limite où le patient acquitte un prix nul. La demande de soins s'établit donc au niveau maximum correspondant à ce prix. À ce niveau, les patients désirent soigner, dans les meilleures conditions possibles, la totalité de leurs affections, du moment que les désagréments de la cure sont inférieurs à la valeur hédonique de la guérison. C'est le niveau de satiété des soins, celui qui correspond à la vision selon laquelle « la santé n'a pas de prix ». Quant à l'offre, elle s'établit au niveau correspondant au tarif fixé par la T2A. On a donc deux cas de figure :

- soit le tarif est élevé, et l'offre est supérieure au niveau de saturation de la demande. Dans ce cas la quantité de soins s'établit au niveau de cette demande, et la tarification transfère des rentes de la collectivité vers le secteur de la santé ; une baisse des tarifs réduirait les dépenses sans avoir d'effets sur la quantité des soins. Le secteur de la santé, cependant, est en sous-capacité : les acteurs voudraient augmenter leur activité, et pourraient induire une demande plus élevée en prescrivant des actes inutiles ;
- soit le tarif est faible, et l'offre de soins est alors inférieure au niveau de satiété de la demande. Ce sont alors les patients qui sont rationnés, conséquence inévitable de la gratuité. La quantité de soins est celle qui égalise le coût marginal de production avec le tarif T2A. Toute baisse du tarif se traduit par une diminution du volume des soins.

Supposons maintenant que l'on essaye de fixer le tarif en estimant la propension moyenne des consommateurs à payer pour les soins. Le cas de figure le plus « canonique » est celui où cette propension est estimée dans une situation initiale qui correspond à l'équilibre concurrentiel entre l'offre et la

(\*) Comme, par exemple, l'externalité positive que représente le fait de se faire vacciner, sur le risque d'autrui d'être contaminé.

demande. Cette situation est purement hypothétique dans la mesure où un marché concurrentiel pour la santé n'a pas été observé depuis plusieurs décennies ; mais elle permet de fixer les idées. Dans une telle situation, dans la mesure où l'utilité marginale de la santé est décroissante, une évaluation de la propension à payer pour les soins nous donnerait une estimation moyenne supérieure à la propension marginale. Or, c'est cette dernière qui est pertinente, car il est optimal de fixer le prix au niveau de la propension marginale à payer. En fixant le tarif au niveau de la propension moyenne, on stimule l'offre de soins au-delà du niveau « optimal » où le coût marginal des soins est égal à la propension marginale à payer des consommateurs. On déboucherait sur une situation où il y aurait « trop » de soins.

Supposons maintenant que la situation initiale diffère de la situation « d'équilibre » décrite plus haut – et il y a des raisons de penser que ce serait le cas, étant donné qu'historiquement, le secteur n'a pas été organisé selon des principes concurrentiels. Supposons également, pour faire abstraction de l'écart entre propension moyenne et propension marginale déjà analysé, que l'on puisse mesurer correctement la propension marginale à payer pour les soins. Alors, si la situation initiale est telle qu'il y a excès de soins par rapport au niveau d'équilibre (par exemple à la suite d'une tarification inadéquate pratiquée pendant de nombreuses années), la propension marginale estimée sera plus faible qu'au niveau « d'équilibre », et le régulateur sera naturellement conduit à mettre en place des tarifs bas qui conduiront à une nouvelle situation où le niveau des soins sera plus faible qu'à l'équilibre. Inversement, si le niveau des soins dans la situation initiale est plus faible qu'à l'équilibre (par exemple à la suite de recrutements trop restrictifs des professionnels de santé), alors on observera une propension marginale à payer plus élevée qu'à l'équilibre, ce qui conduira à des tarifs élevés et à une nouvelle situation où l'offre s'établit à un niveau supérieur au niveau d'équilibre.

Bien entendu, dans la nouvelle situation il convient de réestimer la propension marginale à payer pour les soins, ce qui conduit à un processus de tâtonnement cyclique. Le mécanisme graduel de fixation des prix que nous venons de discuter n'est donc pas très satisfaisant dans la mesure où il conduit à des fluctuations dans la fourniture totale de soins. Bien entendu, il est possible qu'un tel ajustement cyclique lent se produise également dans un marché concurrentiel standard, mais dans ce dernier cas ce sont les consommateurs qui fixent leur demande et il n'y a pas lieu pour l'État de réestimer constamment les propensions à payer.

Par ailleurs, le fait que la procédure de tarification que nous venons de discuter conduise à un niveau de soins qui diffère du niveau d'équilibre n'est pertinent que si c'est bien ce niveau d'équilibre que le système de santé essaie de reproduire. Cela sera le cas si par exemple les dépenses sont socialisées pour des raisons d'équité uniquement, si les préférences des agents sont telles pour la distribution tandis qu'il n'y a pas de raison de penser que le niveau d'équilibre global des dépenses soit inefficace. Si ce n'est pas le cas, par exemple à cause de phénomènes d'aléa moral du côté des professionnels, ou à cause des externalités associées à l'état de santé de la population, le problème de la tarification ne fait que s'aggraver, puisqu'il faut mesurer l'impact de ces distorsions pour être capable de calculer les tarifs optimaux.

sur la base du coût moyen de l'ensemble des hôpitaux ou d'un échantillon représentatif d'hôpitaux. La thérapie est donc bénéficiaire pour les établissements dont les coûts sont inférieurs à la moyenne mais déficitaire pour ceux dont les coûts sont supérieurs à la moyenne. On s'attend donc à ce que ces derniers renoncent à cette activité, tandis que les premiers auront tendance à accroître leur volume. On voit donc que la T2A réalloue l'offre de soins des établissements les moins efficaces vers les établissements les plus efficaces, ce qui est souhaitable du point de vue de l'efficacité économique. On notera ici que certains pays étrangers (Angleterre, Australie) ont pratiqué des expérimentations de tarifs plus « sévères » dans la mesure où plutôt que sur la moyenne des coûts, les tarifs de certains actes sont fondés sur les établissements les plus performants, par exemple le tarif est égal au coût moyen des 10 % d'établissements les moins coûteux (voir Or et Renaud, 2009). Cependant, à long terme<sup>(23)</sup> le système français exerce une pression sélective aussi forte. En effet, une fois que les établissements les moins performants se sont retirés du marché, le coût moyen baisse et la réévaluation des tarifs devrait conduire à une baisse de ceux-ci et à un nouvel écrémage des fournisseurs les moins productifs. On voit donc qu'à long terme, l'indexation des tarifs sur les coûts moyens ne devrait laisser subsister que les fournisseurs les plus efficaces. C'est précisément ce que voudrait un planificateur économique qui tenterait d'atteindre l'efficacité productive.

Une telle vision des choses se heurte évidemment à certaines objections. D'une part, on peut se demander s'il est « socialement acceptable » de fermer certains établissements ou tout au moins d'en retrancher certaines activités. Si ladite activité est transférée vers un établissement voisin, on voit mal quelles difficultés cela soulèverait<sup>(24)</sup>. Si par contre l'accès à ce type de soins disparaît dans une zone géographique donnée parce que l'établissement « inefficace » était le seul à le pratiquer, on peut s'interroger sur l'opportunité de cette disparition. En effet, la T2A ne prend alors pas en compte le coût supplémentaire de transport des patients vers les établissements « efficaces »<sup>(25)</sup>. Fondamentalement, ce problème vient du fait que le même traitement disponible à deux endroits différents constitue deux biens différents sur le plan économique, et il n'est donc pas logique d'inclure ces

(23) Les tarifs des différents groupes homogènes de malades (GHM), ainsi que leur classification, donnent lieu à des réajustements annuels, sur la base de l'Étude nationale des coûts à méthodologie commune (ENCC).

(24) Le cas où l'hôpital voisin ne peut accueillir les patients supplémentaires du fait de ses contraintes de capacité entre dans le cas de figure des coûts marginaux croissants qui est traité plus bas.

(25) L'aide à la contractualisation en France (DREES, 2009) permet cependant de disposer de quelques marges de manœuvre en termes de financement des établissements de santé, pourvu que la permanence des soins soit garantie. Aux États-Unis, les tarifs associés aux DRG (*diagnosis related groups*) équivalents américains des GHM peuvent donner lieu à des ajustements.



deux biens dans le même groupe homogène de malades (GHM). On voit là qu'il existe un arbitrage délicat dans la formulation de la T2A :

- soit les GHMs sont « grossiers » et la tarification sur la base du coût moyen conduit à la disparition de services utiles pour lesquels il n'existe pas de substituts évidents ;
- soit ils sont « fins » et le système devient complexe, le nombre de tarifs à fixer élevés et le nombre d'établissements utilisés pour la tarification d'un GHM est faible.

Dans ce dernier cas, on obtient une mesure bruitée du coût réel et rien ne garantit que, parmi le petit nombre d'établissements impliqués dans un GHM donné, il y en ait au moins un qui se situe sur la frontière d'efficacité productive. Le tableau 4 présente le nombre de DRG (*diagnosis-related groups*, soit la dénomination internationale pour les GHM) pour différents pays, dont la France. Il apparaît que le nombre de GHM est sensiblement plus élevé en France que dans les autres pays cités.

#### 4. Principales caractéristiques de la classification par CMD et GHM en 2008

	France	Allemagne	Royaume-Uni	États-Unis	Pays nordiques
Nombre de DRG (GHM)	2 297	1 200	1 389	1 258	794
Nombre de catégories majeures de diagnostics	28	26	23	25	28

Source : Busse et al. (2011).

Pour en revenir à l'exemple ci-dessus, le problème gît dans le fait que la tarification en fonction du coût moyen ignore la propension à payer des patients (ou de la société) pour le soin considéré, c'est-à-dire ignore le côté de la demande, comme on l'a relevé plus haut. On peut concevoir qu'il soit socialement désirable de maintenir un établissement de type A dans la région 1 alors qu'il en existe un plus efficace de type B dans la région 2, parce que la différence d'efficacité productive entre ces deux types d'établissement est inférieure à la fois aux coûts de transport des patients de la région 1 vers la région 2 et aux coûts d'installation d'un établissement de type B dans la région 1. Un autre aspect du problème est que la T2A ne prend en compte que le coût des soins supporté par l'établissement considéré, et pas le coût des prises en charges alternatives. Ce problème est bien identifié dans la littérature. Par exemple, dans l'exemple précédent, si la somme du coût des soins dans l'établissement B et du coût de transport de la région 1 vers la

région 2 est supérieure au coût des soins dans l'établissement A, alors la suppression de l'établissement A sous l'effet de la T2A conduit finalement à une hausse des dépenses de santé, à services rendus inchangés ! On voit là toute la difficulté de mise en œuvre d'un système de pseudo-marché au sein d'une économie planifiée où les mécanismes concurrentiels de fixation des prix ont été neutralisés.

Les choses se compliquent encore si le coût marginal de production diffère du coût moyen. Dans ce cas, des difficultés supplémentaires apparaissent car c'est en principe sur le coût marginal et non sur le coût moyen que les tarifs devraient être indexés. Si les rendements sont décroissants, le coût marginal est supérieur au coût moyen, ce qui signifie que, si le prix est fixé au coût moyen, on sera amené à tarifier à un niveau trop bas et à avoir une offre de soins trop faible. Dans ce cas, une application mécanique de la T2A devrait même conduire à la disparition des soins ! En effet, les établissements devraient constamment réduire leur activité parce que les soins marginaux ne sont pas rentables ; ils devraient faire en sorte que le coût marginal du dernier soin proposé coïncide avec le tarif. Mais alors ils se mettent dans une situation où le coût moyen est à nouveau inférieur au tarif, et une réévaluation des coûts devrait conduire à une baisse des tarifs et à une nouvelle vague de réduction de l'activité. On voit là qu'il est difficile de mettre en place un système de tarification réglementée en l'absence d'observations sur la structure de coûts des entreprises considérées<sup>(26)</sup>. Par ailleurs, à moyen terme, si les établissements sont hétérogènes, la contraction des établissements les moins efficaces peut conduire à une hausse excessive de la demande de soins dans les établissements les plus efficaces. Ceux-ci, plutôt que de voir leurs coûts marginaux augmenter au-delà du tarif T2A sous l'effet des rendements décroissants, préféreraient rationner les soins. Dans un tel cas de figure, la T2A aurait l'effet « pervers » de réduire la quantité totale de soins<sup>(27)</sup>.

Si les rendements sont croissants, le coût moyen est supérieur au coût marginal, et on a alors affaire à une situation de monopole naturel où, à prix fixé, les établissements de santé ont intérêt à accroître leur activité sans limites pour amortir leurs coûts fixes. On s'attend alors à ce que les établissements dont le coût marginal est supérieur au tarif ferment leur porte (une réduction de l'activité ne ferait qu'accroître leurs pertes), tandis que les autres finiraient par produire la quantité maximale de soins (celle correspondant à la satiété des besoins). Une telle situation est donc peu susceptible de conduire à une réduction des dépenses de santé, d'autant que tant qu'il subsiste plus d'un établissement profitable, ceux-ci ont intérêt à utiliser leurs

(26) Ces difficultés font l'objet de la vaste littérature sur la réglementation optimale des entreprises en présence d'asymétrie d'information, voir par exemple Laffont (2000). Suivant cette littérature, on pourrait envisager la mise en œuvre d'un « contrat optimal » entre État et hôpitaux mais un tel contrat serait sans doute bien plus complexe que celui inhérent à une tarification uniforme par GHM.

(27) Cet effet n'est effectivement pervers que si la quantité de soins dans la situation initiale n'est pas excessive.

ressources pour attirer des patients supplémentaires au détriment de leurs concurrents, ce qui conduit à une dissipation de ressources. Par ailleurs, en présence de rendements croissants, l'optimum consisterait à fixer le tarif à la valeur du coût marginal, mais ceci conduit alors à des pertes ce qui est en contradiction avec les principes de la T2A. On voit donc que la T2A ne fonctionnerait pas non plus très bien en présence de rendements croissants.

Enfin, il existe des cas de figure plus complexes où les rendements sont croissants ou décroissants selon le niveau de production. Dans une telle situation on a tendance à considérer que le coût marginal est croissant avec le niveau d'activité, tandis que le coût moyen décroît puis croît au-delà d'un certain seuil. Ces variations reflètent la présence de coûts fixes associés à des rendements marginaux décroissants. Dans une telle situation, chaque établissement a une taille « efficace » qui est celle qui minimise son coût moyen. Supposons que nous appliquions un tarif égal au coût moyen entre établissements. Tous les établissements dont le coût moyen minimal est inférieur ou égal à ce tarif maintiendront leur activité et choisiront un niveau de soins tel que le coût marginal est égal à ce tarif. Les autres cesseront leur activité. Dans la nouvelle situation, le coût moyen est inférieur au tarif pour tous les établissements. Une réévaluation des tarifs devrait conduire à une nouvelle baisse et à une sortie du marché de quelques établissements. Le processus devrait continuer jusqu'au point où les seuls établissements qui subsistent ont un coût moyen minimal, un niveau de soins correspondant à la taille efficace et font des profits nuls. Une fois de plus, seuls les établissements les plus efficaces subsistent, c'est-à-dire ceux dont le coût moyen minimal est le plus faible. Le problème est qu'ils produisent exclusivement la quantité de soins correspondant à ce minimum, alors qu'en présence de rendements constants ils étaient capables d'accommoder n'importe quelle quantité de soins. Rien ne garantit donc que dans cette situation la quantité de soins soit adéquate. En principe on pourrait laisser aux établissements une marge en fixant le tarif au-dessus du coût. Mais dans ce cas on conservera certains établissements inefficaces et il faudrait en principe calculer le niveau de la marge qui, compte tenu de la répartition des établissements selon leurs différentes structures de coûts, nous donnerait la quantité « optimale » de soins. Cette discussion illustre le paradoxe selon lequel si l'on est capable de calculer exactement le tarif optimal dans un contexte de pseudo-marché, alors on a sans doute suffisamment d'information pour administrer directement le secteur considéré et donc on peut se passer d'un système de tarifs. Une autre solution, plus intelligente, est de répliquer les établissements les plus efficaces, puisque ceux-ci ont la taille optimale, afin d'obtenir le volume désiré de soins au niveau global. Ceci pourrait résulter par exemple d'un mécanisme de libre entrée des cliniques privées<sup>(28)</sup>.

(28) Cependant cette approche ne résout pas la question, mentionnée plus haut, de la déconnexion entre la tarification T2A et la demande de soins. Le nombre total de telles cliniques reste donc problématique.

La discussion qui précède concerne un système de tarification sur la base des seuls coûts observés. En pratique, il est exact que les coûts sont les principaux déterminants du tarif, cependant ils ne sont pas les seuls déterminants. À la marge, les autorités ajustent, de façon bien souvent discrétionnaire, les tarifs de façon à créer des incitations en faveur de certains types ou catégories de soins<sup>(29)</sup>. Un exemple typique concerne les changements de tarification depuis 2004 dans le but de favoriser la chirurgie ambulatoire au détriment de l'hospitalisation proprement dite. Ces mesures ont consisté, dans un premier temps, à aligner le tarif pour une hospitalisation d'une durée d'un jour sur celui de l'ambulatoire, en confondant les deux catégories. Ceci signifie, dans la mesure où la chirurgie ambulatoire est évidemment moins coûteuse, que les praticiens ont intérêt à la mettre en œuvre plutôt qu'une hospitalisation de courte durée chaque fois que cela est possible. Parallèlement, un processus de baisse des tarifs pour l'hospitalisation de plus de deux jours se poursuivait, à tel point que l'alignement du tarif avec celui de l'ambulatoire a été atteint en 2009 pour un certain nombre de maladies (tandis que, pour des raisons peu évidentes, l'hospitalisation d'une journée sort de la catégorie ambulatoire pour rejoindre à nouveau celle des hospitalisations supérieures à deux jours).

On peut s'interroger sur une telle pratique. Si on traite un GHM comme un bien homogène, alors il n'y a pas lieu de distinguer entre différents traitements de la même maladie selon la durée d'hospitalisation (voir discussion ci-dessous). Le fait que deux tarifs sont pratiqués signifie donc que l'on opère une distinction entre deux états de gravité pour la même maladie et que l'on considère que les traitements pour ces deux états constituent deux biens économiques différents, qui doivent être rémunérés différemment. Mais si l'on adopte cette perspective, que signifie alors la volonté d'aligner les prix de ces deux biens et de favoriser l'ambulatoire plutôt que l'hospitalisation de plus longue durée ? Il est évident que l'on considère ces deux traitements comme fortement substituables et que la raison essentielle pour laquelle on veut privilégier l'ambulatoire est que celle-ci est moins coûteuse. Selon cette interprétation, la T2A serait utilisée comme contrat « incitatif » en présence d'aléa moral de la part des hôpitaux, qui auraient intérêt à classer chaque maladie dans la case qui lui rapporte le plus, et donc à en exagérer la gravité. Cependant, cette interprétation purement incitative n'est pas complète, car on sait aussi que l'ambulatoire a de meilleures propriétés « hédoniques » – en effet, cette pratique réduit l'incidence des maladies nosocomiales et est préférée par les patients qui la jugent moins démoralisatrice que l'hospitalisation de nuit. Selon cette interprétation, le décideur public « sait » que la « demande » de chirurgie ambulatoire est élevée, relativement à l'hospitalisation, et utilise ce savoir pour refléter cette demande dans les tarifs. Bien entendu, se pose alors la question de savoir d'où vient cette connaissance ; sans doute d'enquêtes qualitatives, mais, en l'absence de mécanisme de marché, on voit mal comment traduire de façon systématique le résultat de telles enquêtes sous la forme de tarifs

(29) Ces mesures sont analysées en particulier par Cash et al. (2011).

On notera aussi que tous les GHM n'ont pas fait l'objet de mesures tarifaires en faveur de l'ambulatoire, comme par exemple le traitement des hémorroïdes. Cela signifie-t-il que le problème d'aléa moral ne concerne pas cette thérapie, ou encore que les patients valorisent moins l'ambulatoire pour cette maladie que pour les autres ? Sans doute pas. Il s'agit plus probablement d'une illustration supplémentaire du caractère discrétionnaire des tarifs, dont la fixation ne semble obéir à aucun principe systématique.

### 3.2.3. T2A et redéploiement de l'offre de soins

La tarification T2A est également un vecteur de la mise en œuvre des « plans santé ». Cash et *al.* (2011) dressent une liste impressionnante des circulaires dont l'objet est la revalorisation des tarifs de certaines maladies, dans le but d'accomplir les objectifs des divers plans santé. Clairement, il s'agit à nouveau de l'utilisation de la tarification comme « signal de demande » plutôt que comme instrument incitatif de régulation fondé sur les coûts<sup>(30)</sup>. On peut se demander s'il est possible d'utiliser le même instrument pour atteindre plusieurs objectifs ; à savoir, d'une part l'incitation à utiliser les thérapies les plus efficaces pour une maladie donnée, et d'autre part, la réallocation de l'offre de soins vers ce qui est considéré comme prioritaire dans le cadre d'un plan santé ou d'un autre projet politique. Si un marché alloue efficacement l'offre et la demande au moyen d'un seul prix, c'est parce que celui-ci s'ajuste sous l'effet de la concurrence, comme on l'a décrit plus haut. Si par exemple la demande pour le traitement de l'épilepsie – qui a fait l'objet d'une revalorisation en février 2007 – augmente, on s'attend à une hausse du prix de ces soins à court terme. Comme ces soins sont relativement bien rémunérés, il devient profitable de les offrir, et c'est l'augmentation de l'offre qui fait baisser les prix par le jeu de la concurrence et les ramène au niveau des coûts (marginiaux) de production. Clairement, la hausse des tarifs du traitement de l'épilepsie de février 2007 tente d'imiter l'effet d'impact sur les prix qu'aurait une hausse de la demande dans un cadre concurrentiel. Mais la différence est que le tarif s'écarte alors du coût et qu'il n'y a pas de mécanisme clair pour le ramener vers le coût<sup>(31)</sup>. En principe le gouvernement devrait réindexer le tarif sur le coût

(30) Notons cependant que la tarification n'est pas le seul vecteur de transmission des politiques de santé. Pour de nombreux objectifs, les ressources supplémentaires sont versées aux établissements au titre des missions d'intérêt général (MIG). Il s'agit là de transferts directs, « inframarginaux », d'une somme d'argent qui doit être utilisée par l'établissement pour satisfaire à la mission correspondante. C'est par exemple le cas des soins palliatifs dans la circulaire du 21 février 2007. La répartition des soins ciblés par les politiques de santé entre ceux qui font l'objet de réajustements tarifaires et ceux qui bénéficient d'une enveloppe au titre des MIG a de quoi laisser songeur. La logique à laquelle cette répartition obéit, n'est pas très claire. L'analyse économique suggère que l'instrument tarifaire permet de stimuler l'offre lorsque celle-ci est élastique au prix, tandis que le versement d'une somme globale est plus adapté au cas où l'offre est peu réactive au prix et où la mission considérée correspond à une charge bien définie imposée à l'établissement. Mais rien ne permet de penser, dans les arbitrages qui sont faits entre les deux instruments, que la répartition correspond à cette distinction.

(31) Lorsque celui-ci est constant ou particulièrement peu élastique à la quantité de soins produite.

dès lors qu'il a observé une réponse adéquate de l'offre. Mais il n'est pas clair que le gouvernement possède l'information nécessaire pour estimer le surcroît d'offre qui justifierait un retour à l'alignement des tarifs sur les coûts. À court terme, il lui est également difficile de calculer le niveau de prix optimal lui permettant d'atteindre ses objectifs. Plus le prix est élevé, plus la réponse de l'offre l'est également et plus on atteindra rapidement la situation où on a réalloué les soins selon ce qu'on voulait au départ. Mais, par ailleurs, plus le prix est élevé, plus l'impact sur les dépenses de santé est défavorable et plus la rente transférée aux établissements est importante. Ajoutons qu'en principe, comme le patient ne paie pas, une hausse du tarif conduira à une hausse de l'offre de soins uniquement dans le cas où la demande est rationnée. Dans la configuration où le niveau « de satiété » de la demande est atteint, la hausse des tarifs est un pur transfert de rente au profit de l'établissement. Ce genre de phénomène illustre à nouveau la difficulté de mettre en place un pseudo-marché avec des tarifs administrés et pourrait contribuer à expliquer pourquoi les études qui tentent de mettre en évidence un effet incitatif des tarifs sont peu concluantes (voir Cash et *al.*, 2011).

#### 3.2.4. L'articulation T2A/ONDAM

Les choses sont rendues encore plus complexes du fait que la T2A est censée répondre à un troisième objectif, à savoir la maîtrise budgétaire des dépenses de santé. L'idée est que si la progression des dépenses au cours d'une année dépasse l'objectif fixé par le décideur politique (dit « ONDAM », objectif national des dépenses d'assurance-maladie), alors une baisse des tarifs est mise en place afin de rééquilibrer les dépenses pour les ramener à moyen terme dans le cadre fixé par l'ONDAM. On voit comment ce système, qui s'inspire des mécanismes de redistribution à contributions définies, constitue un outil puissant de régulation des dépenses. Mais il a le défaut d'assigner à la tarification un troisième objectif, outre celui de l'efficacité productive et de l'implémentation des politiques de santé : le contrôle des dépenses totales. Le tableau 5 présente, à des fins de comparaison, la nature de la contrainte budgétaire pesant sur les hôpitaux dans différents pays de l'OCDE en 2009.

La question de la demande induite – le fait que les praticiens de la santé soient susceptibles d'accroître la production de biens de santé au-delà du volume minimal nécessaire – suggère que certaines dépenses pourraient être écartées, sans nuire à la santé des patients. C'est pour éviter ce problème que l'on s'accorde à penser qu'un mécanisme de contrôle des dépenses est nécessaire. Cependant, le système actuel soulève deux questions : faut-il jouer sur les tarifs et faut-il indexer négativement ceux-ci sur les dépenses globales ? Comme le relèvent Or et Renaud (2009), ce système peut avoir des effets pervers importants. Si ce que l'on veut limiter est effectivement l'excès de prescriptions, alors on doit mettre en place un mécanisme individuel et non pas collectif. En effet, à moins d'être sous le joug d'un fort impératif catégorique, un établissement donné considérera que s'il prescrit

trop d'actes, cela aura un effet très faible sur les dépenses de santé globales, dont il ne représente qu'une part infime. Sa propre surfacturation ne lui inflige donc qu'un coût négligeable, alors qu'il en retire le plein bénéfice. Par ailleurs, la baisse des tarifs, quoique modeste, qui résulte des dépenses additionnelles qu'il inflige à la collectivité, s'applique à tous les établissements. En surfacturant, il inflige donc une externalité négative sur les autres établissements, externalité qu'il ne prend pas en compte dans sa décision. Bien entendu, tous les établissements font face à des incitations similaires à se comporter en « passager clandestin », et on s'attend donc à un excès général de prescription. Ce problème bien connu de « tragédie des biens communs »<sup>(32)</sup> s'applique à tout système de régulation fondé sur les dépenses globales, qu'il utilise la tarification ou tout autre instrument.

Le fait de revoir la tarification à la baisse lorsque l'ONDAM est dépassé est problématique. Ce que l'on veut en principe, c'est infliger des pénalités aux établissements qui surprescrivent, mais on voit mal l'intérêt que celles-ci prennent la forme d'une baisse générale des tarifs. À terme, certains tarifs pourraient se retrouver en dessous de leur coût de production, ce qui pourrait se traduire par un rationnement des soins et est peu cohérent avec les autres objectifs assignés à la T2A. Certes, il se peut que le rationnement des soins, en présence de gratuité pour le patient et d'aléa moral du côté de l'offre, soit inévitable si l'on se fixe des limites en termes de dépense. Mais un tel rationnement devrait alors être pensé et organisé, et non reposer sur les effets d'une baisse automatique des tarifs. De plus, une telle baisse est politiquement peu crédible si le tarif est inférieur au coût, car les établissements pratiquant les soins correspondants feraient alors des pertes et l'annulation de ces pertes par l'État, ou encore un réajustement à la hausse du tarif incriminé, aurait une légitimité considérable.

La discussion qui précède suggère que d'un côté, la T2A n'est pas très efficace pour transmettre aux établissements des signaux-prix qui rendraient compte de la désirabilité sociale des divers traitements, mais que, d'un autre côté, on lui a assigné de trop nombreux objectifs et qu'il importe en particulier de repenser le système de contrôle global des dépenses d'assurance-maladie. Le tandem ONDAM/T2A ne crée en effet aucune incitation à limiter la surprescription. Lui substituer un autre système permettrait de renforcer la cohérence de la T2A en réduisant le nombre d'objectifs qui lui sont assignés.

---

(32) La tragédie des biens communs correspond au dilemme auquel sont confrontés les agents pouvant librement avoir un accès commun à une ressource en quantité finie. La consommation de chaque agent est trop faible pour que son effet se fasse sentir sur le bien commun, mais chaque agent est néanmoins tenté de sur-consommer le bien, car il ne tient pas compte de l'effet néfaste de son comportement sur les autres agents (on parle d'externalité négative). Parmi les exemples les plus fréquents, on trouve les pêcheurs qui, dans certains cas, doivent être contraints réglementairement à limiter leur activité pour éviter la disparition de certaines espèces.

### 5. Nature de la contrainte budgétaire pesant sur les hôpitaux publics dans plusieurs pays de l'OCDE en 2009

	France	Allemagne	Royaume-Uni	Italie	Japon	Suède	Canada
Contrainte budgétaire définie annuellement au niveau national	Oui Cible avec possible dépassement	Oui Cible avec possible dépassement	Oui Budget strict	Oui Budget strict	Non	Oui Cible avec possible dépassement	Oui Cible avec possible dépassement
Cible globale divisée en cibles par catégorie de dépense	Pour différents services de santé	Pour différents services de santé	Pour différents services de santé Par région/secteur	Par région/secteur	—	Pour différents services de santé Par région/secteur	—
Conséquences du dépassement de la cible de dépense au cours des cinq dernières années	Déficit créé ou accru Politiques de maîtrise des coûts	Politiques de maîtrise des coûts	—	Déficit créé ou accru Augm. des taxes et cotisations sociales Politiques de maîtrise des coûts	—	Déficit créé ou accru Politiques de maîtrise des coûts	Augm. des taxes et cotisations sociales

Source : Paris, Devaux et Wei (2010).



### *3.2.5. Vers un mécanisme de régulation davantage fondé sur les incitations individuelles ?*

Cette question soulève le problème plus général, et essentiel, de celui du contrôle des dépenses globales de santé et de sa pertinence. De façon caricaturale, on peut penser qu'historiquement, les objectifs globaux de dépenses ont été mis en place pour pallier les mauvaises propriétés incitatives du système. Cependant, satisfaire à ces objectifs globaux implique tout de même la mise en place d'un mécanisme incitatif au niveau individuel (ne serait-ce qu'un mécanisme de rationnement) susceptible d'aligner les décisions individuelles d'engagement de dépenses avec l'objectif financier global. Sans cela ce dernier restera évidemment lettre morte. Mais, dès lors que l'on ne peut pas faire l'économie d'un tel mécanisme, autant essayer de s'attaquer aux fondements du problème et mettre en place un système qui incite les agents à internaliser correctement les coûts et les bénéfices de leurs décisions. C'est ce que fait en partie la T2A (et nous discutons plus bas comment elle pourrait être renforcée) ainsi que les systèmes franchises/plafonds que nous discutons et recommandons à la section suivante. Dès lors que de tels systèmes sont mis en place et qu'on pense qu'ils sont satisfaisants, leur superposer un contrôle des dépenses globales est redondant voire nuisible. Un tel contrôle est justifié dans la mesure où l'on pense que le système d'incitations est inadéquat et conduit à des dépenses trop élevées. Mais si l'on a mis en place un système adéquat et qu'on observe une dérive des dépenses, cela signifie simplement que la demande de biens de santé augmente (à cause, par exemple, du progrès technique comme on l'a vu plus haut), ce qui n'est pas en soi inefficace. Bien entendu cela peut impliquer une hausse des dépenses publiques et signifie qu'il faut faire des économies, mais il est alors contre-productif de réaliser ces économies sur le poste de la santé. En d'autres termes si l'objectif poursuivi est la maîtrise des dépenses publiques c'est le niveau total de ces dernières qu'il faut piloter et non celui d'un poste particulier qui est susceptible d'évoluer légitimement en fonction des besoins et préférences des individus.

### **3.3. Le problème de l'homogénéité des GHM**

Comme on l'a relevé plus haut, un système de prix fonctionne d'autant mieux que chaque prix correspond à un bien homogène. Dans le cadre de la santé, ce principe se heurte au fait que « chaque malade est différent ». Il est évidemment très difficile de prendre en compte les spécificités de chaque malade tout en maintenant un mécanisme incitatif de régulation des coûts. La même maladie peut toucher deux patients différents avec des degrés de gravité différents. En général, soigner le patient le plus atteint est plus coûteux. Si la tarification est indépendante du degré de gravité du malade, il est moins profitable de soigner le cas le plus grave. On courrait donc le risque de refus de soin (dans un marché concurrentiel, et en faisant abstraction des problèmes de solvabilité, le malade le plus atteint paierait une somme plus élevée et le problème ne se poserait pas), refus d'autant plus grave que c'est

précisément le patient qui aurait le plus besoin d'être soigné qui ne le serait pas<sup>(33)</sup> ! En particulier, le coût moyen estimé, sur la base duquel les tarifs sont fixés, représente une moyenne entre plusieurs degrés de gravité, et est donc inférieur au coût de traitement des cas les plus graves. Ainsi, traiter ces cas les plus graves serait systématiquement déficitaire.

Pour pallier cette difficulté, on pourrait vouloir mettre en place une grille plus fine de définition des maladies et distinguer les degrés de gravité. Ceci permettrait de pratiquer un tarif d'autant plus élevé pour une maladie donnée que l'état du patient serait grave. Cette distinction est déjà présente dans les tarifs actuels puisqu'ils distinguent les hospitalisations inférieures à un jour de celles supérieures à un jour (mais, par ailleurs, à la suite du processus de réalignement décrit plus haut, le tarif est devenu indépendant de la durée d'hospitalisation pour nombre de maladies). Mais alors réapparaîtrait le problème d'aléa moral, puisqu'il serait profitable pour l'établissement de classer ses patients comme étant dans un état grave, afin de percevoir une rémunération plus élevée. Bien entendu, on pourrait toujours mettre en place un système de vérification *a posteriori*, qui permettrait de vérifier que les ressources mobilisées dans le cadre d'un traitement correspondent bien à sa catégorie tarifaire. Mais cela ne résoudrait que partiellement le problème car l'hôpital pourrait vouloir effectuer des actes superflus dès lors que leur tarif est strictement supérieur au coût.

Ces difficultés procèdent du fait que c'est le praticien et non le patient qui définit la nature des services que ce dernier doit consommer. Elles sont incontournables quelle que soit l'organisation des soins. Par exemple, dans un système purement privé, les praticiens ont également intérêt à accroître artificiellement le nombre d'actes, tout comme dans le cadre du paiement à l'acte qui prévaut encore pour la médecine de ville et qui était la norme à l'hôpital avant l'introduction de la T2A. Dans des systèmes de tarification incitative tels que celui discuté plus haut, c'est le risque inverse de sous-prescription qui est important : en effet, celle-ci permet de réduire le coût du traitement alors que son prix fixé par la T2A ne dépend que de la pathologie et non des ressources engagées. Soit l'on s'en remet à la déontologie des professionnels et l'on s'attend à des prescriptions « justes » de leur part, et dans ce cas il est inutile de s'appuyer sur la tarification pour inciter à une pratique prescriptive correcte. Soit on pense qu'il existe un biais certain vers la surprescription, et dans ce cas on peut vouloir mettre en place des mécanismes de contrôle qui ne passent pas par la tarification. Ceci permettrait de réduire le nombre d'objectifs contradictoires qu'on a tendance à lui assigner. On pourrait par exemple mettre en place des systèmes d'information et de responsabilisation du patient, ainsi que d'évaluation des différents praticiens (comme c'est en partie le cas aujourd'hui avec le site Ameli)<sup>(34)</sup>.

(33) Il n'y a pas à l'heure actuelle de raison de penser que ce type de comportement prévaut dans la pratique, mais il pourrait émerger à l'avenir dans la mesure où la logique économique prendrait plus de place dans les décisions.

(34) À l'heure actuelle, le site Ameli permet de comparer les tarifs des différents praticiens. Par ailleurs, certains sites privés proposent aux patients de noter leurs médecins.

#### 4. Le site Ameli

Ameli, pour assurance-maladie en ligne, est l'interface électronique à laquelle tous les assurés ont accès. Ameli regroupe les fonctionnalités suivantes :

- possibilité de consulter son dossier de remboursement de frais de soin ;
- possibilité d'avoir accès aux différents praticiens par spécialité (ameli-direct), ainsi qu'aux tarifs pratiqués par ces praticiens, avec à titre de comparaison la base sur laquelle est défini le remboursement du régime général de la Sécurité sociale. Le cas échéant, les dépassements figurent clairement, avec parfois des intervalles de tarifs possibles ;
- possibilité d'avoir accès aux différents centres de soin (ameli-direct), avec une grille de recherche semblable à celle des praticiens.

Ameli dispose également d'un espace dédié aux professionnels de la santé qui leur permet, entre autres, de disposer de l'état des remboursements (tiers payant) et d'informations sur la patientèle.

Un autre aspect de la question est celui de l'hétérogénéité entre établissements en ce qui concerne la *qualité* des soins. Les tarifs T2A ne tiennent pas compte de la qualité. Dans la mesure où celle-ci est coûteuse et ne peut être facturée au patient, la T2A impose une borne supérieure à cette qualité. Par ailleurs, on peut s'attendre à ce que les établissements les moins performants sur le plan des coûts jouent sur la qualité des soins. Une certaine forme de concurrence par la qualité (plutôt que par les prix qui sont réglementés et ne sont pas acquittés par le patient) pourrait voir le jour. Les établissements les plus performants offriraient une qualité plus élevée et attireraient plus de patients. À taille donnée, cependant, cela signifierait un accès aux soins plus difficile dans ces établissements. Les patients auraient ainsi le choix entre des soins « rapides », mais de qualité médiocre, et des soins « rationnés », mais de bonne qualité. En quelque sorte, une telle concurrence par la qualité émule la concurrence par les prix, et comme elle, elle devrait pouvoir conduire à une réallocation des soins vers les établissements les plus efficaces, pour autant que ceux-ci puissent croître pour accommoder une demande supplémentaire.

La question de savoir comment fixer les tarifs pour cibler une certaine qualité « acceptable » dans les établissements qui se trouvent sur la frontière d'efficacité est plus délicate. En théorie, on pourrait définir un seuil d'acceptabilité pour chaque maladie, mesurer le coût correspondant à ce seuil dans les meilleurs établissements, et fixer le tarif à ce coût. Les établissements publics n'auraient alors, évidemment, aucun intérêt à accroître la qualité au-delà du seuil, mais les patients désireux d'un niveau de qualité supérieur à celui garanti par le service public pourraient se faire soigner dans le sec-

teur privé, où le supplément de qualité leur serait facturé et resterait à leur charge. En pratique, on voit bien que les modalités actuelles de fixation des tarifs ne font aucune référence à la qualité, peut être parce qu'il ne serait pas politiquement acceptable de fonder la formation des tarifs officiels sur l'hypothèse que certains patients sont mieux soignés que d'autres, parce que les premiers ont eu accès à un meilleur établissement. Le coût moyen sur lequel sont indexés les tarifs reflète donc sans doute une qualité « moyenne », et non une qualité « optimale » qui serait choisie en connaissance de cause par le décideur public. Certes on peut supposer que la perception d'un déficit qualitatif pour un GHM donné puisse conduire, quoiqu'à l'aveuglette, à une hausse discrétionnaire des tarifs pour ce GHM, mais comment s'assurer qu'il existe un mécanisme économique fiable qui garantirait que cette hausse de tarifs se traduise dans les faits par une amélioration de la qualité, plutôt que par un transfert de rentes vers les établissements ou par une augmentation du nombre d'actes<sup>(35)</sup> ?

Ceci nous conduit au troisième volet de notre problématique, à savoir l'absence de mécanisme clair de réallocation.

### **3.4. Le problème de l'absence de mécanisme de réallocation**

Un système de prix, oui, mais pour quoi faire ? Voilà l'objection qu'un sceptique serait amené à formuler à l'encontre de la T2A. Dans un mécanisme d'allocation fondé sur le marché, le système de prix n'est pas seulement un signal permettant aux agents d'intégrer l'utilité et le coût des biens qu'ils désirent produire ou consommer. Il crée aussi des incitations fortes à ce que la production soit allouée aux producteurs les plus efficaces et à ce que les consommateurs qui dérivent le plus de « bien-être » (exprimé en termes monétaires) d'une unité du bien considéré le consomment en priorité. Ces incitations sont elles-mêmes le résultat de la poursuite de l'intérêt personnel des participants au marché, c'est-à-dire que le prix est acquitté par le consommateur au producteur, de sorte que chacun a un intérêt personnel à se comporter de manière optimale. Le bien-être du consommateur augmente d'autant plus que le prix est faible et que l'utilité du bien consommé est élevée. Le profit du producteur est d'autant plus élevé que ses coûts sont faibles et le prix de vente élevé. Chaque participant au marché a intérêt à consommer d'autant plus (respectivement produire d'autant plus) que la différence entre le bien-être engendré par le bien et le prix (respectivement le prix et le coût de production) est élevée. Enfin, le marché converge vers un prix d'équilibre qui, sous certaines conditions, est socialement optimal, grâce à un mécanisme concurrentiel qui permet aux offreurs désirant vendre plus de pratiquer un prix plus faible.

(35) Il a par exemple été mentionné en auditions que certains départements ont un niveau anormalement élevé d'opérations de la cataracte.

La T2A repose sur un système de prix, ce qui suggère que l'on voudrait s'appuyer dessus pour répliquer les « bonnes » propriétés du marché en termes d'allocation. Mais le mécanisme d'allocation des ressources est très différent de celui du marché. D'une part, le prix n'est pas acquitté par le patient, mais par l'assurance-maladie et les mutuelles, bien que ce soit tout de même le patient qui choisisse l'endroit où il se fait soigner. D'autre part les hôpitaux publics ne sont pas des entreprises capitalistes qui maximiseraient le profit. La question de savoir comment ils se comportent face aux tarifs, en d'autres termes quelle est la « fonction objectif » qu'ils cherchent à maximiser, reste donc ouverte. Enfin, les tarifs sont administrés et la concurrence par les prix ne joue pas, sauf peut-être dans le secteur privé à travers les dépassements d'honoraires.

En l'absence d'un objectif de maximisation des profits, il est difficile d'avoir une vision claire du comportement d'un hôpital (ou de tout autre établissement public). À une extrémité du spectre, on peut penser que celui-ci pourrait se contenter de vouloir rentrer dans ses frais. Il continuerait à opérer tant que ses revenus restent supérieurs à ses coûts, accumulant au fil du temps des réserves financières sans stratégie très claire pour les utiliser. Il n'aurait pas d'incitation particulière à réduire ses activités déficitaires, ni à développer celles qui lui sont profitables, puisque seul l'équilibre budgétaire global lui importerait<sup>(36)</sup>. On voit bien qu'un hôpital qui se comporte de cette façon sera peu réactif aux incitations tarifaires et que dans un monde peuplé de tels hôpitaux, la raison d'être d'un système de prix n'est pas claire. Dans les faits, la mise en place des pôles ainsi que le passage à la T2A ont toutefois donné lieu, quoique très progressivement, à des réorganisations dont l'un des objectifs est d'accroître l'efficacité dans les démarches comptables – embauches de nouveaux contrôleurs de gestion, de statisticiens – ainsi que les incitations – mesures d'intéressement (Moison et Pépin, 2010). Les changements de comportement des différents acteurs, et la meilleure articulation des responsabilités entre la direction des établissements hospitaliers et le personnel médical, sont en cours, mais nécessitent à l'évidence du temps. La loi a introduit depuis 1991<sup>(37)</sup> la possibilité de l'intéressement pour les personnels des établissements publics de santé, mais cet intéressement ne pouvait pas initialement donner lieu à des rétributions pécuniaires. C'est désormais possible, et un arrêté publié en 2010<sup>(38)</sup> fixe les modalités de versement de prime aux chefs de pôles des établissements publics de santé. Cette prime comporte une partie variable, dont le montant, ne pouvant excéder 2 400 euros annuels, dépend de la réalisation ou non des objectifs figurant dans le contrat de pôle.

(36) Bien évidemment, la seule hypothèse d'équilibre des comptes est peu structurante : il existe un grand nombre de portefeuilles d'activités et de modes d'organisation compatibles avec cet équilibre. Il est naturel de supposer qu'une telle entité se comporterait selon le modèle du « *business as usual* », reproduisant son portefeuille d'activité initial tout en accommodant, dans la mesure du possible, les fluctuations de la demande pour chacune des activités.

(37) Loi n° 91-748 du 31 juillet 1991.

(38) Arrêté du 11 juin 2010 fixant le montant et les modalités de versement de l'indemnité de fonction des chefs de pôles.

À l'autre extrémité, on peut considérer un établissement qui se développerait selon une logique bureaucratique, en tentant de maximiser la taille de ses activités, sous la contrainte de l'équilibre budgétaire. Un tel établissement est relativement efficace : les activités déficitaires grèvent son budget et limitent donc la taille de ses opérations ; les activités excédentaires dégagent des marges qui permettent de financer une augmentation du niveau général d'activité. L'établissement a donc intérêt à réduire les premières et à développer les secondes. Qualitativement, il se comporte comme une entité qui maximiserait le profit. Relativement à cette dernière, il a cependant intérêt à pratiquer un volume total d'activité trop élevé<sup>(39)</sup>.

Enfin, un établissement qui maximiserait le profit serait incité, face à un mécanisme comme la T2A où il ne peut influencer sur les tarifs, à choisir un niveau d'activité tel que le tarif pour chaque pathologie est égal au coût marginal de production. Une telle décision conduit à l'efficacité productive, et à une répartition efficace de l'activité entre établissements, mais elle ne garantit pas *a priori* que la quantité totale de soins offerte soit la bonne, à moins que les tarifs reflètent de façon optimale les besoins, ce qui n'est pas généralement le cas comme on l'a indiqué plus haut.

On voit donc que l'utilisation d'une grille de tarifs a d'autant plus de sens que les établissements maximisent le profit, mais qu'elle peut conduire à des situations inefficaces si ce n'est pas le cas. Il est naturel de penser que les cliniques privées sont dans cette situation, mais ce n'est sans doute pas le cas pour les hôpitaux publics.

Une autre marge importante à travers laquelle le marché atteint une allocation efficace est l'entrée et la sortie d'établissements. Cela signifie que les hôpitaux les moins efficaces doivent être fermés tandis que de nouveaux hôpitaux pourraient être ouverts, en particulier pour imiter les pratiques de ceux qui sont profitables. Ce mécanisme de « destruction créatrice » accélère la diffusion des meilleures pratiques, générant ainsi des gains de productivité. Mais, une fois de plus, il ne contribue sans ambiguïté à l'efficacité que si les prix sont calculés correctement. S'ils sont trop bas, la sortie d'établissements même déficitaires ne fait qu'accroître le rationnement des soins. S'ils sont trop élevés, l'entrée d'établissements même efficaces est susceptible d'induire une demande supplémentaire inutile et préjudiciable à l'équilibre des comptes publics.

Par ailleurs, ce mécanisme de destruction créatrice se heurte à la logique politique en ce qui concerne les hôpitaux publics. Lorsque ceux-ci sont déficitaires, ils peuvent se livrer à des activités de lobbying afin de faire supporter leurs pertes par la collectivité. On peut s'attendre à ce que de telles activités soient d'autant plus efficaces que la fermeture de l'établissement

(39) La maximisation des profits conduit à égaliser le prix de chaque activité avec son coût marginal. La maximisation du niveau d'activité conduit à choisir un coût marginal supérieur au prix pour chaque activité, d'où le biais vers un excès d'activité. Cependant, le ratio coût marginal/prix est le même dans toutes les activités, d'où l'efficacité relative de ce type de producteur.

est perçue comme un événement politiquement sensible, par exemple par les élus locaux. Les effets bénéfiques de la T2A sur la productivité seront particulièrement atténués si le politique ne peut s'engager de façon crédible à fermer les établissements déficitaires. On voit donc que le système de la T2A, en tant que pseudo-marché dont le but est d'obtenir une allocation efficace des soins sans que leurs coûts ne soient supportés par les bénéficiaires de ces soins, pâtit de deux carences :

- l'absence d'un mécanisme permettant de trouver le « bon prix », c'est-à-dire reflétant la demande et pas seulement les coûts – l'ajustement des tarifs en fonction des différents « plans santé » ne jouant ce rôle qu'imparfaitement, puisque surtout déterminé par les objectifs publics ;
- l'absence d'une courroie de transmission entre les signaux de profitabilité envoyés par les tarifs et les décisions des hôpitaux publics, en l'absence d'une fonction objectif claire pour ces derniers.

En ce qui concerne le premier problème, une solution évidente consisterait à augmenter le « reste à charge » du patient, ce qui ouvrirait de plus la possibilité d'une concurrence tarifaire entre les hôpitaux. Mais une telle vision procède d'un changement radical de la philosophie de notre système de santé, et mérite d'être discutée dans un contexte plus large, ce que nous faisons plus bas.

Une autre solution consisterait à reproduire les mécanismes de « tâtonnement » pour trouver graduellement les bons prix. Par exemple, on pourrait envisager que les autorités procèdent à des mesures de « pénurie » dans l'accès aux diverses thérapies, à partir par exemple de données sur les délais. Les traitements pour lesquels de telles pénuries apparaissent se verraient revalorisés<sup>(40)</sup>, tandis que les tarifs baisseraient pour les traitements pour lesquels un excédent apparaîtrait. Notons cependant que dans ce dernier cas, la mesure de l'offre excédentaire est problématique à cause de la demande induite. De plus, le mécanisme proposé délivre une quantité de soins correspondant au niveau de la demande à *prix nul*, et ceci est inévitable tant qu'une partie des frais n'est pas à la charge du patient. Si cette quantité est jugée trop élevée, alors une certaine forme de rationnement est inévitable, et il faudrait alors calculer la durée d'attente correspondant à la quantité optimale de soins, et calibrer le mécanisme de fixation des prix de manière à cibler cette durée optimale.

Concernant le second problème, une solution évidente serait, à nouveau, une privatisation des hôpitaux publics (à l'exclusion de leurs activités de recherche et d'enseignement). Une telle préconisation peut sembler choquante, compte tenu de l'attachement des Français au service public.

(40) Dans les faits, la gestion des phénomènes de file d'attente, à la mise en place de la T2A, a été opérée dans plusieurs pays, notamment en Australie. Un classement selon le degré d'urgence de l'intervention a été effectué, et des remboursements supplémentaires ont été conditionnés au fait de traiter prioritairement les cas considérés comme urgents (Or et Renaud, 2009).

Pourtant, elle est l'aboutissement de la logique de la T2A, puisque, comme notre discussion l'a montré, un système de prix n'a d'intérêt que si les producteurs auxquels il s'applique ont un objectif de rentabilité. Par ailleurs de nombreux exemples de délégation de la fourniture de services publics au privé existent : ramassage des ordures, transports publics, contrôle technique automobile, enseignement privé sous contrat, etc. Ces méthodes ont l'intérêt de ménager une distance entre le décideur public et les producteurs, ce qui accroît la transparence, permet de mieux faire jouer la concurrence et de se débarrasser des producteurs inefficaces. Et elles ne remettraient pas en cause la réglementation des tarifs et l'égalité d'accès aux soins, qui resteraient inscrites au cahier des charges. On peut cependant s'attendre à ce que le coût politique d'une telle privatisation la rende inenvisageable à moyen terme. On peut alors envisager des solutions alternatives, telles que la mise en place de mécanismes d'intéressement des décideurs hospitaliers aux résultats comptables de leur établissement dans le public. De tels mécanismes augmenteraient les incitations à la réduction des coûts. À court terme, de telles réductions accroîtraient les revenus des dirigeants de l'hôpital ; mais, à moyen terme, elles se traduiraient par une baisse générale des coûts et donc des tarifs, et donc des économies pour le contribuable.

#### **4. Perspectives de politique économique**

Il y a plus de dix ans, le CAE avait produit un rapport sur la régulation du système de santé, rédigé par Michel Mougeot (1999). L'auteur y constatait la piètre performance de la France, si l'on met en regard la part du PIB consacrée aux dépenses et son rang dans divers indicateurs de niveau de santé. Qu'en est-il, dix ans après ?

Malgré les nombreuses réformes et ajustements, la part des dépenses de santé dans le PIB n'a cessé de progresser et la France se trouve désormais au deuxième rang mondial pour cet indicateur, avec 11,7 %, tandis qu'elle n'occupait que la quatrième place en 1996. Ceci est d'autant moins encourageant que la pyramide des âges en France est globalement plus favorable que pour nos voisins immédiats dans ce classement (Suisse, Allemagne, Belgique, Canada, Autriche et Pays-Bas). Cette observation doit cependant être nuancée : la part des plus de 80 ans est en effet particulièrement élevée en France, ce qui contribue au niveau élevé des dépenses<sup>(41)</sup>. Le dérapage n'en est pas moins réel, et il existe un risque important à moyen terme d'un ajustement brutal dans l'urgence sous la pression de la crise budgétaire. Un tel ajustement serait bien plus coûteux pour les usagers du système de santé que la mise en place de mécanismes raisonnables de régulation même si ces mécanismes réduisent quelque peu le niveau de protection sociale dont bénéficie la population.

(41) Toutefois, comme le montre Dormont (2009), le vieillissement n'est pas nécessairement à l'origine de la plus grande part de l'augmentation des dépenses de santé. Selon elle, l'augmentation des dépenses de santé à tout âge, en raison de soins plus performants, mais également plus coûteux, explique une part conséquente de l'accroissement des dépenses de santé totales.



#### 4.1. La difficile conciliation d'objectifs contradictoires

Bien que des réformes comme la T2A constituent un progrès indéniable vers l'amélioration des incitations et de l'efficacité productive, un problème fondamental subsiste, à savoir que les décideurs ne sont pas les payeurs. Cela est évident au niveau décentralisé, puisque les décisions sont prises par les patients et les médecins. Mais cela est également vrai à un échelon plus global, puisque ni la CNAM, ni les dirigeants d'hôpitaux, ni bien sûr les hommes politiques ne profitent des économies éventuelles qu'ils parviendraient à mettre en œuvre (inversement, ils ne sont pas pénalisés pour les dérapages dans les dépenses). La chose est moins vraie pour les mutuelles dans la mesure où elles sont en concurrence et doivent équilibrer leur bilan.

Dans ces conditions, les patients font face à un prix « nul » et soit on accepte que la quantité totale de soins soit celle qui correspond à ce prix, soit on met en place des mécanismes quantitatifs de rationnement.

Dans tous les cas, la question du « niveau » optimal des soins doit être posée. À une extrémité, on peut considérer la santé comme un bien « primal » et donc envisager la gratuité totale des soins<sup>(42)</sup>. Cette posture éthique pose que le niveau « optimal » est celui qui correspond à un accès gratuit et illimité de toute la population. Elle a le mérite d'évacuer tout problème d'incitation puisqu'il n'est pas concevable de « trop » consommer de soins. À l'autre extrême, on peut laisser le coût des soins à la charge des patients, ce qui leur permet de déterminer eux-mêmes leur niveau de soin optimal, tout en les laissant s'assurer librement. Cette logique est implicitement rejetée par le système actuel qui considère qu'elle est en contradiction avec l'égalité d'accès aux soins. Par ailleurs, comme nous l'avons vu, elle est génératrice d'inefficacités puisqu'il n'est pas possible de s'assurer contre les conditions préexistantes.

Le système français se situe entre ces deux extrêmes ce qui rend nécessaire de formuler une cible quantitative. Il semble devoir concilier des objectifs contradictoires. Le ticket modérateur et les dépassements d'honoraires semblent remettre en question la vision de la santé comme bien « primal ». Mais par ailleurs une couverture complète est offerte aux bénéficiaires de la CMU ou aux patients atteints d'affections de longue durée. Là-dessus se greffent des mécanismes de prise en compte d'objectifs macroéconomiques d'encadrement des dépenses publiques. Ceux-ci devraient impliquer, lorsqu'ils sont contraignants, soit une hausse des tarifs acquittés par les

(42) Il est plus facile de traduire la notion de « bien primal » en termes de politique économique (sacralisation des dépenses de santé) que de lui donner des fondements rigoureux. En effet, si l'on suppose que les consommateurs accordent une priorité absolue à la consommation d'une quantité  $q$  minimale de santé, en deçà de laquelle ils ne consommeraient que du bien santé, alors toute personne dont le revenu est supérieur ou égal à la valeur de  $q$  acquerra ce niveau minimal, et la politique de santé devrait se ramener à s'assurer, au moyen de politiques redistributives, que le revenu de l'ensemble de la population est supérieur à ce niveau. Ceci est bien différent des arguments selon lesquels la santé ne se vend pas, etc.

patients, soit du rationnement des soins<sup>(43)</sup>, ce qui dans les deux cas remet *de facto* en cause le principe de la gratuité.

Les évolutions récentes témoignent d'une volonté de mettre en place de meilleurs mécanismes incitatifs, sans pour autant transférer les coûts des décisions aux décideurs eux-mêmes. Ainsi, si la T2A est un facteur de transparence comptable, il n'est pas évident que les perspectives de profits et de pertes qu'elle engendre pour les hôpitaux se traduisent par des décisions plus efficaces, car la fonction objectif de l'hôpital ne permet pas clairement aux décideurs d'internaliser ces profits et ces pertes (voir section précédente)<sup>(44)</sup>. De même, comme on l'a relevé section 2, le patient qui arbitre entre deux établissements ne prend pas en compte le coût total de chaque possibilité<sup>(45)</sup>. Il est cependant exact que sous la pression des déficits, on a assisté à une hausse progressive des franchises et des tickets modérateurs, et une poursuite de ces hausses est le scénario principal pour les ajustements à venir (voir le complément B pour une présentation des effets de telles mesures au moyen d'un modèle de microsimulation). Il n'en reste pas moins que, jusqu'ici, ces restes à charge sont couverts par les complémentaires, d'où leur effet limité sur les incitations individuelles et la progression des dépenses de santé.

## 4.2. L'arbitrage entre incitation et égalité dans l'accès aux soins

Ces contradictions résultent du conflit entre incitations et objectif d'égalité dans l'accès aux soins, dans la mesure où toute prise en charge par les assurés du coût de leurs soins remet en question le second objectif. Au regard de la hausse continue des dépenses de santé, on peut penser que cette approche est insuffisante et qu'une meilleure internalisation des coûts de la part des décideurs est nécessaire. Si l'on veut espérer résoudre ce conflit, il convient de s'interroger sur la signification de cette notion d'égalité.

Une première interprétation est qu'il n'est pas juste que les patients atteints de mauvaises conditions préexistantes paient plus cher pour leur assurance. Cette interprétation fait intervenir la notion d'inégalité entre états de santé initiaux. Si cette notion se manipule aisément en théorie, elle pose le problème de la frontière, floue, entre ce qui est à proprement parler hérité à la naissance, et ce qui est acquis par la suite. En particulier, si des prédispositions à être victime de telle ou telle affection sont transmises génétiquement,

---

(43) Ce qui est le cas avec le système actuel de réajustement des tarifs T2A en fonction de l'ONDAM : une baisse des tarifs pour raisons d'équilibrage des comptes se traduit par une réduction de l'offre hospitalière et par un plus grand rationnement dans la mesure où celle-ci est inférieure à la demande.

(44) C'est sans doute ce qui explique les conclusions relativement pessimistes du rapport de Cash *et al.* (2011).

(45) Néanmoins, la part des dépassements non remboursés par la complémentaire incite le patient à arbitrer, mais comme on l'a mentionné cela conduit à un choix qui n'est pas nécessairement le bon du point de vue du bien-être social ni même d'un simple point de vue budgétaire.

il est délicat de distinguer ce qui ressort du comportement à risque du malade atteint de ladite affection, de ce qui est dû aux conditions préexistantes, dont il n'est pas responsable. On peut alternativement considérer l'inégalité des états de santé initiaux comme un problème d'efficacité dû au fait que les marchés sont incomplets et ne permettent pas de s'assurer contre de telles conditions préexistantes. Ce problème peut être résolu, comme cela a été souvent proposé, dans un marché concurrentiel réglementé où il est interdit aux assureurs de pratiquer des tarifs discriminants selon certaines caractéristiques observables de leurs clients. On peut également envisager un marché de l'assurance-maladie purement concurrentiel, assorti du remboursement par l'État des surcoûts associés aux conditions préexistantes (par définition, de telles conditions ne résultent pas du choix des individus et le versement d'une telle indemnité n'implique aucun phénomène d'aléa moral). Cette interprétation ne plaide donc pas en faveur de la gratuité.

Une seconde interprétation de la notion d'égalité est qu'il n'est pas juste d'exiger la même prime d'assurance des assurés les plus pauvres car il est plus coûteux pour eux, en termes de bien-être, de s'acquitter de cette somme. C'est cette interprétation qui sous-tend l'existence de la CMU. La CMU complémentaire, elle, implique que la mise en place de mécanismes incitatifs ne s'appliquerait qu'au-delà d'un certain seuil de revenu, tandis que ceux qui se situeraient en deçà auraient un accès « illimité » à la santé. Une telle situation a peu de chances d'être efficace même du point de vue d'un « planificateur social » parfaitement égalitaire ; il n'y a pas de raison qu'il soit optimal que la consommation de biens de santé décroisse avec le revenu pour certaines catégories de ménages. Par ailleurs elle évacue le problème des inégalités d'accès entre les ménages non exonérés, question qui ne peut être que de plus en plus saillante au fur et à mesure que l'on augmente les tickets modérateurs. Il est naturel de concilier redistribution et incitation en mettant en place des franchises ou tickets modérateurs qui dépendraient eux-mêmes du revenu (nous ne faisons là que reprendre des propositions d'autres économistes, par exemple Geoffard, 2006). Les systèmes de franchise/plafond ont été maintes fois proposés, notamment à l'étranger, voir le complément B. La quantité de redistribution qui correspond aux « préférences sociales » serait ciblée en calibrant la dépendance de la franchise par rapport au revenu. Le rapport de Briet et Fragonard (2007) sur le bouclier sanitaire retient d'ailleurs, parmi les scénarios testés du bouclier (modulant le ticket modérateur et un plafond de reste à charge annuel), une variante où le plafond dépend du revenu de l'agent. Les surcoûts afférents aux maladies graves seraient donc traités grâce à un plafond qui pourrait aussi dépendre du revenu. D'une manière générale, de nombreux systèmes peuvent être envisagés ; par exemple la franchise pourrait être définie en termes annuels : aucun remboursement n'interviendrait avant que cette franchise soit épuisée. Les patients internaliseraient donc 100 % des coûts tant qu'ils resteraient en deçà de la franchise. Un tel système a cependant peu de chances d'être optimal. En principe un ticket modérateur inférieur à 100 % est optimal y compris pour les petits risques, encore que

le niveau de ce ticket est difficile à évaluer, puisqu'il dépend entre autres du degré d'aversion pour le risque et de l'acuité du problème d'aléa moral.

On notera cependant que de telles mesures ne peuvent être évaluées indépendamment de l'ensemble du système de prélèvements, puisqu'elles augmentent mécaniquement les taux marginaux d'impositions<sup>(46)</sup>, qui sont déjà très élevés si l'on prend en compte toutes les formes de taxation.

Par ailleurs, la simplification du système de remboursement que permettrait un tel système rendrait le système plus lisible pour les assurés qui connaîtraient notamment le risque maximum de reste à charge qu'ils courent. Elle permettrait, ensuite, la tenue de débats plus transparents sur la participation financière des patients. En effet, face aux besoins de financement de la sécurité sociale, les débats sur le niveau de la franchise et du plafond seraient plus concrets que ceux sur le niveau du ticket modérateur. Lorsqu'une modification du ticket modérateur est envisagée l'assuré ne peut percevoir clairement l'effort supplémentaire qu'il supportera car celui-ci sera proportionnel à sa dépense. (cf. complément B)<sup>(47)</sup>.

On pourrait également envisager une variante dynamique du système franchise/plafond fondée sur l'existence d'un « compte santé »<sup>(48)</sup>. L'intérêt d'une telle approche est de mobiliser la capacité d'auto-assurance des individus à travers l'épargne et l'emprunt<sup>(49)</sup>. Un tel compte santé serait débité d'un certain nombre de « points » (et la valeur d'un point pourrait dépendre plus ou moins du revenu selon les considérations redistributives que l'on voudrait incorporer au système) proportionnel au coût des soins consommés par l'assuré.

Au-delà d'un certain seuil, un solde créditeur serait convertible en euros et utilisable librement. Un solde débiteur ne permettrait pas une telle opération,

---

(46) En effet, la partie restant à charge pour le patient peut être interprétée comme un impôt ; plus celle-ci croît avec le revenu, plus le taux de prélèvement total sur ce dernier est élevé. La CMU-C constitue en quelque sorte une forme extrême de plafond associé à une franchise nulle en deçà d'un certain seuil de revenu. C'est d'ailleurs pour réduire les taux marginaux d'imposition très élevés associés à de tels seuils qu'une aide complémentaire partielle a été mise en place pour les ménages immédiatement au-dessus du seuil d'éligibilité à la CMU complémentaire. Une franchise plafonnée en fonction du revenu ne crée pas de taux d'imposition si élevés car elle n'implique pas d'effets de seuil, mais il n'en reste pas moins que le plafond exerce un effet d'autant plus distorsif sur l'offre de travail qu'il est indexé sur le revenu. Il n'y a donc pas de « recette miracle » qui permettrait d'accroître simultanément redistribution et incitation.

(47) Je remercie Michel Houdebine de la Direction du Trésor pour avoir attiré mon attention sur ces points.

(48) Ce type de système a été proposé pour l'assurance-chômage, voir, par exemple, Orszag et Snower (1997).

(49) Hassler et Rodriguez Mora (1999) ont montré, dans le cas de l'assurance chômage, que l'épargne personnelle a des propriétés d'assurance presque aussi bonnes que l'assurance proprement dite, tandis que dans le premier cas les agents internalisent bien entendu correctement les effets de leur effort.

mais n'aurait pas à être réglé et jouerait donc ainsi le même rôle qu'un plafond sur le reste à charge total de l'assuré. L'intérêt d'un tel système est que les assurés sont incités, pour chaque pathologie, à choisir la thérapie la moins coûteuse (à résultats donnés)<sup>(50)</sup>. Même s'ils sont dans le rouge, ils ont toujours intérêt à améliorer leur solde car cela augmente la probabilité de retour à un solde créditeur dans le long terme (tandis qu'avec un plafond annuel, il n'y a pas d'incitations à réduire ses dépenses à la marge lorsque ce plafond est crevé). Ils sont également incités à investir dans la prévention puisqu'elle leur permettra un solde créditeur à l'avenir<sup>(51)</sup>. On pourrait d'ailleurs envisager d'appliquer ce système à d'autres champs de la protection sociale et de mettre en place un compte social unique pour tous les citoyens. Notons cependant qu'un tel système devrait être calibré avec précision pour éviter que la multiplication des soldes créditeurs ne s'avère trop coûteuse en termes de dépenses publiques. Ces méthodes permettraient sans doute de réaliser des économies importantes. Ainsi, le dispositif « tiers payant contre générique », qui impose un coût minimal aux patients qui refuseraient de choisir un générique, semblerait avoir eu des effets considérables. On est donc en droit d'attendre des réductions importantes de dépenses d'un transfert des coûts vers les assurés.

### 4.3. De meilleures incitations à l'endroit des professionnels de santé

Cependant, il est bien connu que les décisions en matière de santé n'appartiennent pas seulement au patient, mais aussi à son médecin qui est, de par ses connaissances, mieux informé. Il y a d'ailleurs un débat sur la marge de manœuvre que l'on doit laisser au patient (voir Le Pen, 2009). On peut néanmoins penser qu'une meilleure information ne peut pas lui nuire et que la hausse du niveau général d'éducation et les nouvelles technologies permettent d'envisager une plus grande autonomie des malades (cf. Gimbert et Lemoine, 2010). Bien entendu, ceci ne peut qu'améliorer les propriétés incitatives des franchises et tickets modérateurs. Il est néanmoins également souhaitable de renforcer les incitations au niveau des praticiens et des gestionnaires d'hôpitaux. Le système Ameli est un premier pas important puisqu'il permet une mise en concurrence plus grande des établissements et des praticiens au niveau des tarifs de service. Cependant

(50) On peut citer la République tchèque qui applique un système où les assurés qui n'ont pas demandé de remboursements sur une période récente bénéficient de récompenses non financières (cf. Paris et al., 2010).

(51) On peut objecter à une telle proposition que certains agents sont « myopes » et qu'une fois dans le rouge ils ne prendront aucune mesure visant à limiter leur recours aux soins. Si cette myopie frappe la majorité de la population, alors un système plus simple de franchise/plafond avec apurement annuel est sans doute plus efficace, encore que pour n'importe quel système d'incitations on peut envisager l'existence d'agents aux limitations cognitives ou comportementales suffisamment fortes pour éliminer les vertus de ce système. On peut également envisager de laisser le choix aux assurés entre l'apurement annuel ou le report du solde d'une année sur l'autre. Quoi qu'il en soit, le mérite du compte santé préconisé ici est de pouvoir reverser aux contribuables une partie des économies que leur comportement « vertueux » fait faire à la CNAM. Quelle que soit la méthode utilisée, il serait sain d'envisager la mise en place de mécanismes s'inspirant de cette philosophie.

un tel système ne peut fonctionner *in fine* que si les patients internalisent une part suffisante des coûts, sans quoi ils choisiraient tous les praticiens de la meilleure qualité possible, générant ainsi des problèmes de congestion. En ce qui concerne les hôpitaux, la T2A restera peu efficace si les décideurs n'internalisent pas ses conséquences comptables (voir discussion de la section 3). Une mesure extrême serait la privatisation de certains hôpitaux ou de certains de leurs services. Une telle privatisation n'écarterait pas les missions de service public ni les préoccupations d'équité dans l'accès aux soins, dans la mesure où elle resterait soumise à un cahier des charges. Elle inciterait les gestionnaires à rationaliser leur production en fonction des signaux économiques inhérents aux tarifs hospitaliers. Dans la mesure où elle rapprocherait le revenu du médecin tiré de ses activités hospitalières publiques des honoraires qu'il pourrait percevoir dans le privé pour des actes équivalents, elle impliquerait également une réduction des distorsions dans les choix d'établissement. Précisons également que la privatisation n'est pas forcément synonyme d'activité lucrative, bien au contraire. Dans nombre de pays, au premier rang desquels les États-Unis, le privé non lucratif domine nettement le privé lucratif, aussi bien par le volume de soins produits que par la qualité des soins (cf. Devereaux et al., 2004, et Silber et al., 2005). Il est à cet égard intéressant de remarquer que la France est l'un des rares pays à disposer d'une offre de soins privée lucrative si développée (cf. tableau 6). Sans aller jusqu'à la privatisation, on peut également envisager la mise en place de mécanismes d'intéressement qui lieraient la rémunération des gestionnaires d'hôpitaux à la bonne santé budgétaire de leurs établissements (voir section 3).

## 6. Répartition des soins hospitaliers aigus dans plusieurs pays de l'OCDE en 2009

	France	Allemagne	Royaume-Uni	Italie	Japon	Suède	Canada
Proportion des lits en soins hospitaliers aigus							
• à l'hôpital public	66	49	96	81,5	26,3	98	100
• à l'hôpital privé non lucratif	9	36	4	16,7	73,7	0	0
• à l'hôpital privé lucratif	25	15	0	1,8	0	2	0

Source : Paris, Devaux et Wei (2010).

Concernant la médecine de ville, les économistes ont longtemps déploré que le système actuel comprenne peu d'incitations. Si la réglementation des tarifs est censée prévenir un dérapage des prix, elle semble avoir l'effet

pervers de favoriser une multiplication des actes, à titre compensatoire. D'un point de vue strictement allocatif, cette inefficacité est plus importante que celle qu'on observerait avec des prix excessifs. En effet (en admettant pour simplifier l'analyse que les patients se conforment aux décisions de leurs médecins), dans le premier cas, la quantité d'actes prescrits est relativement efficace et le prix « excessif » est un transfert non distorsif entre les assurés et le corps médical. Dans le second cas, les médecins prescrivent trop d'actes et font trop de consultations pour chaque pathologie. Il en résulte soit un nombre excessif de médecins, soit du rationnement dans l'accès à ceux-ci (et les deux phénomènes peuvent coexister selon les spécialités considérées). Il est naturel d'envisager une extension du système T2A à la médecine de ville : le montant remboursé par la sécurité sociale et les assurances complémentaires serait alors fonction de la pathologie et non du nombre d'actes. Ce système est cependant une forme particulière de plafond de remboursement qui s'applique à nouveau au patient et non au médecin. On voit bien la difficulté associée à la mise en œuvre d'incitations pour la médecine de ville : tandis qu'un gestionnaire d'hôpital, face à une tarification à l'activité, a intérêt à améliorer l'efficacité de sa gestion grâce à des pratiques organisationnelles appropriées, on peut penser qu'un médecin libéral ne peut générer des économies qu'en réduisant sa propre activité. On notera cependant que de nombreuses personnes auditionnées dans le cadre de ce rapport ont fait état d'une évolution du comportement des jeunes médecins et de leur acceptation croissante du salariat et du regroupement en établissements de taille suffisante. La mise en place d'un système type T2A appliqué à la médecine de ville pourrait catalyser de telles évolutions, ainsi qu'une utilisation plus grande des nouvelles technologies de l'information et de la communication (consultations à distance, etc.) (voir à nouveau Gimbert et Lemoine, 2010 et Reynaudi et Sauneron, 2011). Comme pour la T2A hospitalière, la difficulté réside dans la tension entre le principe d'une tarification uniforme par pathologie et la prise en compte des patients particulièrement difficiles. Dans ce dernier cas, cependant, les dépassements de coûts inhérents aux caractéristiques du patient peuvent être acceptables pour ces derniers s'ils s'inscrivent dans un système franchise/plafond ou compte santé comme décrit plus haut.

Une autre piste de responsabilisation des praticiens pourrait reposer sur un système d'affiliation des médecins à des organismes d'assurance, qui seraient eux-mêmes en concurrence sur le plan des tarifs offerts aux assurés. Ces derniers auraient alors un intérêt financier à mettre en place un système de contrôle des coûts engendrés par l'activité de leurs médecins affiliés.

## Références bibliographiques

- Acemoglu D. et J. Linn (2004) : « Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry », *Quarterly Journal of Economics*, vol. 119, n° 3, pp. 1049-1090.
- Acemoglu D., A. Finkelstein et M.J. Notowidigdo (2009) : « Income and Health Spending: Evidence from Oil Price Shocks », *NBER Working Paper*, n° 14744.
- Adams P., M.D. Hurd, D. McFadden, A. Merrill et T. Ribeiro (2003) : « Healthy, Wealthy, and Wise? Tests for Direct Causal Paths Between Health and Socioeconomic Status », *Journal of Econometrics*, vol. 112, n° 1, pp. 3-56.
- Barlet M. et C. Collin (2010) : « Localisation des professionnels de santé libéraux » in *Comptes nationaux de la santé 2009*, Anne-Marie Brocas (éd.), DREES.
- Boyd J.H. et E.C. Prescott (1986) : « Financial Intermediaries Coalitions », *Journal of Economic Theory*, vol. 38, n° 2, pp. 211-232, avril.
- Briet R. et B. Fragonard (2007) : *Mission bouclier sanitaire*, La Documentation française.
- Busse R., A. Geissler, W. Quentin et M. Wiley (eds) (2011) : *Diagnosis Related Groups in Europe. Moving Towards Transparency, Efficiency and Quality in Hospitals*, Open University Press, Maidenhead.
- Caisse Nationale d'Assurance-Maladie (CNAM) (2011) : *Propositions de l'assurance-maladie sur les charges et produits pour l'année 2012*, Rapport pour le Conseil CNAMTS du 7 juillet.
- Cash E., R. Cash et C. Dupilet (2011) : « Étude sur la réactivité des établissements de santé aux incitations tarifaires », *Document de Travail de la DREES*, n° 106, juillet.
- Comité d'Histoire de la Sécurité Sociale (CHSS) (2006) : *L'esprit de réforme dans la Sécurité sociale à travers son histoire*, Actes du colloque du mercredi 16 novembre 2005, La Documentation Française.
- Cutler D.M. et M. McClellan (2001) : « Is Technological Change in Medicine Worth It? », *Health Affairs*, vol. 20, n° 5, septembre/octobre.
- Devereaux P.J., D. Heels-Ansdell, C. Lacchetti, T. Haines, K. EA Burns, D.J. Cook, N. Ravindran et al. (2004) : « Payments for Care at Private For-Profit and Private Not-For-Profit Hospitals: A Systematic Review and Meta-Analysis », *Canadian Medical Association Journal*, vol. 170, n° 12, pp. 1817-1824.



- Dormont B. (2009) : « Les dépenses de santé : une augmentation salubre », *Opuscule du CEPREMAP*, n° 15.
- DREES (2009) : Second rapport d'activité du Comité d'évaluation de la T2A (rapporteurs : N. Blanchard, S. Pichetti et D. Raynaud), *Document de Travail*, n° 94, septembre.
- Fédération hospitalière de France (2010) : *Rapport au Parlement sur la tarification à l'activité (T2A)*, Ministère de la Santé et des Sports
- Frijters P., J.P. Haisken-DeNew et M.A. Shields (2005) : « The Causal Effect of Income on Health: Evidence from German Reunification », *Journal of Health Economics*, vol. 24, n° 5, pp. 997-1017.
- Geoffard P-Y. (2006) : *La lancinante réforme de l'assurance-maladie*, Éditions de la rue d'Ulm.
- Gimbert V. et S. Lemoine (2010) : « Médecine de ville : quelles nouvelles pratiques pour quels gains d'efficience ? », *Note d'analyse du CAS*, n° 204,
- Grignon M. (2010) : « Le financement d'un système de santé et le partage obligatoire-volontaire », *Revue Française des Affaires Sociales*, n° 4, pp. 53-73.
- Hassler J. et J.V. Rodriguez Mora (1999) : « Employment Turnover and the Public Allocation of Unemployment Insurance », *Journal of Public Economics*, vol. 73, n° 1, pp. 55-83, juillet.
- Haut-Conseil pour l'Avenir de l'Assurance-Maladie (HCAAM) (2005) : *Notes sur les couvertures complémentaires maladie*, Réunion du jeudi 24 février.
- Henriet D. et J-C. Rochet (1999) : « Régulation et intervention publique dans les systèmes de santé » in *Régulation du système de santé*, Rapport du CAE n° 13, La Documentation française, pp. 115-129.
- Lachaud C. et L. Rochaix (1995) : « Équité dans le financement et la prestation de soins de santé en Europe », *Revue d'Économie Financière*, n° 36, pp. 367-394.
- Laffont J-J. (2000) : *Incentives and Political Economy*, Clarendon Press, Oxford.
- Lange O. (1936) : « On the Economic Theory of Socialism: Part One », *The Review of Economic Studies*, vol. 4, n° 1, pp. 53-71.
- Le Pen C. (2009) : « Patient ou personne malade ? Les nouvelles figures du consommateur de soin », *Revue Économique*, vol. 60, n° 2, pp. 257-274.
- Moison J-C. et M. Pépin (2010) : « Les impacts de la T2A sur les modes d'organisation et de fonctionnement des établissements de santé », *Document de Travail de la DREES*, n° 97, mai.
- Mougeot M. (1999) : *Régulation du système de santé*, Rapport du CAE, n° 13, La Documentation française.

- Nyman J.A. (2011) : « Measurement of QALYs and the Welfare Implications of Survivor Consumption and Leisure Forgone », *Health Economics*, vol. 20, n° 1, pp. 56-67.
- O'Donoghue T. et M. Rabin (2003) : « Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes », *American Economic Review*, vol. 93, n° 2, pp. 186-191.
- Or Z. et T. Renaud (2009) : « Principes et enjeux de la tarification à l'activité à l'hôpital (T2A). Enseignement de la théorie économique et des expériences étrangères », *Document de Travail de l'IRDES*, n° 23, mars.
- Orszag M. et D. Snower (1997) : *From Unemployment Benefits to Unemployment Support Accounts*, Miméo.
- Paris V., M. Devaux et L. Wei (2010) : « Health Systems Institutional Characteristics », *OCDE Health Working Papers*, n° 50.
- Reynaudi, M. et S. Sauneron (2011) : « Quelles opportunités pour l'offre de soins de demain (volet 2) : la télésanté », *Note d'Analyse du CAS*, n° 255.
- Saint-Paul G. (2011) : *The Tyranny of Utility*, Princeton University Press.
- Sambuc C. (2011) : « Transplantation d'organes et comparaisons interpersonnelles. Une discussion critique sur le recours aux QALYs », *Document de Travail du GREQAM*, Université d'Aix-Marseille, n° 8.
- Scarf H. (1960) : « Some Examples of Global Instability of the Competitive Equilibrium », *International Economic Review*, n° 1, pp. 157-172.
- Silber D., L. Castra, P. Garel et A. Risk (2005) : *Hôpital: le modèle invisible*, Institut Montaigne.
- Thaler R. et C. Sunstein (2003) : « Libertarian Paternalism », *American Economic Review*, vol. 93, n° 2, pp. 175-179.

## Commentaire

**Brigitte Dormont**

*Professeuse à l'Université Paris Dauphine*

L'évaluation d'un système de santé est-elle soluble dans la théorie économique ? En avant-propos, l'auteur de ce rapport revendique la fraîcheur de sa perspective. Il est vrai que beaucoup d'acteurs de notre système considèrent qu'intégrer des arbitrages économiques à la politique de santé est contraire à l'éthique. Selon eux, les considérations économiques et la recherche d'une efficacité de la dépense doivent venir en aval des décisions relatives aux objectifs de la politique de santé.

L'économiste que je suis rejoint sans réserve Gilles Saint-Paul sur ce point : l'analyse économique doit justement permettre de clarifier les objectifs du système de soins étudié, et d'examiner si son organisation permet de les atteindre.

L'exercice effectué par Gilles Saint-Paul est une sorte de visite d'un candide formaté à la théorie économique dans l'édifice alambiqué du système de santé français. Le résultat est une réflexion continue, très intéressante mais assez peu structurée, au fil de la visite. La logique est poussée, ce qui conduit à des passages fascinants et un peu vertigineux, qui ne pourront peut-être pas améliorer la réputation des économistes. Ainsi cette note dans les premières pages du rapport, où l'objection des maladies ou handicaps préexistants que l'on peut opposer à un projet de libre marché de l'assurance-maladie est levée avec l'évocation d'une possibilité d'assurance volontairement contractée par les parents avant la naissance de leur enfant. D'autres réflexions sont éclairantes, qui mettent à jour l'implicite de la société française, où l'on n'accepte pas de différences géographiques de tarifs pour les soins, alors que la tolérance pour des inégalités spatiales de prix des logements est totale.

Quelques vérités importantes sont dites : qu'il faut séparer les champs d'intervention de la sécurité sociale (CNAM) et des assurances complé-

mentaires ; que la tarification à l'activité (T2A) mise en place en 2004 pour réformer le financement des hôpitaux est un instrument qui ne peut servir trois objectifs ; qu'elle ne peut donc à la fois permettre d'améliorer l'efficacité productive et l'efficacité allocative et de respecter la contrainte budgétaire définie par l'objectif national de dépenses de l'assurance-maladie (ONDAM) qui est une enveloppe fermée ; que le respect de l'ONDAM induit l'existence d'un point flottant pour la définition des tarifs hospitaliers, ce qui ruine l'impact vertueux des mécanismes mis en place ; qu'il peut y avoir une hétérogénéité dans la qualité des soins délivrés mais que cette idée n'est pas intégrée par les instances régulatrices.

Mais on est loin d'avoir ici un examen de notre système de soins avec un énoncé de ses objectifs, un état des lieux de son fonctionnement et des propositions d'amélioration. Il n'y a pas non plus d'évaluation des changements intervenus depuis le diagnostic dressé en 1999 par le rapport de Michel Mougeot<sup>(1)</sup>, avec des précisions sur les blocages qui demeurent et les nouveaux problèmes identifiés.

Le rapport de Gilles Saint-Paul aborde dans trois sections l'assurance-maladie, la tarification hospitalière et la responsabilisation des acteurs. Sur la question de l'hôpital, il faudrait souligner que le principe de la tarification par pathologie, qui crée des incitations à l'efficacité productive, est fortement dénaturé dans la version française de la tarification à l'activité, où la finesse et le contenu de la nomenclature utilisée font dériver la tarification vers un système de paiement à l'acte, avec les risques de demande induite associés. Par ailleurs, la question de la convergence tarifaire entre les secteurs public et privé, abandonné récemment dans le cadre du PLFSS 2013, devrait être discutée. Dans le dernier chapitre, Gilles Saint-Paul souligne avec justesse le conflit existant entre incitations et objectif d'égalité dans l'accès aux soins. Sur la responsabilisation des acteurs, il faudrait cependant discuter de l'importance respective, dans l'impact sur les dépenses, de l'aléa moral du côté de l'offre et de l'aléa moral du côté de la demande.

Examinons maintenant la question de l'assurance-maladie. La question centrale ne porte pas sur le statut public ou privé des intervenants, mais sur la régulation de l'assurance : doit-elle être, ou non, universelle et obligatoire, avec un seul assureur comme en France pour l'assurance de base, ou une pluralité de compagnies comme en Suisse, Pays-Bas ou Allemagne ? Dans le cas où il y a plusieurs assureurs, la définition des règles de la concurrence est cruciale : doit-elle porter sur l'étendue de la couverture, ou seulement sur le niveau des primes pour un contrat standard défini par le régulateur ? La sélection des risques et l'ajustement des primes aux risques doivent-ils être autorisés ou non ? L'assurance de base doit-elle être obligatoire ?

---

(1) Mougeot M. (1999) : *Régulation du système de santé*, Rapport du CAE, n° 13, La Documentation française.

Dans les pays où interviennent plusieurs assureurs, le modèle dit de la « concurrence régulée » s'est imposé : obligation de contracter une assurance-maladie, interdiction de la sélection, fixation de primes indépendantes des caractéristiques de la personne, standardisation des contrats et compensation des risques<sup>(2)</sup>. Ces principes sont appliqués, avec des variantes nationales, en Suisse, aux Pays-Bas, en Allemagne et en Israël. Ils inspirent la récente réforme de l'assurance-maladie aux États-Unis (*Affordable Care Act*, mars 2010).

Selon Gilles Saint-Paul, l'obligation de contracter une assurance-maladie résulte d'un raisonnement paternaliste. Oui et non. Oui, dans la mesure où l'on considère généralement que les individus sous-estiment le risque auquel ils sont exposés. Non, si l'on considère que la société cherche à préserver sa cohésion en évitant de se trouver dans la situation où elle devrait refuser des soins à une personne non assurée pour tirer toutes les conséquences de sa liberté d'avoir choisi de ne pas s'assurer. Cet argument vaudrait, même si les individus étaient pleinement clairvoyants sur leur degré d'exposition au risque.

L'analyse de la couverture offerte par l'assurance-maladie doit tenir compte de la concentration des dépenses, un élément souvent négligé alors qu'il est déterminant. Les dépenses de soins se caractérisent en effet par une distribution extrêmement concentrée : chaque année, 50 % des dépenses couvertes par la sécurité sociale sont causées par les soins prodigués à une minorité de patients, qui représentent seulement 5 % des assurés. Cette caractéristique est observée dans tous les pays développés. Elle ne découle pas d'un dysfonctionnement de l'assurance-maladie, mais simplement de la valeur des traitements : leur coût peut dans certains cas atteindre des valeurs extrêmes qui dépassent les capacités financières des ménages les plus fortunés. Tout individu encourt un risque non nul d'être exposé à un niveau de dépenses qu'il ne peut assumer. Cette caractéristique de la distribution des dépenses de soins implique qu'une société qui veut garantir à tous ses membres un accès aux soins de base doit rendre obligatoire l'assurance-maladie. Cette obligation permet d'instaurer un partage des risques à l'échelle de la population, réduisant le coût de l'assurance pour chacun.

Outre cet argument de mutualisation des risques, il existe une autre raison, déjà mentionnée ci-dessus, de priver les citoyens de la liberté de ne pas s'assurer : dans une société comme la nôtre, l'État ne peut pas s'engager à ne pas soigner les personnes non assurées. Pour des raisons éthiques évidentes, les individus sont *de toute façon* pris en charge aux frais de la collectivité. Il est logique de les contraindre à contribuer *ex ante* au pot commun.

---

(2) Pour décourager la sélection des risques qui passerait par une stratégie de ciblage de publics particuliers, ces systèmes de concurrence régulée comprennent la création d'une caisse de compensation, qui vise à transférer *ex post* des sommes perçues par les assureurs disposant d'une forte proportion d'assurés jeunes et bien portants vers ceux dont le portefeuille comporte plus de personnes à risque.

Une autre caractéristique de notre système est que l'assurance-maladie obligatoire n'offre qu'une couverture partielle des soins. Ses concepteurs ont prévu dès l'origine une couverture incomplète des dépenses de soins. Le ticket modérateur visait à limiter la dépense en faisant directement payer une partie du coût des soins par le patient. Cependant, il a été pris en charge progressivement par les organismes d'assurances complémentaires, conduisant à un copaiement de fait : la sécurité sociale et les assurances complémentaires se partagent maintenant la couverture des soins. L'intervention des assurances complémentaires est loin d'être négligeable, puisqu'elles couvrent 13,7 % de la totalité des dépenses de soins en 2011, et 22,3 % des soins ambulatoires. En France, les organismes complémentaires cumulent les fonctions d'assurances complémentaire et supplémentaire. Complémentaire car ils participent à la couverture des soins de base, soins fondamentaux auxquels tous les membres de la société doivent avoir accès. Supplémentaire car ils couvrent aussi d'autres soins comme le confort à l'hôpital, la médecine alternative et les dépassements d'honoraires. Le fait que des assurances facultatives interviennent de façon importante pour compléter la couverture des soins de base est une originalité française. Dans les autres pays européens, les assurances facultatives sont cantonnées à la fonction d'assurance supplémentaire, c'est-à-dire qu'elles couvrent des biens non essentiels pour la santé.

Sur ce point, nous rejoignons le constat de Gilles Saint-Paul : le fait que les assurances complémentaires interviennent pour compléter la couverture sur les soins de base est à la source de nombreuses inefficacités. Mais il faut aller plus loin pour étudier comment les modalités actuelles de la concurrence dans le secteur des complémentaires nuisent à la fois à l'efficacité et à l'équité de notre système.

Comme on le sait, les organismes d'assurance complémentaire regroupent en France les mutuelles, les sociétés d'assurance et les institutions de prévoyance. Environ deux cinquièmes des détenteurs d'une assurance complémentaire en bénéficient grâce à un contrat collectif d'entreprise. Ces contrats de groupe sont assortis d'une exonération sociale et d'une déductibilité fiscale s'ils sont obligatoires<sup>(3)</sup>. Pour les autres assurés, l'accès à la complémentaire est facultatif. Il passe par le paiement d'une prime souvent indépendante du revenu et croissante avec l'âge. Dans ce cas le financement est dégressif : en 2006, la part du revenu consacrée à l'achat d'une complémentaire hors employeur est de 2,9 % pour les ménages les plus aisés mais peut atteindre 8 % pour les plus modestes. Un tel effort pour la seule complémentaire santé peut sembler prohibitif aux ménages modestes et les conduire à renoncer à l'assurance.

Des compagnies privées en concurrence sur un marché d'assurance facultatif ne peuvent offrir qu'un partage des risques limité. Même des mutuelles qui affichent un idéal de solidarité ne peuvent se permettre d'éga-

(3) Pour en bénéficier, ils doivent aussi respecter les règles des contrats « responsables » décrits plus loin.

liser les primes de leurs adhérents. Ce faisant, elles risqueraient de perdre leurs affiliés jeunes, réduisant l'étendue de la mutualisation et renchérissant d'autant la prime moyenne. Dans ce contexte, les compagnies doivent calculer les primes au plus près des dépenses prévues pour chaque individu. La concurrence leur impose notamment de définir des primes plus élevées pour les personnes âgées. Les écarts de prime entre les jeunes et les seniors peuvent ainsi varier du simple au triple.

Les complémentaires peuvent-elles fixer leur tarif en fonction de l'état de santé du souscripteur ? Les dispositifs fiscaux concernant le secteur visent à décourager cette pratique en favorisant les contrats dits « solidaires », sans tarification en fonction de l'état de santé ni questionnaire médical à la souscription<sup>(4)</sup>. Pour respecter cette contrainte tout en ajustant leurs tarifs au plus près des dépenses de santé des individus, les compagnies développent actuellement des stratégies de segmentation des contrats : des produits de plus en plus différenciés sont proposés, contrats modulaires, contrats *low cost*, avec franchise cautionnée, surcomplémentaires et contrats à options. Cette stratégie permet de définir des groupes d'affiliés avec des niveaux de dépenses homogènes, réalisant *de facto* une sélection des risques et une tarification selon l'état de santé. Par exemple, des jeunes en bonne santé seront attirés par les contrats *low cost* : on pourra proposer à ce groupe une prime bien plus attractive que ce qu'on pourrait leur offrir avec une tarification à l'âge à l'intérieur d'un groupe plus large couvert par un contrat plus généreux. Ici, l'avantage concurrentiel dont bénéficie la compagnie qui développe une segmentation des contrats vient du fait qu'elle peut sélectionner les « bons risques » en offrant un contrat avec une faible couverture. En revanche, une mutuelle qui se bornerait à tarifier à l'âge risquerait de perdre ses affiliés jeunes et bien portants : ceux-ci ont intérêt à souscrire des contrats *low cost* ou peuvent carrément renoncer à s'assurer.

En France, le marché de l'assurance complémentaire est ainsi caractérisé par la dégradation progressive de la mutualisation des risques entre malades et bien portants et une esquivance de la pression concurrentielle par les stratégies de sélection des risques développées par les compagnies via la segmentation des contrats. La multiplicité des contrats qui en résulte engendre une complexité de l'offre qui devient difficilement lisible pour le consommateur, érigeant un obstacle supplémentaire à une véritable concurrence.

Dès maintenant, cette situation se traduit concrètement par des niveaux de primes déjà très élevés pour certains individus, et ce pour une couverture qui est loin d'être complète. Les données individuelles actuellement disponibles sur les couvertures offertes par les assurances complémentaires sont très parcellaires. Elles ne permettent pas de connaître la distribution réelle des taux de couverture des dépenses correspondant à la somme du reste à charge et des dépassements, mais on sait que les contrats offrent des couvertures très hétérogènes.

---

(4) Cet avantage fiscal a été réduit en 2010, ce qui donne lieu à de fortes protestations de la part des mutuelles.

La faible régulation des complémentaires fait que la concurrence se joue sur l'étendue de la couverture, ce qui est à la source, non seulement d'inégalités, mais aussi de fortes inefficacités. Alors que des médicaments avaient eu leur taux de remboursement par la Sécurité sociale abaissé à 15 % à cause d'un trop faible service médical rendu, la Mutualité Française avait conseillé à ses adhérents de ne pas rembourser le complément. Mais les assurances privées ont offert ce remboursement et les mutuelles exposées à la concurrence d'autres assureurs ont bien dû suivre le mouvement ! Pareillement, de nombreux contrats de groupe (subventionnés par des exonérations fiscales) offrent une couverture généreuse, voire intégrale, des dépassements d'honoraires, ce qui permet une solvabilisation de la demande pour les médecins pratiquant d'importants dépassements. Un législateur désireux de réduire, voire contenir le développement des dépassements, doit impérativement cesser de les favoriser au travers d'un dispositif d'exonérations fiscales visant des contrats de groupe offrant ce type de couverture.

La distribution très concentrée des dépenses de soins a aussi pour conséquence qu'une couverture partielle par l'assurance-maladie obligatoire expose les individus à des restes à charge considérables, en l'absence de mécanisme de plafonnement. Le Haut-Conseil pour l'avenir de l'assurance-maladie (HCAAM) a ainsi évalué que 1 % des assurés sociaux ont eu un reste à charge moyen approchant les cinq mille euros en 2008<sup>(5)</sup>. Ce résultat montre les limites de la couverture du risque offerte par notre système d'assurance. À cela s'ajoute le fait que les assurés ne bénéficiant pas d'une couverture majorée au titre d'une affection de longue durée (ALD) commencent à voir leur taux de couverture très dégradé, surtout depuis l'introduction ces dernières années de multiples dispositifs visant à augmenter la contribution financière des patients. Pour la même année 2008, le HCAAM calcule que le taux de couverture par la sécurité sociale des soins de ville n'est que de 51,3 % pour les assurés qui ne sont pas en ALD, lesquels représentent 85 % d'assurés. Ce constat montre l'importance du rôle des assurances complémentaires dans la couverture des soins. Il fait craindre un glissement progressif de l'assurance-maladie obligatoire vers une assurance catastrophique, conception contraire aux objectifs affichés de notre système de santé.

Face à ces évolutions, il est impératif de réfléchir à une réforme de l'assurance-maladie. Deux évolutions principales sont envisageables : l'introduction d'une concurrence régulée pour les assurances complémentaires ou la mise en place d'un bouclier sanitaire.

Dans le premier cas, il faut séparer de façon stricte les activités d'assurance complémentaire et supplémentaire et définir une concurrence régulée pour le secteur des complémentaires, avec assurance obligatoire, interdiction de la sélection, primes indépendantes des risques, standardisation des

(5) Le reste à charge moyen calculé pour le centile supérieur est exactement de 4 928 euros, dont la moitié environ était liée aux dépassements et à la liberté tarifaire (cf. HCAAM, Rapport 2011).



contrats et compensation des risques. Ceci impliquerait une architecture complexe pour l'assurance-maladie, avec trois étages correspondant à la sécurité sociale, les assurances complémentaires (régulées) et les assurances supplémentaires (pour les soins situés en dehors du panier de base). Deux simplifications sont possibles : une extension à 100 % de la couverture des soins de base par la sécurité sociale (ce qui rapprocherait les régions françaises du régime hérité de Bismarck encore en place en Alsace-Moselle) ; ou à l'inverse sa réduction à 0 % pour laisser toute la couverture des soins de base à des caisses d'assurance soumises aux règles de la concurrence régulée, garantes d'une solidarité dans le financement<sup>(6)</sup>. Les avantages d'un assureur unique sont la réduction des coûts administratifs et le fait que les questions de sélection des risques sont résolues simplement et sans coût. L'avantage de faire appel à des caisses multiples est qu'elles peuvent développer des stratégies de contractualisations avec les offreurs de soins pour gagner en efficacité dans la délivrance des soins.

Le bouclier sanitaire consiste à plafonner les restes à charge, c'est-à-dire à faire en sorte que la sécurité sociale couvre à 100 % les dépenses de soins de l'individu, une fois que la somme des tickets modérateurs qu'il a payés durant l'année a atteint un certain plafond. Les promoteurs d'une telle réforme proposent d'en financer le surcoût en introduisant une franchise, ce qui revient à limiter pour tous la couverture des premières dépenses. Autrement dit, on modifie le partage du risque, en améliorant beaucoup la couverture des plus malades et en dégradant encore un peu la couverture de tous. Dans un souci d'équité, le plafond et la franchise peuvent être définis de façon proportionnelle aux revenus des ménages<sup>(7)</sup>. Cette hypothèse du bouclier sanitaire montre qu'améliorer l'accès aux soins n'est pas simple. Bien qu'elle vise à modifier la couverture dans un sens plus équitable en termes de protection contre le risque de dépenses catastrophiques, une telle réforme peut en effet décourager une partie des patients en imposant une dépense non remboursée lors du premier contact avec le médecin. Elle aurait aussi des conséquences très importantes sur le marché des complémentaires, lequel pourrait être fortement réduit, sinon disparaître. Le risque étant plafonné par le bouclier, tous les individus de revenus moyens à supérieurs pourraient en effet préférer ne pas s'assurer. Dans cette hypothèse, les acteurs actuels du marché des complémentaires devraient se reporter sur l'activité d'assurance supplémentaire.

On peut être attaché au caractère facultatif de l'assurance pour favoriser l'expression des préférences. Mais il y a un conflit entre la liberté individuelle de s'assurer et l'égalisation des conditions d'accès à l'assurance et aux soins grâce à la mutualisation des risques à l'échelle de la population. L'obligation d'assurance doit concerner les soins du panier de base, pour

(6) L'organisation de l'assurance en Allemagne montre qu'une concurrence régulée entre caisses est compatible avec des cotisations proportionnelles aux salaires.

(7) Voir Geoffard P-Y. et G. de Lagasnerie (2012) : « Réformer le système de remboursement pour les soins de ville : une analyse par microsimulation », *Économie et Statistique*, à paraître.

lesquels la société décide de faire jouer la solidarité. C'est pour les soins du panier de base que les complémentaires doivent être régulées ou le bouclier sanitaire défini.

Ces propositions laissent entière la question de la définition du panier de base et de son évolution avec le progrès technique médical. Quels sont les biens et services de santé auxquels tous doivent avoir accès ? Pour ces biens, une société qui donne la priorité à la solidarité ne peut pas accorder la liberté de ne pas s'assurer. Si les dépenses associées à ce panier sont en croissance, les prélèvements obligatoires (ou les primes obligatoirement acquittées en cas de concurrence régulée avec pluralité de caisses) doivent augmenter suffisamment pour les financer. Certes, augmenter les prélèvements obligatoires pour couvrir les dépenses de santé peut nuire à la dépense privée, limiter des dépenses de protection sociale comme les retraites ou évincer d'autres dépenses publiques, touchant par exemple à l'éducation ou aux transports. Des arbitrages doivent être rendus pour définir la dépense d'assurance-maladie obligatoire (ONDAM) par rapport à d'autres dépenses publiques et pour définir le panier de soins correspondant à cette enveloppe.

Jusqu'à présent une hostilité marquée en France à l'égard de l'utilisation de critères économiques dans la sélection des traitements a conduit à une grande opacité dans les arbitrages rendus. Depuis octobre 2012 toutefois, un décret relatif aux missions de la Haute Autorité de Santé précise les cas dans lesquels une évaluation médico-économique est requise. Cette évolution constitue certainement un progrès dans la formalisation des critères de priorisation. Mais comme le font les Britanniques du NICE, l'agence chargée de définir les soins pris en charge par le National Health Service, l'outil utilisé sera vraisemblablement le Qaly (voir complément C), qui ignore à la fois l'hétérogénéité des préférences individuelles et les conséquences distributives des arbitrages rendus. Des outils comme le revenu équivalent santé développé par Marc Fleurbaey permettent de construire des pondérations qui prennent en compte à la fois les préférences des individus en matière d'arbitrage santé-revenu et les priorités découlant de la fonction de bien-être social (par exemple, la priorité aux défavorisés si cette fonction est rawlsienne)<sup>(8)</sup>.

Dans un système de santé qui repose sur un principe de solidarité, il est impératif d'améliorer les procédures de définition du panier de soins de base. Concernant le progrès médical, en effet, l'avenir est à la fois radieux et orageux. De très nombreuses innovations sont anticipées pour les prochaines décennies, toutes aussi efficaces pour les soins qu'elles seront coûteuses pour la collectivité dès lors qu'elles seraient intégrées au panier de soins remboursés par la Sécurité sociale. Des arbitrages rigoureux devront être rendus en matière de couverture ; ils devront être sans concession sur les critères d'efficacité et d'utilité sociale des procédures ou des produits

(8) Voir Fleurbaey M., S. Luchini, E. Schokkaert et C. Van de Voorde (2012) : « Pondérations équitables dans l'analyse coût-bénéfice des politiques de santé », *Économie et Statistique*, à paraître.

## Commentaire

### **Franck von Lennep**

*Directeur de la recherche, des études, de l'évaluation et des statistiques,  
ministère des Affaires sociales et de la Santé*

L'analyse économique du système de santé sur laquelle repose l'argumentaire du rapport tient en un principe : notre système souffre de l'absence de recours au signal prix, d'une part car les patients acquittent une part insuffisante du coût des soins, ce qui empêche la révélation de leur propension à payer pour les soins, d'autre part car les prix des soins sont fixés par la puissance publique et ne tiennent pas (assez) compte des fonctions d'utilité (des patients) et de production (des professionnels de santé).

Le rapport apporte des solutions plus ou moins tranchées à ces problèmes, que l'on résume ici de manière succincte et donc sans doute un peu caricaturale. Du côté des patients, il faudrait augmenter les restes à charge, idéalement sous la forme de l'introduction d'une franchise et d'un plafond de reste à charge, voire sous la forme d'un compte santé (on regrettera qu'une proposition aussi radicale, inspirée du modèle singapourien, soit abordée aussi elliptiquement). Du côté des professionnels, les propositions sont moins clairement dessinées. A l'hôpital, si l'on devait renoncer à la privatisation des hôpitaux publics (dont l'auteur semble conclure implicitement qu'elle n'est pas portable politiquement), il faudrait introduire des mécanismes d'intéressement liant la rémunération des gestionnaires d'hôpitaux à la santé budgétaire de leurs établissements (on pourra se demander au passage si l'introduction de diverses formes de rémunérations variables des dirigeants dans les grandes entreprises a réellement eu les effets bénéfiques attendus par la théorie économique). En ville, il faudrait aller vers des systèmes de rémunération à la pathologie, sur le modèle de la T2A.

Le fondement même de ce raisonnement mérite questionnement. Une meilleure régulation du système de santé passe-t-elle par un recours accru au signal prix ?

## 1. Augmentation du reste à charge des patients et recours aux soins

Du point de vue du patient, la gratuité empêcherait la détermination de la quantité optimale de soins en conduisant à une demande infinie. La piste évoquée par le rapport consisterait donc à revenir sur l'architecture actuelle base+complémentaire, et à recentrer les assureurs privés sur un panier de soins qu'ils assureraient au premier euro (optique, dentaire). En parallèle, le ticket modérateur, qui ne serait plus remboursé par les complémentaires, resterait donc à la charge des patients, mais serait intégralement pris en charge par l'assurance maladie au-dessus d'un montant fixé chaque année (« bouclier »), montant pouvant varier avec le revenu et avec l'état de santé.

Cet argumentaire appelle plusieurs réflexions.

L'idée avancée plusieurs fois dans le rapport selon laquelle la participation monétaire du patient au financement de ses soins de santé le responsabiliserait ex ante, en l'incitant à adopter des comportements plus favorables à sa santé, est peu crédible. Si on ne constate pas de renoncement à des comportements à risque alors même que les personnes concernées savent que c'est leur vie qui est en jeu, on voit mal pourquoi le fait de devoir payer une petite partie du traitement suffirait à modifier les comportements. En réalité, on distingue généralement en économie de la santé le risque moral ex-ante du risque moral ex-post, et on considère que le risque moral ex-ante ne pose pas vraiment de problème, car on ne peut raisonnablement pas imaginer que le seul fait d'être assuré conduise les individus à adopter des comportements à risque.

Dans la théorie économique, la participation financière des assurés est donc avant tout un moyen de lutter contre le risque moral ex-post (sachant que je suis malade, ou sachant qu'il y a un risque que je le sois, combien vais-je dépenser). L'augmentation des co-paiements réduit certes un peu la consommation de soins de ville selon certaines études. Mais d'une part, l'élasticité est de toute façon assez faible ; d'autre part, le risque est évidemment que la réduction de la consommation de soins fasse porter des risques sur la santé des assurés. Un exemple : en cas d'absence de détection précoce de carcinome ou de mélanome, c'est la vie même du patient qui est en jeu. Il y a donc un arbitrage à faire entre aléa moral et recours aux soins, et, au-delà de cas abusifs relativement limités (par exemple, faire venir le médecin généraliste à domicile quand on pourrait se déplacer), même pour les soins courants l'assuré n'est souvent pas en situation de déterminer seul quelle est l'utilité de l'expertise médicale. Résultat : aucune étude ne prouve que seule la consommation de soins « inutile » est diminuée en cas d'augmentation des co-paiements – tout simplement parce qu'il est extrêmement difficile de répartir ce qui est un juste recours aux soins de ville de ce qui est un recours « inutile ou excessif ». Au fond, augmenter les restes à charge sur les soins de ville pour révéler les préférences, c'est dire aux assurés qu'ils sont des joueurs de poker et qu'ils n'ont qu'à payer pour voir. On rappellera

ici, sans originalité, qu'un moindre recours à des soins courants aujourd'hui peut se traduire par un recours à des soins plus coûteux demain ; ce qui à la fois est néfaste au système de soins dans son ensemble, mais en outre est porteur d'un risque d'accroissement des inégalités sociales de santé.

Le rapport n'est pas clair au sujet du périmètre des soins sur lequel il faudrait accroître les restes à charge des patients. Les options habituellement étudiées (par exemple dans le rapport Fragonard-Briet sur le bouclier sanitaire, cité par l'auteur) n'envisagent d'accroître les co-paiements que sur le champ des soins de ville. L'augmentation des co-paiements sur les soins hospitaliers n'est que très rarement évoquée – c'est pourtant une voie implicitement retenue ici dans la section sur la T2A. Mais l'argument de la réduction de l'aléa moral ne peut pas vraiment s'appliquer aux soins coûteux, qui dépendent surtout des professionnels et qui génèrent des désagréments non assurables pour les patients (personne n'aime passer un séjour à l'hôpital ou souffrir des effets secondaires des traitements médicamenteux lourds). Ainsi, pour une large part, les dépenses de santé découlent d'une expertise médicale et ne sont pas à la discrétion des assurés. L'idée du rapporteur est, peut-être (car elle n'est pas exprimée), de « peser » sur les choix des professionnels en les conduisant à prendre en compte les capacités financières de leurs patients. Si telle est l'idée, elle est suffisamment structurante pour mériter un débat et ne pas rester au stade implicite.

En conclusion, il est clair que l'introduction ou l'augmentation de co-paiements fait porter le risque de réduire davantage l'aléa moral et donc le recours aux soins chez les ménages les plus modestes. Cela dépasse le simple cadre de l'analyse économique et du jeu des préférences individuelles, et cela renvoie à un cadre d'analyse éthique et politique. Là où on voit bien qu'il s'agit aussi de questions éthiques, c'est que bien souvent, ceux qui prônent l'introduction de franchises en excluent les enfants ; même des penseurs très libéraux ont du mal à accepter (ou en tout cas à écrire) que des parents puissent renoncer à soigner leurs enfants en raison du coût des soins.

Pour limiter les dépenses de santé, les mécanismes de régulation de l'offre, très peu évoqués dans le rapport au-delà de l'analyse de la T2A, semblent nettement plus efficaces que les mécanismes de régulation de la demande.

## **2. Les leviers pour agir sur l'offre de soins**

S'agissant des producteurs de soins, le rapport présente une analyse intéressante de la T2A, de ses objectifs et de ses limites ; mais certains points essentiels ne sont pas abordés.

La section sur la T2A met bien en lumière les difficultés posées par un système tarifaire qui tarifie au coût moyen et non au coût marginal, et les

ambiguïtés liées aux différents objectifs assignés à la T2A : tarification des séjours au coût réel (moyen), incitations de santé publique et régulation prix-volume pour respecter l'enveloppe budgétaire de l'ONDAM. On fera ici trois remarques : d'une part, l'introduction d'incitations de santé publique via les tarifs peut en effet sembler peu pertinente, il existe d'autres outils pour cela ; d'autre part, les incitations à la qualité pourraient être développées, d'autant que qualité et efficacité ne s'opposent pas, les études disponibles montrant plutôt que les hôpitaux les plus efficaces sont aussi ceux qui offrent la meilleure qualité des soins ; enfin, la régulation prix-volume pourrait prendre d'autres formes qu'un abaissement généralisé de l'ensemble des tarifs : ainsi, en Allemagne, les tarifs des GHS sont dégressifs au-delà d'un volume d'actes fixé annuellement.

On voudrait ici ouvrir la réflexion vers les questions non abordées dans ce rapport. La question essentielle porte sur l'alternative à la vision « économiste » portée par la théorie de l'équilibre de l'offre et de la demande par l'ajustement des prix : quelles peuvent être les références et les leviers d'action du régulateur du système de santé en l'absence de recours systématique au signal prix ?

Il s'agit là d'un champ de réflexion très vaste. Il faut d'abord partir d'un constat : les dépenses de soins de ville connaissent depuis plusieurs années un ralentissement, expliqué à la fois par le ralentissement des prix et des volumes ; il reste néanmoins des gains d'efficacité possibles, notamment en matière de prescription médicamenteuse (à cet égard, la mise en place récente du paiement à la performance des médecins généralistes constitue une innovation que le rapport aurait pu mentionner) ; d'autre part, l'évolution de la démographie médicale et le renouvellement générationnel débouchent sur des pratiques professionnelles renouvelées, plus collectives. À l'hôpital, les volumes restent relativement dynamiques sur la période récente ; une partie importante de ce dynamisme s'explique par le vieillissement de la population (la hausse du recours à l'hôpital est forte chez les plus de 75 ans), mais les comparaisons internationales laissent penser que des marges existent pour réduire le recours à l'hôpital en accroissant la qualité du système de soins.

Les leviers à la disposition du régulateur sont de plusieurs ordres pour mieux réguler l'offre et aligner stratégiquement les offreurs par le biais d'incitations financières et non financières.

Certes, le levier tarifaire est réel : à l'hôpital, des améliorations pourraient être apportées à la T2A (plus grande transparence sur le calcul des coûts, dégressivité des tarifs à l'allemande, etc.) ; en ville, le paiement à la performance pourrait être étendu (à d'autres critères ou à d'autres professions), et des rémunérations nouvelles doivent être inventées (on y revient plus bas).

Il est en revanche illusoire de croire qu'en augmentant le reste à charge des patients (d'ailleurs, le rapport se garde de prédire ce que serait le ni-

veau de reste à charge optimal) on aboutira au « juste » recours aux soins puisque le patient ne sait pas déterminer ce qu'est un juste recours ; c'est donc l'organisation du système de soins globalement, et les professionnels de santé plus spécifiquement, qui doivent faire en sorte que le recours aux soins s'établisse au niveau optimal en minimisant les recours non pertinents.

Pour cela, des leviers existent. L'élaboration, assise sur la connaissance scientifique (à la fois médicale et médico-économique), du panier de soins et de recommandations de bonne pratique, et la mise en place de la gouvernance et des outils techniques pour assurer le respect de ces recommandations par les professionnels, constitue une alternative pour déterminer le juste recours aux soins. On n'aboutit pas alors à un équilibre déterminé par les préférences individuelles, mais à une situation tenant compte à la fois de l'état de santé de la population et des recommandations de bonne pratique définies pour chaque état de santé. Le système de santé a peu à peu investi ce champ depuis une vingtaine d'années, trop lentement sans doute, trop frileusement parfois, mais il s'agit bien d'une voie fertile. Ce mouvement doit s'accompagner du développement plus ambitieux et plus rapide des indicateurs de qualité, notamment à l'hôpital.

Autre levier, à côté duquel le rapport passe presque complètement, celui de la coordination des soins. Le temps n'est plus, en effet, à considérer de manière séparée la régulation des soins de ville et de l'hôpital. L'objectif du régulateur, aujourd'hui, c'est bien de faciliter la prise en charge complète du patient, en ville, à l'hôpital et, le cas échéant, dans le secteur médico-social<sup>(\*)</sup>. Cela passe par la définition de recommandations de bonnes pratiques (pas seulement au niveau de chaque professionnel, mais pour l'ensemble du parcours du patient), la mise en place d'outils facilitant l'échange d'informations, la définition de nouvelles organisations des professionnels, en ville comme à l'hôpital, et l'introduction de rémunérations nouvelles, plus collectives, davantage liées à la qualité des soins.

C'est de l'amélioration de l'organisation des soins, autour du parcours des patients, que pourra naître *a posteriori* le « juste » niveau de recours aux soins. L'introduction de nouvelles rémunérations permettra d'accompagner ce mouvement en favorisant la coordination et la qualité ; le niveau des tarifs lui-même restera en partie le résultat de choix collectifs plus ou moins contraints, prenant en compte la confrontation entre les contraintes budgétaires et les attentes des professionnels libéraux et hospitaliers.

---

(\*) Il est d'ailleurs révélateur que le rapport ne mentionne pas les Agences régionales de santé, dont la création en 2009 reposait sur le nécessaire rapprochement des trois secteurs.





## Complément A

# Régulation de la demande de soins : une analyse de la littérature

**Grégoire de Lagasnerie**

*Direction générale du Trésor*

L'assurance-maladie, comme tout système d'assurance, est confrontée au dilemme de la recherche d'une couverture la plus large sans induction de comportements opportunistes.

Plus la couverture offerte par l'assureur est généreuse, plus le prix des soins pour les assurés est diminué. Les patients peuvent ainsi adopter des comportements non optimaux susceptibles d'entraîner une surconsommation médicale car ils ne sont plus sensibilisés au coût réel des soins. Cette situation impose alors un arbitrage à l'assureur qui doit trouver un équilibre entre couverture du risque maladie et incitation financière afin de limiter la surconsommation médicale et les soins inutiles.

En contrepoint du risque d'une surconsommation médicale se pose le problème du renoncement aux soins ou de la sous-consommation médicale. Des incitations financières trop fortes peuvent dresser des barrières importantes à l'accès aux soins. Récemment, le renoncement aux soins a fait l'objet d'un intérêt particulier en France comme à l'étranger. Une étude internationale datant de 2010 montre que 39 % des ménages américains qui ont un revenu inférieur à la moyenne nationale ont renoncé à des soins pour raisons financières contre 27 % en Allemagne, 22 % en Australie, 21 % en Norvège, 18 % au Canada et 17 % en France<sup>(1)</sup> (cf. graphique 1).

(1) Cf. 2010 Commonwealth Fund International Health Policy Survey dans onze pays. Le renoncement aux soins est entendu dans cette enquête comme « n'a pas consulté un médecin lorsqu'il était malade », « n'a pas suivi les prescriptions du médecin » et « a abandonné des tests et traitements ».

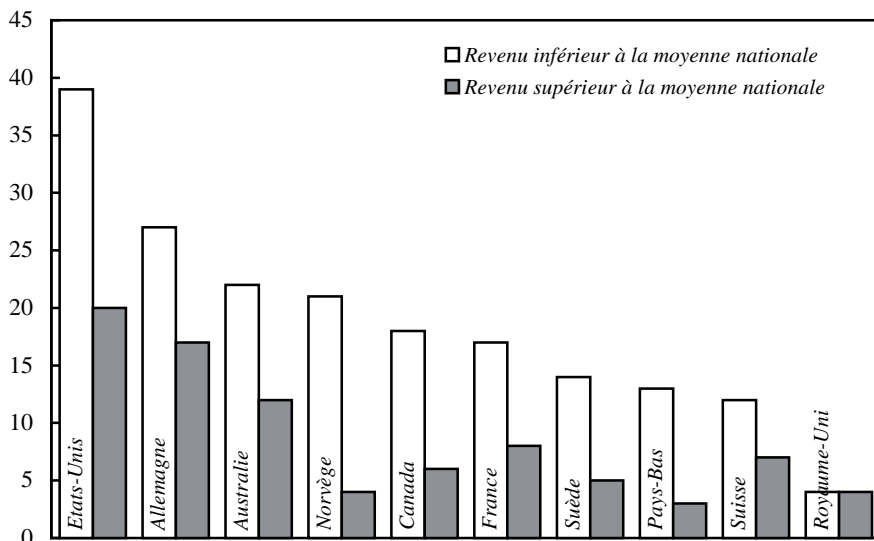
## 1. Mesures modifiant la participation financière des assurés en France depuis 2005

	Barème
2005	1 euro de 13 à 14 euros
<ul style="list-style-type: none"> <li>• Instauration d'une participation forfaitaire</li> <li>• Hausse du forfait journalier</li> </ul>	
2006	de 30 à 40 % 18 euros de 14 à 15 euros
<ul style="list-style-type: none"> <li>• Hausse du ticket modérateur hors parcours de soins</li> <li>• Instauration d'un ticket modérateur sur les actes lourds (K50 ou au-delà de 91 euros)</li> <li>• Hausse du forfait journalier</li> </ul>	
2007	de 40 à 50 %
<ul style="list-style-type: none"> <li>• Hausse du ticket modérateur hors parcours de soins</li> <li>• Extension de la participation forfaitaire d'un euro aux consultations externes des hôpitaux mais plafonnement des participations forfaitaires à 4 euros par jour</li> <li>• Hausse du forfait journalier</li> </ul>	de 15 à 16 euros
2008	0,50 euro par acte/ 2 euros par transport 0,50 euro par boîte de 0 à 65 %
<ul style="list-style-type: none"> <li>• Instauration des franchises médicales plafonnées à 50 euros par an : <ul style="list-style-type: none"> <li>– auxiliaires médicaux et transports</li> <li>– médicaments</li> </ul> </li> <li>• Hausse du ticket modérateur sur les médicaments à vignette bleue des affections de longue durée</li> </ul>	
2009	de 50 à 70 %
<ul style="list-style-type: none"> <li>• Hausse du ticket modérateur hors parcours de soins</li> </ul>	
2010	de 65 à 70 % de 35 à 40 %
<ul style="list-style-type: none"> <li>• Hausse du ticket modérateur sur les médicaments à vignette bleue</li> <li>• Hausse du ticket modérateur sur les dispositifs médicaux</li> </ul>	

*Note* : Seules les modifications du taux de remboursement pour les soins qui restent dans le panier de soins remboursables sont résumées dans ce tableau. Ainsi, les déremboursements et les baisses de prix des médicaments ne sont pas pris en compte. Un résumé détaillé de l'ensemble des mesures prises entre 2005 et 2009 est disponible dans les comptes nationaux de la santé 2009.

Source : DG Trésor.

## 1. Renoncement à des soins médicaux et médicaments pour raisons financières



*Lecture* : Le renoncement aux soins est entendu dans cette enquête comme « n'a pas consulté un médecin lorsqu'il était malade », « n'a pas suivi les prescriptions du médecin » et « a abandonné des tests et traitements ».

*Source* : 2010 Commonwealth Fund International Health Policy Survey dans onze pays.

Avec la nécessité accrue de maîtriser les dépenses publiques, de nombreux pays ont pris des mesures afin d'infléchir la dynamique des dépenses de santé. Certains ont notamment augmenté la participation financière des assurés : mise en place de la participation forfaitaire (2005), des franchises médicales (2009), différentes hausses du ticket modérateur pour la France (cf. tableau 1) ; introduction d'un ticket modérateur de 10 euros par trimestre en 2003 en Allemagne ; introduction d'un paiement de 10 euros par consultation sur certains actes pratiqués par des spécialistes en Italie en 2011. Ces mesures augmentent le coût des soins pour les patients et ont pu, par conséquent, contribuer à accroître les phénomènes de renoncement aux soins. L'efficacité de ces mesures doit être jugée au regard des objectifs qui leur sont conférés afin de ne pas restreindre inutilement l'accès aux soins. Ces mécanismes incitatifs sont efficaces s'ils se traduisent par une diminution des dépenses inutiles, mais peuvent poser question s'ils conduisent des patients à renoncer, pour des raisons financières, à des soins nécessaires.

Entre risque de renoncement aux soins et surconsommation médicale, entre efficacité économique et efficacité sanitaire, la définition d'un système de remboursement équilibré, optimal au regard des arbitrages auxquels l'assureur doit faire face, apparaît délicate. Elle passe d'abord par la compréhension et la mesure de la réaction de la demande de soins aux incitations financières.

Dans ce contexte et dans la lignée de l'article publié par Geoffard (2000) sur l'« hypothèse d'aléa moral », nous dressons ici un bilan des conclusions des études nationales et internationales mesurant la sensibilité de la demande de soins aux incitations.

Dans un premier temps seront identifiées les différentes questions théoriques soulevées par l'étude de l'élasticité-prix de la demande de soins. Ensuite, seront détaillés les résultats des études à la fois sur la demande globale de soins mais également sur des soins particuliers mais centraux dans le système de santé (consultations d'un médecin, consommation de médicaments). Enfin, seront présentées des études qui, s'appuyant sur les enseignements théoriques et empiriques de la littérature, ont cherché à définir des contrats d'assurance optimaux.

## **1. Les questions soulevées par la mesure de l'élasticité-prix**

Afin d'interpréter correctement les différentes études qui seront présentées et de comprendre les enjeux, assurantiels notamment, de l'estimation de l'élasticité-prix de la demande de soins, une présentation synthétique du champ des différentes études et de ce que l'élasticité mesure, est menée.

### **1.1. Quelles mesures de la quantité et du prix ?**

Lorsque l'analyse porte sur un soin simple et clairement identifié, la quantité de soins consommés est facilement identifiable (nombre de consultations, nombre de boîtes de médicaments consommées). Quand la mesure de l'élasticité-prix<sup>(2)</sup> porte sur un service en particulier, la plupart des analyses s'intéressent donc au changement de la quantité de services utilisés (variation de la durée de séjours dans les hôpitaux, du nombre de consultations en ambulatoire, du nombre d'ordonnances) suite à une variation de prix.

En revanche, lorsque l'analyse porte sur un ensemble complexe de biens comme la consommation totale de soins, la consommation exprimée en volume est plus délicate à appréhender concrètement. En effet, la consommation totale de soins est un terme générique qui renvoie à des consommations de soins très différents tels que les soins hospitaliers et ambulatoires, les consultations chez un généraliste, spécialiste la consommation de médicaments. Les études relatives à la demande totale de soins s'appuieront donc sur l'analyse de chaque sous-poste de consommation avant d'agrèger les différents résultats afin d'estimer l'élasticité-prix globale de la demande de soins. On pourra combiner, par exemple, la mesure de la consommation de médicaments avec les consultations.

Parallèlement, déterminer le prix des soins est une tâche difficile. Le prix payé par l'individu dépend de la générosité de son contrat d'assurance.

(2) On rappelle que l'élasticité-prix de la demande de soin représente la sensibilité de la demande en dépenses de santé à la variation du prix de ces dépenses. Précisément, si l'élasticité vaut  $\epsilon$ , cela signifie que lorsque le prix des dépenses de santé s'accroît de 1 %, la demande en biens santé diminue de  $\epsilon$ .

Il est fonction des différents mécanismes de participation financière mais également de protection (cf. encadré). Le prix des soins peut donc dépendre de la quantité de services consommés précédemment.

Afin de déterminer le prix effectif des soins consommés, deux hypothèses de comportement individuel ont été envisagées :

- un comportement rationnel à anticipations parfaites. L'assuré anticipe l'effet de ses dépenses présentes sur le prix futur des soins. Il inclut cet effet dans la perception du prix des soins. L'exemple canonique est une femme enceinte qui dispose d'un plafond de dépense au-delà duquel les soins sont intégralement pris en charge par l'assureur. Elle anticipe un dépassement du plafond durant l'année du fait des dépenses élevées liées à sa grossesse. Une fois les soins pour sa grossesse reçus, les soins seront donc gratuits. Elle peut par conséquent considérer le coût des soins comme constant tout au long de l'année (avant et après le plafond) car ses dépenses n'auront aucun impact sur son reste à charge annuel. Ce comportement est donc équivalent à un comportement rationnel à anticipations parfaites. L'assuré fait ses choix comme s'il faisait face à un prix annuel unique. Ce prix peut être, par exemple, le prix de début d'année. La dépense de soins n'est alors plus sensible au système de remboursement des soins ;
- un comportement myope. L'assuré prend ses décisions de consommation selon le taux de remboursement du moment.

Keeler et Rolph (1988) ont testé ces deux hypothèses de comportement sur les données de la *Randomized Health Insurance Experiment* (RAND). La RAND reste la plus importante expérience jamais menée dans l'analyse de la demande de soins. Elle consistait en la création d'un échantillon aléatoire d'environ 5 800 personnes appartenant à 2000 familles américaines sans personnes âgées. Les dépenses de santé de ces individus étaient analysées sur une période de cinq ans. Keeler et Rolph concluent que l'hypothèse à retenir serait celle d'un comportement myope. Le choix d'utilisation des soins d'un consommateur dépendrait donc des différents taux de remboursement auxquels il doit faire face au moment du soin. C'est l'hypothèse retenue dans une majeure partie des études présentées. Pour autant, cette hypothèse ne fait pas l'unanimité. Certaines études privilégient l'hypothèse d'un comportement rationnel. Eichner (1998) utilise, par exemple, le prix constaté de fin d'année dans l'étude de l'élasticité-prix de la demande de soins aux Pays-Bas.

Par ailleurs, le prix ressenti des soins peut être augmenté par le prix du temps notamment passé pour bénéficier effectivement des soins. En effet, dans certains pays, des phénomènes de files d'attente existent ou les temps de transport sont très importants à cause des déserts médicaux. Pour autant, cette part du coût des soins n'est souvent pas prise en compte dans les études estimant l'élasticité-prix de la demande de soins.

## Définitions des mécanismes de participation financière et de protection des patients

### 1. Définition liminaire

*Tarif de référence* : l'assurance ou le financeur public prend en charge un montant fixe, les patients devant payer la différence entre ce montant et le prix réellement facturé.

### 2. Définitions des mécanismes de participation financière

*Ticket modérateur* : le patient paie une part du coût total, le reste étant à la charge de l'assureur ou du financeur public. En France, la Sécurité sociale prend pour base de calcul de remboursement la dépense valorisée sur la base du tarif de référence, auquel elle applique un taux qui diffère suivant les actes et prestations de soins, de l'état de santé de l'assuré, ou du respect ou non du parcours de soins coordonnés. Le ticket modérateur représente la différence entre la dépense valorisée sur la base du tarif de référence et le remboursement de l'assurance-maladie. La dépense totale pouvant être supérieure à la base de calcul (dépassements, médicaments non remboursables...), le ticket modérateur peut n'être ainsi qu'une part du reste à charge supporté par l'assuré avant éventuelle prise en charge par les organismes complémentaires.

*Franchise* : le mécanisme de couverture, et donc de remboursement, des dépenses de santé n'intervient qu'à partir du moment où les dépenses cumulées des patients dépassent le montant de la franchise. En deçà de ce montant, les patients supportent la totalité de leurs dépenses. Les franchises sont le plus souvent définies sur des périodes de douze mois (non glissantes).

*Remarque sur la participation forfaitaire d'un euro et sur les franchises médicales sur les boîtes de médicaments en France* : la participation forfaitaire et la franchise médicale sont déduites des remboursements effectués par l'assurance-maladie ; ainsi, ce sont des franchises à l'acte. L'assurance-maladie continue donc de rembourser une partie du coût des soins soumis à la participation forfaitaire ou à la franchise médicale ce qui n'est pas le cas avec la franchise.

### 3. Définition des mécanismes de protection des patients

*Plafond de dépense ou de reste à charge* : un plafond de dépense est défini, généralement sur une base annuelle calendaire (non glissante), en fonction de la dépense remboursable (dépense engagée dans la limite du tarif de référence) ou du ticket modérateur. Au-delà de ce plafond, l'assurance prend en charge la totalité du coût des soins, toujours dans la limite de la dépense remboursable. Le ticket modérateur marginal est alors nul.

## 1.2. Que mesure l'élasticité-prix de la demande de soins ?

### 1.2.1. Aléa moral sur le comportement du patient

L'asymétrie d'information entre l'assureur et l'assuré (sur le comportement de ce dernier pour prévenir la survenance du risque maladie ou face à la maladie) crée les conditions d'un aléa moral.

On distingue deux types d'aléa moral, *ex ante* et *ex post*, dans la relation principal (assureur)/agent (assuré). L'aléa moral *ex ante* influence la probabilité d'occurrence d'un risque. Dans la sphère de la santé, l'individu couvert peut être plus susceptible d'adopter des comportements à risque (fumer, mauvais régime alimentaire...). Comme l'individu valorise sa santé dans son capital humain (Grossman, 1972) et du fait des conséquences sur son bien-être, ce type d'aléa est supposé faible. Il existe, d'ailleurs, peu de preuves empiriques de la présence d'un tel aléa (Stanciole, 2008).

L'aléa moral *ex post* survient lorsque l'assuré traverse un épisode de maladie. Un assuré mieux couvert dépenserait davantage à état de santé donné. En diminuant le prix net des soins médicaux, le contrat d'assurance mènerait à une surconsommation des soins (Pauly, 1968, Ehrlich et Becker, 1972 et Feldstein, 1973). Lorsque l'on mesure l'élasticité-prix de la demande de soins, on mesure cet aléa moral, c'est-à-dire la variation de consommation de soins consécutive à une modification de la couverture, à état de santé donné.

L'assureur intervient en remboursant une partie du coût, ce qui introduit une distorsion dans le prix payé par l'assuré, pour deux raisons principales. Tout d'abord, l'assureur ne peut observer l'état de santé de l'assuré. Ensuite, il n'existe pas de référentiel de coût pour chaque maladie. Si ces deux conditions étaient remplies, l'assureur pourrait verser une somme forfaitaire aux patients en fonction de leur maladie et éviter ainsi tout risque d'aléa moral.

### 1.2.2. Caractéristiques et comportements individuels : aléa moral et effet de sélection

L'identification de l'aléa moral est rendue également délicate par la présence d'une seconde asymétrie d'information : la sélection adverse. Les individus qui anticipent des dépenses élevées choisissent un niveau de couverture plus élevé. La sélection adverse se traduit donc par une corrélation positive, comme l'aléa moral, entre niveau d'assurance et niveau de risque. Cependant, cette corrélation n'est pas liée à l'action de l'assuré mais à une caractéristique privée de celui-ci et notamment à sa perception de son exposition au risque maladie. Rothschild et Stiglitz (1976) ont mis en avant ce phénomène en démontrant qu'un individu qui se perçoit comme plus exposé au risque santé contractera une assurance plus généreuse qu'un individu qui s'estime en bonne santé. Dans le cas français, il choisira de s'assurer auprès d'une assurance complémentaire d'autant plus généreuse qu'il se perçoit comme fortement exposé au risque maladie.

Afin d'isoler l'élasticité-prix, il faut séparer ce qui tient de la sélection adverse et de l'aléa moral. En effet, il convient de ne mesurer que les changements de comportements des assurés au regard de la générosité de l'assurance et non les différences de comportements liées aux caractéristiques individuelles relatives à l'état de santé. Si la sélection adverse n'est pas contrôlée, l'élasticité mesure à la fois les différences de consommation entre des individus avec des états de santé (et donc des contrats d'assurance) différents et la sensibilité de la quantité demandée face à une variation du prix des soins à état de santé donné. La différence de consommation ne relèvera d'un phénomène d'aléa moral que si, pour l'individu qui a choisi une couverture plus généreuse, c'est l'augmentation de sa couverture assurantielle qui est à l'origine d'un changement de son comportement de consommation de soins. À noter que dans une logique purement budgétaire, il n'est pas forcément nécessaire de distinguer ce qui tient de la sélection adverse et de l'aléa moral. En effet, ce qui importe dans ce cas, c'est uniquement le changement de consommation globale suite à un changement de couverture du risque maladie.

Cette difficulté pratique de l'estimation de l'élasticité-prix explique notamment le faible nombre d'études qui lui sont consacrées. Nous effectuons une revue de littérature de ces différentes études depuis 1973, date des premières estimations.

## 2. Principales conclusions des études empiriques

Toute étude de l'impact de la couverture assurantielle sur la demande de soins ne peut pas se prévaloir de la même robustesse. Celle-ci dépend en effet souvent des données disponibles. Après une exposition des données utilisées par ordre croissant de qualité, les conclusions des études les plus robustes sur la demande totale de soins dans le monde, puis en France, seront détaillées. Enfin, la question de la sensibilité aux prix de la demande de consultation, en tant que soins primaires, et de médicament, comme poste majeur des dépenses de santé, a suscité un intérêt particulier. Les résultats de ces études plus spécifiques seront également présentés.

### 2.1. Les différentes méthodologies utilisées

Les méthodologies utilisées sont fortement influencées par les données disponibles. Les données idéales comporteraient les caractéristiques personnelles des individus, des détails sur l'utilisation des services médicaux et le prix effectif payé pour une unité supplémentaire de soins. Surtout, il faut que les variations de prix soient indépendantes des caractéristiques individuelles afin de ne mesurer que la modification de consommation du fait d'un changement du prix. Les différentes données utilisées sont ici présentées par ordre croissant de qualité et de robustesse des résultats :

- *données administratives ou d'enquêtes* : dans le cas des données administratives ou d'enquêtes, des modèles structurels sont utilisés



pour estimer la demande de soins. L'élasticité-prix de la demande de soins est alors obtenue en comparant les consommations de soins selon la générosité des contrats d'assurance dans la population. Ces modèles tentent de traiter les problèmes inhérents aux données non expérimentales tels que les biais de sélection (l'échantillon d'enquête ne constitue pas un groupe représentatif de la population), le biais de variables omises (certaines variables ne sont pas observées) et l'endogénéité (est-ce le niveau de remboursement qui détermine les dépenses de santé ou les dépenses de santé anticipées qui expliquent le choix de l'étendue de la couverture ?). En pratique, afin de neutraliser l'endogénéité des choix d'assurance, problème le plus récurrent et le plus problématique, deux stratégies principales d'identification ont été utilisées dans la littérature. La première méthode consiste à poser comme hypothèse que la corrélation entre choix d'assurance et consommation médicale ne tient qu'à des facteurs observables. En utilisant ces variables comme variables de contrôle, le niveau d'assurance n'est plus endogène dans l'équation de consommation (Caussat et Glaude, 1993 et Genier, 1998). La seconde méthode est l'estimation simultanée d'un système d'équations expliquant à la fois le choix d'assurance et le choix de consommation. L'enjeu principal de ces estimations est de déterminer une ou plusieurs variables, utilisées comme instrument, qui influencent le choix d'assurance mais pas directement la consommation de soins afin que le modèle soit identifiable (par exemple, le fait d'être fonctionnaire et des variables liées à la profession pour Buchmueller et *al.*, 2004)<sup>(3)</sup> ;

- *données quasi expérimentales* : les données quasi expérimentales, ou expériences naturelles, sont les données qui se rapprochent le plus des données expérimentales. Elles contiennent, comme ces dernières, des informations sur un groupe qui est le sujet de l'expérience (groupe de traitement). L'expérience prend la forme d'une modification exogène de l'environnement économique des individus. Cependant, ces données se distinguent des données expérimentales par la composition du groupe de contrôle, qui ne subit pas l'expérience, et du groupe de traitement. Les individus ne sont pas affectés aléatoirement dans chacun des groupes (Cook et Campbell, 1979). Par exemple, Chiappori et *al.* (1998) ont analysé l'impact de l'augmentation du ticket modérateur de 5 % suite à la mise en place de la loi Veil de 1993. Dans ce cadre, les auteurs disposent de données issues d'un organisme de prévoyance dans lesquelles figure la consommation de soins sur deux ans des assurés. L'assurance pour les assurés étudiés étant souscrite par l'entreprise, les effets de sélection adverse sont réputés bien contrôlés. L'étude porte sur la comparaison de deux groupes d'assurés au sein de cette population : ceux pour qui l'assurance complémentaire s'ajuste et prend en charge cette augmentation du ticket modérateur et ceux pour lesquelles elle ne

(3) Voir Geoffard (2000) pour une présentation plus détaillée des enjeux économétriques.

s'aligne pas et diminue même son propre taux de remboursement de 5 %. La baisse de leur couverture est alors de 10 %. Cette étude, en particulier, et l'ensemble des études quasi expérimentales reposent sur la méthode de la « double différence ». Cette méthode permet de contrôler l'ensemble des évolutions temporelles qui pourraient influencer l'objet de l'étude et qui seraient survenues même si la réforme n'avait pas eu lieu. Pour autant, il convient que les groupes de contrôle et de traitement aient des caractéristiques proches afin de comparer des individus comparables pour éviter de devoir en complément redresser les profils ce qui affaiblirait la robustesse des résultats ;

- *données expérimentales* : les données expérimentales utilisent l'affectation aléatoire d'individus entre un groupe de contrôle et un groupe de traitement qui subit une modification de ses conditions de remboursement lors d'une expérience. L'affectation aléatoire des individus permet d'éviter le biais de sélection. Cette méthode permet de déterminer l'impact de la générosité du système d'assurance tout en contrôlant les modifications de l'environnement socio-économique par la comparaison des effets mesurés dans chacun des groupes. Les résultats obtenus lors de ces expériences servent de référence. Bien que les données issues de ces expériences soient idéales, leur mise en œuvre est problématique. Tout d'abord, elles sont coûteuses : les données expérimentales les plus connues sont celles obtenues par la RAND. Ce projet a coûté 50 millions de dollars pour 5 800 participants. Ensuite, en dépit des redressements statistiques, ces données ne portent que sur un nombre limité d'individus et les résultats obtenus peuvent ne pas être généralisables. Les effets mesurés sur une petite population encadrée ne seront peut-être pas les mêmes lors d'une implémentation plus large.

## 2.2. Les enseignements des principales études empiriques

Les premiers travaux qui se sont attelés à mesurer l'élasticité-prix de la demande de soins datent des années soixante-dix. Les principales études concluent généralement que la demande de soins est faiblement élastique. La valeur centrale est  $-0,2$  : une hausse de 10 % du prix des soins induit en moyenne une réduction de 2 % de la dépense de soins. Les différentes études présentées s'attachent également à décrire le comportement de l'élasticité-prix de la demande de soins en fonction du taux de remboursement ou encore selon le type de soins ou de recours au système de soins. En effet, l'élasticité varie selon le prix et la quantité pour lesquels elle est mesurée, mais également selon le type de soins consommés.

### 2.2.1. Les premières études confirment la théorie

Les deux études pionnières (Rosett et Huang, 1973 et Phelps et

Newhouse, 1974) cherchant à déterminer l'élasticité-prix de la demande de soins ont été effectuées sur des données administratives ou d'enquêtes issues d'assureurs privés aux États-Unis. Dans le système de santé américain, les deux tiers des employés sont couverts par un assureur privé par le biais de leur employeur (Cohu, 2007).

Ces études concluent que la relation entre prix effectif des soins consommés et consommation de soins est bien décroissante. La demande de soins est donc sensible aux incitations monétaires.

Rosett et Huang (1973) ont estimé à  $-0,35$  l'élasticité-prix de la demande de soins lorsque le taux de remboursement est de 80 %. Lorsque le taux de remboursement diminue, les élasticités sont bien plus importantes, jusqu'à un maximum de  $-1,5$ .

La seconde étude, menée par Phelps et Newhouse (1974), montre que lorsque le taux de co-paiement varie de 20 à 25 % (respectivement de 15 à 20 % et de 10 à 15 %), l'élasticité-prix est de  $-0,12$  (resp.  $-0,07$  et  $-0,04$ ). Ces résultats semblent indiquer que des participations financières peu élevées infléchissent faiblement la demande de soins. Par ailleurs, pour une hausse du taux de co-paiement<sup>(4)</sup>, l'élasticité-prix observée est croissante avec le niveau de ce taux.

### 2.2.2. La RAND, une étude ancienne mais toujours de référence

Afin d'obtenir des estimations fiables de l'élasticité-prix de la demande de soins, le gouvernement américain a financé une expérience : la *Randomized Health Insurance Experiment* (RAND). Ces données expérimentales ont permis de s'affranchir de tous les problèmes de sélection et d'endogénéité qui sont présents dans les données administratives ou d'enquête. À chaque famille participante était affecté aléatoirement un contrat d'assurance. Les contrats différaient par leur taux de couverture des dépenses (5 %, 50 %, 75 %, 100 %) et par le montant maximal des dépenses médicales laissées à la charge des familles (plafond). Cette expérience unique a été critiquée sur un aspect du déroulement de l'expérience. Afin d'inciter à participer à l'enquête, les participants recevaient en effet une somme forfaitaire qui correspondait au montant du plafond. Ce dispositif pouvait donc influencer la consommation de soins des patients puisqu'il disposait d'une somme couvrant les dépenses de santé à venir dès le début de l'expérience. Cependant, comme cette somme s'ajoute au revenu du ménage, le dispositif devrait donc influencer plus, *a priori*, l'élasticité-revenu que l'élasticité-prix de la demande de soins.

En suivant pendant une période de cinq ans de 1974 à 1979 ces familles, la RAND a largement contribué à approfondir la compréhension des déterminants de la demande de soins. L'élasticité-prix pour la demande de soins était obtenue en comparant l'utilisation des services médicaux selon les différents contrats. Trois enseignements principaux peuvent être tirés de cette étude.

(4) Le taux de co-paiement représente la part de la dépense remboursable – qui n'est qu'une partie de la dépense totale – qui n'est pas prise en charge par l'assurance-maladie obligatoire.

Tout d'abord, la RAND confirme la décroissance des dépenses de santé lorsque le taux de remboursement diminue. Newhouse et *al.* (1993) estime que l'élasticité-prix de la demande de soins est de  $-0,2$  pour l'ensemble des soins. Cette estimation suggère donc que la demande de soins est faiblement élastique à son prix. Parallèlement, l'étude montre que la diminution de la consommation est due, principalement, à une baisse du nombre de recours au système de soins et non à une intensité moins importante des traitements lors d'un épisode de soins.

Ensuite, la plus forte diminution des dépenses de soins de ville est mesurée lorsque le taux de remboursement diminue de 100 à 75 %. La variation des dépenses de soins en ambulatoire en fonction du taux de remboursement est donc dégressive. Entre 100 et 75 %, l'élasticité-prix de la demande de soins est de  $-0,17$  alors qu'elle s'élève, en moyenne, à  $-0,22$  lorsque le taux de remboursement passe de 75 à 5 %. Ainsi, l'efficacité incitative sur le comportement des assurés d'une politique de déremboursement d'un même soin diminuera progressivement.

Enfin, la RAND a permis de mettre en évidence la diversité des élasticité-prix selon les types de recours (soins aigus, chroniques, hospitaliers).

Les soins hospitaliers sont moins sensibles aux prix que les soins ambulatoires. En effet le recours aux soins hospitaliers est peu sensible aux prix car il est souvent justifié par des interventions lourdes sur lesquelles le patient n'a que peu d'emprise. L'effet de la participation financière sur le recours aux soins hospitaliers semble diminuer, même, lorsque le taux de remboursement diminue. Ainsi, alors que l'élasticité-prix de la demande de soins hospitaliers est de  $-0,17$  entre 0 et 25 % de ticket modérateur, elle est de  $-0,14$  entre 25 et 95 % (*cf.* tableau 2). Une très forte hausse de la participation financière à l'hôpital n'a donc que peu d'effet incitatif sur la demande.

## 2. Arc-élasticité<sup>(\*)</sup> prix de la demande de soins selon les types de dépenses

*Écart-type entre parenthèses*

		Soins aigus	Soins chroniques	Soins ambulatoires	Hôpital	Ensemble des soins
Taux de co-paiement	0-25 %	-0,16 (0,02)	-0,2 (0,04)	-0,17 (0,02)	-0,17 (0,04)	-0,17 (0,02)
	25-95 %	-0,32 (0,02)	-0,23 (0,04)	-0,31 (0,04)	-0,14 (0,10)	-0,22 (0,06)

*Lecture* : Pour les soins aigus, lorsque le taux de co-paiement se situe entre 0 et 25 %, une augmentation du prix de 10 % diminue de 1,6 % la quantité consommée.

Note : (\*) L'arc-élasticité est l'élasticité d'une variable par rapport à l'autre entre deux points donnés. Elle est utilisée lorsqu'il n'y a pas de fonction générale pour définir la relation entre les deux variables. L'arc-élasticité est également défini comme l'élasticité entre deux points sur une courbe.

Source : Newhouse (1993).

### 2.2.3. Des études plus récentes confirment les résultats de la RAND

Des études récentes ont permis de confirmer les différents résultats de l'étude de la RAND. Trois études sur données américaines, néerlandaises et japonaises sont présentées. Les deux premières études analysent le comportement d'assurés auprès d'organismes privés. La dernière étude porte sur les assurés sociaux de l'assurance-maladie obligatoire japonaise.

Eichner (1998) se sert de données d'assurance d'entreprises américaines comprenant environ 17 000 employés âgés de 25 à 55 ans. Les différents contrats d'assurance comportent une franchise puis un taux de remboursement qui varie de 70 à 90 %. Une autre caractéristique importante de ces contrats est de couvrir l'ensemble de la famille de l'employé. Du fait de la franchise unique par famille, la consommation de soins d'un ayant droit diminue le prix des soins pour l'employé. Eichner compare alors des individus semblables mais dont la dépense de soins des ayants droit diffère. Si ces dépenses ont dépassé la franchise, des employés semblables ne feront pas face aux mêmes prix. En utilisant cette variation exogène des prix au sein d'une population, il conclut à une élasticité-prix de  $-0,32$ .

van Vliet (2004) utilise des données administratives néerlandaises (1993-1995) comprenant plus de cent mille assurés auprès d'assureurs privés. Au moment de l'étude, le système d'assurance-maladie néerlandais comprenait une assurance sociale obligatoire qui couvrait deux tiers de la population. L'autre tiers était assuré auprès d'une assurance privée. En effet, les employés dont le revenu était supérieur à 32 600 euros avaient le droit de s'affilier à une assurance privée et donc de sortir du système d'assurance public (Bernstein, 2011). Les données dont dispose l'auteur se focalisent donc sur une population aisée. Elles comprennent des informations sociodémographiques mais également sur la couverture assurantielle (la franchise choisie par l'individu) et le type de consommation de soins (dépenses en soins ambulatoires, d'hospitalisation et de médicaments). Les dépenses de médicaments des trois années précédant l'enquête sont aussi disponibles. Ces dépenses étaient entièrement couvertes par un système d'assurance publique indépendant. Elles sont alors considérées par l'auteur comme des bons indicateurs de l'état de santé des individus. Cette hypothèse identifiante permet de contrôler la sélection adverse. L'étude conclut à une élasticité-prix de  $-0,14$ .

Kan et Suzuki (2006) utilisent une expérience naturelle pour évaluer l'élasticité-prix de la demande de soins au Japon. Le Japon possède une assurance-maladie obligatoire et universelle, proche de celle de la France, mais avec un régime d'affiliation professionnel inspiré du modèle allemand. Cette recherche s'appuie sur des données de onze caisses d'assurance-maladie des employés. En 1997, le ticket modérateur au Japon est passé de 10 à 20 % pour les salariés (population âgée d'au plus 70 ans). Par la méthode de la double différence, Kan et Suzuki parviennent à identifier une élasticité de la demande de soins ambulatoires qui se situe entre  $-0,05$  et  $-0,06$ .

Des études plus anciennes menées également au Japon sur des données administratives ou d'enquête avaient évalué l'élasticité-prix de la demande de soins ambulatoires. Yoshida et Takagi (2002) concluent à une élasticité comprise entre  $-0,08$  et  $-0,26$ . Une étude plus ancienne de Bhattacharya et *al.* (1996) avait estimé une élasticité comprise dans un intervalle plus grand entre  $-0,12$  et  $-0,54$ .

Ces différentes études confirment donc en partie les estimations de la RAND. Malgré quelques divergences, elles montrent également la robustesse des résultats dans l'espace et dans le temps de l'estimation de l'élasticité-prix de la demande de soins.

#### *2.2.4. Les résultats sur données françaises*

Les études sur données françaises se sont focalisées sur l'effet de la couverture complémentaire sur les dépenses de soins.

L'analyse de l'effet de la couverture complémentaire est complexe. En effet, si on identifie une relation positive entre niveau de dépenses et niveau de couverture qui relève de l'individu, cette corrélation peut être due à un phénomène d'aléa moral (les assurés mieux couverts dépensent plus) comme de sélection adverse (les personnes qui anticipent des dépenses élevées choisissent un niveau de couverture plus élevé). On peut supposer que les deux phénomènes coexistent. Si, comme en France, le choix du niveau de couverture existe, il faut traiter simultanément la demande d'assurance et celle de soins.

Suite aux travaux de Newhouse (1974), Cameron et *al.* (1988) ont utilisé des modèles à équations simultanées afin d'identifier les phénomènes d'aléa moral. S'inspirant de ces modèles, Caussat et Glaude (1993) ont conclu à un effet positif du niveau de couverture sur la probabilité d'observer une consommation non nulle et dans le cas d'une consommation non nulle sur le niveau de la consommation en ambulatoire. Le fait d'être affilié auprès d'un organisme complémentaire expliquerait une probabilité de consommer supérieure de 12 % et une consommation totale supérieure de 30 %. Le différentiel de dépense porte surtout sur les consultations de spécialistes. Genier (1998) utilise la même base de données (enquête santé 1991-1992) et conclut également à la présence d'aléa moral. L'influence de l'assurance porterait surtout sur la décision de recourir aux soins lorsque survient une maladie et non sur la quantité de soins consommés pour soigner cette maladie en cas de recours.

### **2.3. Description synthétique de l'élasticité-prix de la demande de consultation médicale et de médicaments**

Certaines études ont également cherché à déterminer l'impact de la participation financière du patient non plus sur la demande globale de soins mais sur des types particuliers de soins.

Premier recours en cas de maladie et aiguilleur dans le système de santé, la demande de consultation médicale a été particulièrement étudiée. Un recours insuffisant au médecin est souvent perçu comme un facteur de hausse des dépenses à moyen ou long terme car il peut engendrer une aggravation de l'état de santé. Ainsi, mesurer les effets sur les consultations d'une modification du système de remboursement permet d'appréhender plus précisément l'impact d'un tel changement sur le recours aux soins.

En France, Chiappori et *al.* (1998) concluent à une absence d'aléa moral pour les visites en cabinet, ce qui suggère une élasticité de la demande proche de zéro pour de faibles variations de prix. Au contraire, dans le cas des visites à domicile, ils mettent en évidence une diminution de la demande car les coûts non monétaires (trajet, temps d'attente) sont faibles à la différence des visites en cabinet. Ainsi, lorsque les coûts non monétaires sont importants, une faible variation du prix de la visite ne représente qu'une faible part du coût total (monétaire et non monétaire) de la visite. À l'inverse, pour les visites à domicile, la variation du prix même faible représente une plus forte part du coût total, les coûts non monétaires étant peu élevés. La demande est dans ce cas plus sensible à une variation du prix. Même si aucune estimation n'est disponible en France, des études internationales ont permis d'estimer l'élasticité-prix relative à la demande de consultation médicale. Les estimations vont de  $-0,04$  à  $-0,36$ , avec une valeur médiane de  $-0,16$  (*cf.* tableau 3).

Parallèlement, la consommation de médicaments est l'un des postes les plus importants de la dépense de santé. En France, elle représente 20 % de la consommation de soins et de biens médicaux et 1,9 % du PIB. Afin de rationaliser la demande de médicaments, de nombreuses politiques ont joué sur les taux de remboursement. Cinq taux de remboursement sont actuellement en vigueur<sup>(5)</sup>.

En France, seules des études qualitatives ont tenté de déterminer la modification des comportements suite à la mise en place des franchises médicales sur les boîtes de médicaments (2008). Selon Kambia-Chopin et Perronnin (2010), 12 % des enquêtés déclarent avoir modifié leurs achats de médicaments suite à la mise en place des franchises. La probabilité d'avoir changé de comportement est plus élevée de 2 points chez les individus souffrant d'une pathologie chronique, comparés aux autres. Les études internationales ont permis cependant de mesurer la valeur de l'élasticité-prix de la demande de médicaments. Ainsi, l'élasticité-prix de la demande de médicaments s'étend de  $-0,01$  à  $-0,6$  avec comme valeur médiane  $-0,17$  (*cf.* tableau 3).

---

(5) 100 % pour les médicaments reconnus comme irremplaçables et particulièrement coûteux ; 65 % pour les médicaments dont le service médical rendu (SMR) aura été jugé important ou majeur et présentant un caractère indispensable ; 30 % pour les médicaments destinés aux maladies sans caractère habituel de gravité et dont le SMR n'a été reconnu ni comme majeur, ni comme important ; 15 % pour les médicaments dont le SMR est faible ou insuffisant ; 0 % pour les médicaments qui ne sont pas dans le panier de soins remboursables.

### 3. Élasticité-prix de la demande de consultation médicale et de médicaments

Méthode	Prix	Quantité mesurée	Élasticité prix
Visite chez le médecin <ul style="list-style-type: none"> <li>Fuchs et Kramer (1972)</li> <li>Beck (1974)</li> <li>Scitovsky et McCall (1977)</li> <li>Wedig (1988)</li> <li>Cherkin <i>et al.</i> (1989)</li> <li>Chiappori <i>et al.</i> (1998)</li> <li>Cockx et Brasseur (2003)</li> <li>Kan et Suzuki (2010)</li> </ul>	Taux de remboursement Reste à charge Taux de remboursement Reste à charge Reste à charge Taux de remboursement Taux de remboursement Taux de remboursement	Nombre de visites Nombre de visites Nombre de visites Nombre de visites Nombre de visites Nombre de visites Nombre de visites	-0,10 à -0,36 -0,07 -0,29 -0,16 à -0,35 -0,04 0 sur les consultations en cabinet > 0 pour les consultations à domicile -0,06 -0,055
Consommation de médicaments <ul style="list-style-type: none"> <li>O'Brien (1989)</li> <li>Lavers (1989)</li> <li>Smith (1993)</li> <li>Newhouse <i>et al.</i> (1993)</li> <li>Hughes et McGuire (1995)</li> <li>Contoyannis <i>et al.</i> (2005)</li> <li>Landsman <i>et al.</i> (2005)</li> <li>Fiore et Siciliani (2010)</li> </ul>	Reste à charge Reste à charge Taux de remboursement Taux de remboursement Reste à charge Taux de remboursement Reste à charge Reste à charge	Nombre de prescriptions Nombre de prescriptions Nombre de prescriptions Dépenses de médicaments Nombre de prescriptions Dépenses de médicaments Nombre de prescriptions Nombre de prescriptions	-0,33 -0,15 à -0,20 -0,1 -0,17 -0,35 -0,12 à -0,16 -0,1 à -0,6 -4 % nombre. de prescription par fête pour une augmentation de 1 euro



### 3. Quelles conclusions en termes de politiques publiques ?

Les différentes études présentées ont permis de mieux appréhender le comportement de l'assuré face au coût de recours à des services de santé. Des travaux ont alors été menés afin de déterminer le contrat d'assurance optimal qui permettrait à la fois de garantir une couverture du risque maladie satisfaisante tout en instaurant des incitations financières efficaces face aux problèmes d'aléa moral.

#### 3.1. Les études théoriques d'un schéma d'assurance optimal

Cinq études théoriques de référence ont cherché à déterminer un schéma optimal d'assurance (cf. tableau 4). Les paramètres de l'assurance dépendent des hypothèses faites notamment sur la fonction d'utilité retenue. Les modèles théoriques sont ensuite testés sur des données empiriques (par exemple, Newhouse (1993) sur les données de la RAND).

Depuis Feldstein et Friedman (1977) où le contrat d'assurance optimal ne comprenait qu'un ticket modérateur, les schémas d'assurance optimaux sont composés de différents mécanismes de participations financières. Les contrats de Buchanan (1991) et Newhouse (1993) combinent franchise et ticket modérateur à la différence de celui de Blomqvist (1997) qui se caractérise par des tickets modérateurs qui évoluent en fonction de la dépense. Par ailleurs, certains contrats comprennent des mécanismes de protection des assurés. Ainsi, un plafond de dépense dans le but de protéger les assurés contre des restes à charge élevés fait partie des schémas d'assurance optimal de Newhouse (1993) et de Manning (1996).

#### 4. Schéma d'assurance optimal

	Schéma d'assurance optimal
Feldstein et Friedman (1977)	Ticket modérateur unique de 58%
Buchanan, Keeler et <i>al.</i> (1991)	1. Franchise 200 dollars 2. Ticket modérateur unique de 25%
Newhouse et <i>al.</i> (1993)	1. Franchise de 200 à 300 dollars 2. Ticket modérateur unique de 25% 3. Plafond de dépense de 1 000 dollars
Manning et Marquis (1996)	1. Ticket modérateur unique de 25% 2. Plafond de dépense de 25 000 dollars
Blomqvist (1997)	1. Ticket modérateur de 27 % jusqu'à 100 dollars de dépenses 2. Le ticket modérateur diminue alors progressivement pour atteindre 5% pour une dépense de 30 000 dollars

Certains points communs ressortent de ces cinq études et peuvent être interprétés comme des voies d'amélioration des différents systèmes de remboursement des soins, ce que nous explorons maintenant.

## 3.2. Conclusions en termes de politiques publiques

Dans la plupart des systèmes de santé, l'assurance-maladie a deux fonctions principales. Tout d'abord, elle prend en charge les dépenses de soins courants c'est-à-dire les soins qui découlent d'un épisode de maladie temporaire. Ensuite, elle couvre les dépenses relatives aux maladies chroniques. L'étude des schémas d'assurance optimaux et la connaissance de la demande de soins permettent de définir des régulations efficaces de la dépense de soins courants et chroniques.

### 3.2.1. Quel schéma d'assurance pour les soins courants ?

Quatre schémas optimaux d'assurance sur les cinq étudiés sont non linéaires. Ainsi, l'assuré doit faire face à des taux de prise en charge différents au cours de l'année. La diversité des mécanismes de participation financière permet d'adapter les incitations aux aléas que l'on cherche à modérer. La franchise est reconnue comme un bon instrument pour lutter contre l'aléa moral *ex ante*. En effet, l'individu sera d'autant plus incité à prévenir la maladie qu'il sait que les premières dépenses auxquelles il devra faire face en cas de problème de santé seront entièrement à sa charge. Parallèlement, le ticket modérateur vise plutôt à agir sur l'aléa moral *ex post*. Le patient sera d'autant plus enclin à modérer ses dépenses de soins qu'il doit en payer une part et que ce co-paiement croît avec la dépense engagée.

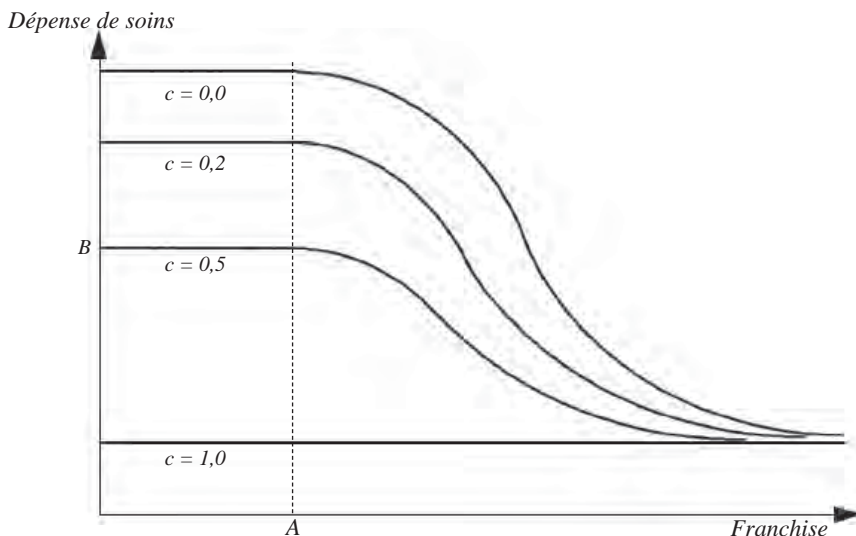
Ensuite, ces schémas d'assurance contiennent un ticket modérateur unique. Sur données américaines, à l'exception de Feldstein et Friedman (1977), les études suggèrent de fixer le taux de ticket modérateur aux alentours de 25 %. Ces résultats concordent avec les conclusions des études de la RAND. L'élasticité-prix de la demande de soins est dégressive et la plus forte diminution des dépenses est observée lorsque le taux de remboursement passe de 100 à 75 %. Ainsi, selon ces études, il semble efficace d'instaurer un taux ticket modérateur aux alentours de 25 %. Même si un taux de ticket modérateur plus important permettrait de réduire les dépenses de soins, les gains en termes d'incitation pourraient être plus faibles que les pertes en termes de couverture du risque.

Enfin, ces études suggèrent qu'une combinaison d'une franchise et d'un plafond de dépense peut garantir un équilibre entre incitation financière efficace et protection des assurés contre des restes à charge élevés.

Des franchises, d'un montant non symbolique, seraient des incitations financières efficaces. Keeler, Newhouse et Phelps (1977) ont conclu que, lorsque la franchise est basse, les individus se comportent comme si elle n'existait pas. En effet, on constate que pour des niveaux de franchise faibles, la dépense de soins reste constante jusqu'à un certain montant de franchise à partir duquel la dépense de soins commence à décroître. Cette même étude montre que les assurés se comportent comme s'ils n'avaient pas d'assurance lorsque la franchise est élevée. L'ensemble des dépenses

de soins se rejoignent à partir d'un certain niveau de franchise. Lorsqu'une franchise à la base est instaurée, son montant doit donc être raisonnable afin de ne pas dresser une barrière trop importante à l'accès aux soins mais non symbolique pour influencer effectivement le comportement des assurés (cf. graphique 2).

## 2. Demande totale de soins en fonction du montant de la franchise et du niveau du ticket modérateur



*Lecture* : Lorsque la franchise est égale à A, la dépense de soins est égale à B pour un individu qui fait face à un taux de ticket modérateur de 50 % une fois la franchise dépassée.

*Note* : «  $c$  » (coinsurance) correspond à la valeur du ticket modérateur.

*Source* : Keeler et al. (1977).

La majorité des contrats d'assurance optimaux comprennent une participation financière nulle (plafond de dépense) ou faible au-delà d'un certain niveau de dépenses. La légitimité d'un plafond de dépenses s'explique, tout d'abord, par l'importance de protéger les assurés contre les risques catastrophiques. Ces risques, à faible probabilité mais qui entraînent des dépenses importantes, sont la raison d'être de tout système d'assurance. En effet, lorsqu'ils surviennent, ils peuvent être à l'origine d'un appauvrissement important et le système d'assurance doit alors jouer à plein pour protéger les assurés. Au-delà du plafond, lorsque les risques catastrophiques se sont réalisés, l'assuré n'a qu'une faible emprise sur la dépense de soins, ce qui réduit fortement toute suspicion d'aléa moral. La gravité de la maladie impose souvent aux patients de déléguer l'ensemble de l'autorité médicale au médecin qui décide alors seul du parcours de soins. Toute incitation finan-

cière pesant sur le patient semble alors superflue. Pour autant, il convient que seuls les assurés traversant des épisodes de maladie coûteux saturant le plafond. En effet, plus le plafond est bas plus le risque d'une surconsommation médicale liée à la gratuité des soins est élevé. Il convient donc de fixer le plafond à un montant qui permette une bonne couverture des risques catastrophiques tout en évitant qu'il contribue à une surconsommation des soins. Dans l'étude de la RAND, le comportement de consommation de soins change très faiblement pour des niveaux de plafond de dépenses supérieurs à 1 000 dollars. Ainsi, le risque d'aléa moral au-delà de 1 000 dollars de dépenses semblait inexistant à l'époque et explique donc que le plafond dans le contrat d'assurance optimal calculé par Newhouse (1993) soit de ce montant. En France, la mise en place d'un tel plafond a été évoquée lors de la discussion sur le bouclier sanitaire (2007).

### 3.2.2. *Quelle prise en charge des soins chroniques ?*

Les progrès médicaux et l'amélioration des traitements sont à l'origine d'une place grandissante des maladies chroniques. Ces transformations imposent d'adapter les systèmes de soins à cette transition afin de contrôler la croissance des coûts de santé.

La présence d'une asymétrie d'information entre l'assureur et l'assuré sur l'état de santé de ce dernier justifie la mise en place d'incitations financières. Or, dans le cas des maladies chroniques, l'assureur peut observer l'état de santé de l'assuré. Ainsi, selon la théorie assurantielle, rien ne justifie la mise en place d'incitations financières pour ce type de soins.

Pour autant, il n'existe pas de référentiel de coût pour chaque maladie chronique. Si tel était le cas, l'assureur pourrait verser une somme forfaitaire aux patients en fonction de leur maladie chronique. Du fait de ce manque, différents systèmes de prise en charge des soins chroniques, plus ou moins généreux, ont été mis en place. En France, le système des affections de longue durée (ALD) permet la prise en charge intégrale des soins liés à la maladie chronique. L'Allemagne, au contraire, ne prend pas en charge intégralement les soins liés aux maladies chroniques. Les assurés allemands souffrant d'une maladie chronique bénéficient d'un plafond de reste à charge inférieur aux autres assurés (1 % des ressources annuels brutes contre 2 % pour les assurés de droit commun).

## **Conclusion**

Les différentes études présentées dans ce complément cherchent à déterminer les caractéristiques d'un contrat d'assurance-maladie et le niveau de co-paiements qui réalisent un arbitrage optimal entre la réduction de l'aléa moral et la couverture du risque maladie. Les conclusions de ces études théoriques et empiriques permettent de juger de l'efficacité des mécanismes de régulation de la demande de soins tels que la franchise, le ticket modérateur unique et le plafond de dépense.

Même si la régulation de la dépense de soins suppose la mise en place d'un système d'assurance-maladie limitant la surconsommation médicale, les incitations ne semblent pas devoir porter que sur les patients.

En effet, un second courant de la littérature étudie les phénomènes de demande induite par le médecin (Rice, 1983). La théorie de la demande induite suppose que ce sont les médecins qui déterminent la consommation médicale des patients. Le médecin est alors, dans certaines circonstances, à l'origine d'une surproduction médicale (Rochaix et Jacobzone, 1997, Delattre et Dormont, 2000, Franc, 2001 et Trinquard, 2006). Par exemple, dans le cadre d'une rémunération à l'acte comme c'est le cas en France, le médecin peut prescrire davantage que nécessaire. Il peut adopter un tel comportement pour atteindre notamment un revenu cible. Les études relatives à la demande induite se sont attachées à définir les mécanismes de régulation de l'offre de soins qui permettraient de réduire les risques de surproduction ou de sous-production de soins par le professionnel de santé. Cette régulation s'appuie généralement sur des modes de rémunération des médecins qui permettent à la fois de garantir une certaine qualité de la prestation fournie et un traitement au juste coût. Une analyse approfondie des avantages et inconvénients de chacun des modes de rémunération en théorie et en pratique dans différents pays européens a été menée par la Direction générale du Trésor (Albouy et Déprez, 2009). Cette étude suggère que la réduction de la demande induite devrait s'appuyer sur un mode de rémunération mixte du médecin, tout en montrant la difficulté de définir un schéma de rémunération optimal.

Si, comme le suggère cette analyse de la littérature sur la demande de soins, une régulation efficace du phénomène d'aléa moral est nécessaire, une politique plus générale visant une responsabilisation symétrique des patients et des médecins permettrait d'avoir de meilleurs résultats sur l'inflexion de la demande de soins qu'une politique asymétrique (Bardey et Lesur, 2006). En effet, des mesures agissant à la fois sur l'offre et la demande de soins permettront de réduire simultanément l'aléa moral et la demande induite. La littérature théorique et empirique s'intéressant simultanément aux incitations destinées aux patients et médecins est encore embryonnaire. L'enrichissement de cette littérature devrait permettre de créer un pont entre ces deux courants dans un souci de régulation efficace des dépenses de soins.

## Références bibliographiques

- Albouy V. et M. Déprez (2009) : « Mode de rémunération des médecins », *Économie et Prévision*, n° 188.
- Bardey D. et R. Lesur (2006) : « Optimal Regulation of Health System with Induced Demand and *ex post* Moral Hazard », *Annales d'Économie et de Statistique*, vol. 83, n° 84, pp. 279-294.
- Beck R.G. (1974) : « The Effects of Copayment on the Poor », *Journal of Human Resources*, vol. 9, n° 1, pp. 129-142.
- Bernstein D. (2011) : « Pays-Bas : la théorie de la concurrence régulée à l'essai », *Les Tribunes de la Santé*, n° 30, pp. 127-148.
- Bhattacharya J., W.B. Vogt, A. Yoshikawa et T. Nakahara (1996) : « The Utilization of Outpatient Medical Services in Japan », *Journal of Human Resources*, vol. 31, n° 2, pp. 450-476.
- Blomqvist A.G. (1997) : « Optimal Non-Linear Health Insurance », *Journal of Health Economics*, vol. 16, n° 3, pp. 303-321.
- Buchanan J.L., E.B. Keeler, J.E. Rolph et M.R. Holmer (1991) : « Simulating Health Expenditures Under Alternative Insurance Plans », *Management Science*, vol. 37, n° 9, pp. 1067-1089.
- Buchmueller T.C., A. Couffinhal, M. Grignon et M. Perronnin (2004) : « Access to Physician Services: Does Supplemental Insurance Matter? Evidence from France », *Health Economics*, vol. 13, n° 7, pp. 669-687.
- Cameron A.C., P.K. Trivedi, F. Milne et J. Piggott (1988) : « A Microeconomic Model of the Demand for Health Care and Health Insurance in Australia », *Review of Economic Studies*, vol. 55, n° 1, pp. 85-106.
- Caussat L. et M. Glaude (1993) : « Dépenses médicales et couverture sociale », *Économie et Statistique*, n° 265, pp. 31-44.
- Cherkin D.C., L. Grothaus et E.H. Wagner (1989) : « The Effect of Office Visit Copayments on Preventive Care Services in an HMO », *Inquiry*, vol. 27, n° 1, pp. 24-38.
- Chiappori P-A., F. Durand et P-Y. Geoffard (1998) : « Moral Hazard and the Demand for Physician Services: First Lessons from a French Natural Experiment », *European Economic Review*, vol. 42, n° 3-5, pp. 499-511.
- Cockx B. et C. Basseur (2003) : « The Demand for Physician Services. Evidence from a Natural Experiment », *Journal of Health Economics*, vol. 22, n° 6, pp. 881-913.
- Cohu S. et D. Lequet-Slama (2007) : « Le système d'assurance santé aux États-Unis. Un système fragmenté et concurrentiel », *Études et Résultats*, n° 600.

- Contoyannis P., J. Hurley, P. Grootendorst, S.H. Jeon et R. Tamblyn (2005) : « Estimating the Price Elasticity of Expenditure for Prescription Drugs in the Presence of Non-linear Price Schedules: An Illustration from Quebec, Canada », *Health Economics*, vol. 14, n° 9, pp. 909-923.
- Cook T.D. et D.T. Campbell (1979) : *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Houghton Mifflin Company, Boston.
- Delattre E. et B. Dormont (2000) : « Induction de la demande de soins par les médecins libéraux français », *Économie et Prévision*, n° 142, pp. 137-161.
- Ehrlich I. et G.S. Becker (1972) : « Market Insurance, Self-Insurance, and Self Protection », *Journal of Political Economy*, vol. 80, n° 4, pp. 623-648.
- Eichner M.J. (1998) : « The Demand for Medical Care: What People Pay Does Matter », *American Economic Review*, vol. 88, n° 2, pp. 117-121.
- Feldstein M.S. (1973) : « The Welfare Loss of Excessive Health Insurance », *Journal of Political Economy*, vol. 81, n° 1, pp. 251-280.
- Feldstein M.S. et B. Friedman (1977) : « Tax Subsidies, the Rational Demand for Insurance and the Health Care Crisis », *Journal of Public Economics*, vol. 7, n° 2, pp. 155-178.
- Fiorio C. et L. Siciliani (2010) : « Co-Payments and the Demand for Pharmaceuticals: Evidence from Italy », *Economic Modelling*, vol. 27, n° 4, pp. 835-841.
- Franc C. (2001) : « Mécanismes de rémunération et incitations des médecins », *Économie Publique, Études et Recherches*, n° 9, n° 3, pp. 13-36.
- Fuchs V.R. et M.J. Kramer (1972) : « Determinants of Expenditures for Physicians' Services in the United States, 1948-1968 », *National Bureau of Economic Research Occasional Paper*, n° 117.
- Genier P. (1998) : « Assurance et recours aux soins : une analyse micro-économétrique à partir de l'enquête Santé 1991-1992 de l'INSEE », *Revue Économique*, vol. 49, n° 3, pp. 809-819.
- Geoffard P-Y. (2000) : « Dépenses de santé : l'hypothèse d'aléa moral », *Économie et Prévision*, n° 142, pp. 123-136.
- Grossman M. (1972) : « On the Concept of Health Capital and Demand for Health », *Journal of Political Economy*, vol. 80, n° 2, pp. 223-255.
- Hughes D. et A. McGuire (1995) : « Patient Charges and the Utilization of NHS Prescription Medicines », *Health Economics*, vol. 4, n° 3, pp. 213-220.
- Kambia-Chopin B. et M. Perronnin (2010) : « Les franchises ont-elles modifié les comportements d'achat des médicaments ? », *Questions d'Économie de la Santé*, n° 158.

- Kan M. et W. Suzuki (2006) : « The Demand for Medical Care in Japan: Initial Findings from a Japanese Natural Experiment », *Applied Economics Letters*, vol. 13, n° 5, pp. 273-277.
- Kan M. et W. Suzuki (2010) : « Effects of Cost Sharing on the Demand for Physician Services in Japan: Evidence from a Natural Experiment Japan and the World Economy », *Revue*, vol. 22, n° 1, pp. 1-12.
- Keeler E.B., J.P. Newhouse et C.E. Phelps (1977) : « Deductibles and Demand: A Theory of the Consumer Facing a Variable Price Schedule under Uncertainty », *Econometrica*, vol. 45, pp. 641-655.
- Keeler E.B. et J.E. Rolph (1988) : « The Demand for Episodes of Treatment in the Health Insurance Experiment », *Journal of Health Economics*, vol. 7, n° 4, pp. 337-367.
- Landsman P.B., W. Yu, X.F. Liu, S.M. Teutsch et M.L. Berger (2005) : « Impact of 3-Tier Pharmacy Benefit Design and Increased Consumer Cost-Sharing on Drug Utilization », *American Journal of Managed Care*, vol. 11, n° 10, pp. 621-628.
- Lavers R. (1989) : « Prescription Charges, the Demand for Prescriptions and Morbidity », *Applied Economics*, vol. 21, n° 8, pp. 1043-1052.
- Manning W.G. et M.S. Marquis (1996) : « Health Insurance: The Tradeoff Between Risk Pooling and Moral Hazard », *Journal of Health Economics*, vol. 15, n° 5, pp. 609-639.
- Newhouse J.P. (1993) : *Free for All? Lessons from the RAND Health Insurance Experiment*, University Press.
- O'Brien B. (1989) : « The Effect of Patient Charges on the Utilization of Prescription Medicines », *Journal of Health Economics*, vol. 8, n° 1, pp. 109-132.
- Pauly M.V. (1968) : « The Economics of Moral Hazard: Comment », *American Economic Review*, vol. 58, n° 3, pp. 531-537.
- Phelps C.E. et J.P. Newhouse (1974) : « Coinsurance, The Price of Time, and the Demand for Medical Services », *The Review of Economics and Statistics*, vol. 56, n° 3, pp. 334-342.
- Rice T. (1983) : « The Impact of Changing Medical Care Reimbursement Rates on Physician-Induced Demand », *Medical Care*, vol. 21, n° 8, pp. 803-815.
- Rochaix L. et S. Jacobzone (1997) : « L'hypothèse de demande induite : un bilan économique », *Économie et Prévision*, n° 129-130, pp. 129-140.
- Rosett R.N. et L. F. Huang (1973) : « The Effect of Health Insurance on the Demand for Medical Care », *Journal of Political Economy*, vol. 81, n° 1.
- Rothschild M. et J. Stiglitz (1976) : « Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information », *Quarterly Journal of Economics*, vol. 90, n° 4, pp. 629-650.



- Scitovsky A.A. et N.M. McCall (1977) : « Coinsurance and the Demand for Physician Services: Four Years Later », *Social Security Bulletin*, vol. 40, n° 5, pp. 19-27.
- Smith D. (1993) : « The Effects of Copayments and Generic Substitution on the Use and Costs of Prescription Drugs », *Inquiry*, vol. 30, n° 2, pp. 189-198.
- Stanciole A.E. (2008) : « Health Insurance and Lifestyle Choices: Identifying *ex ante* Moral Hazard in the US Market », *The Geneva Papers on Risk and Insurance, Issues and Practice, Palgrave Macmillan Journals*, vol. 33, n° 4, pp. 627-644.
- Trinquard S. (2006) : « Demande induite par l'offre ambulatoire : un survol de la littérature théorique et empirique », *Cahiers du LASER*, n° 20.
- van Vliet R. (2004) : « Deductibles and Health Care Expenditures: Empirical Estimates of Price Sensitivity Based on Administrative Data », *International Journal of Health Care Finance and Economics*, vol. 4, n° 4, pp. 283-305.
- Wedig G.J. (1988) : « Health Status and the Demand for Health », *Journal of Health Economics*, vol. 7, n° 2, pp. 151-163.
- Yoshida A. et S. Takagi (2002) : « Effects of the Reform of the Social Medical Insurance System in Japan », *The Japanese Economic Review*, vol. 53, n° 4, pp. 444-465.
- Zweifel P. et W.G. Manning (2000) : « Moral Hazard and Consumer Incentives in Health Care », chapitre 8 in *Handbook of Health Economics*, vol. 1, Culyer et Newhouse (eds), pp. 409-459, 1<sup>re</sup> édition.



## Complément B

# Assurance-maladie : évaluation de modifications du système de remboursement des soins

Grégoire de Lagasnerie  
*Direction générale du Trésor*

La France a défini en 2011 une trajectoire pluriannuelle de retour à l'équilibre des finances publiques à l'horizon 2016 qui se décline pour l'assurance-maladie sous la forme d'un objectif national des dépenses d'assurance-maladie (ONDAM) limité à une évolution annuelle de 2,5 % par an sur la période 2012-2016. Pour ramener la dynamique naturelle des dépenses d'assurance-maladie sur cet objectif, la programmation pluriannuelle présentée le 23 novembre 2011<sup>(1)</sup> prévoit que l'assurance-maladie obligatoire devra réaliser environ 13 milliards d'euros d'économies en dépenses entre 2013 et 2016.

Afin d'améliorer la conduite des politiques publiques en France dans le domaine des finances publiques et des questions de santé, la DG Trésor a développé un outil de microsimulation (ISIS)<sup>(2)</sup> des dépenses de soins de santé et de leur prise en charge par les patients, les organismes obligatoires et complémentaires d'assurance-maladie (*cf.* annexe). L'utilisation de ce type d'outil présente de nombreux avantages et connaît un développement croissant (Jones, 2010). Le modèle de microsimulation ISIS permet notamment d'étudier finement par type de population l'impact de diverses mesures sur le remboursement des soins des assurés et d'évaluer les mesures d'économies qui pourraient porter sur le système de prise en charge des soins par l'assurance-maladie publique.

(1) Mise à jour du rapport économique, social et financier.

(2) Instrument de simulation individualisée des dépenses de santé. Maquette construite par Valérie Albouy et Bérengère Mesqui en collaboration avec la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES) et avec la participation d'Émilie Rodriguez, Claire Montialoux et Grégoire de Lagasnerie.

Le présent complément explicite dans un premier temps les défis auxquels devra faire face l'assurance-maladie obligatoire à l'horizon 2016. Dans ce contexte, sont ensuite présentés et analysés, à partir du modèle de microsimulation ISIS, différents scénarios d'évolution possible du système de remboursement des soins. D'autres évolutions de la participation financière du patient ont déjà été analysées (Briet et Fragonard, 2007, Debrand et Sorasith, 2010 et Geoffard et *al.*, 2011). Cette analyse complète ainsi la littérature en appréhendant concrètement l'impact de différentes réformes possibles du système de remboursement des soins.

## **1. L'assurance-maladie dans le cadre de la stratégie pluriannuelle des finances publiques**

Dans un contexte de déficits publics élevés, la France a décidé de renforcer ses efforts de maîtrise budgétaire en définissant une trajectoire pluriannuelle de retour à l'équilibre des comptes publics d'ici 2016.

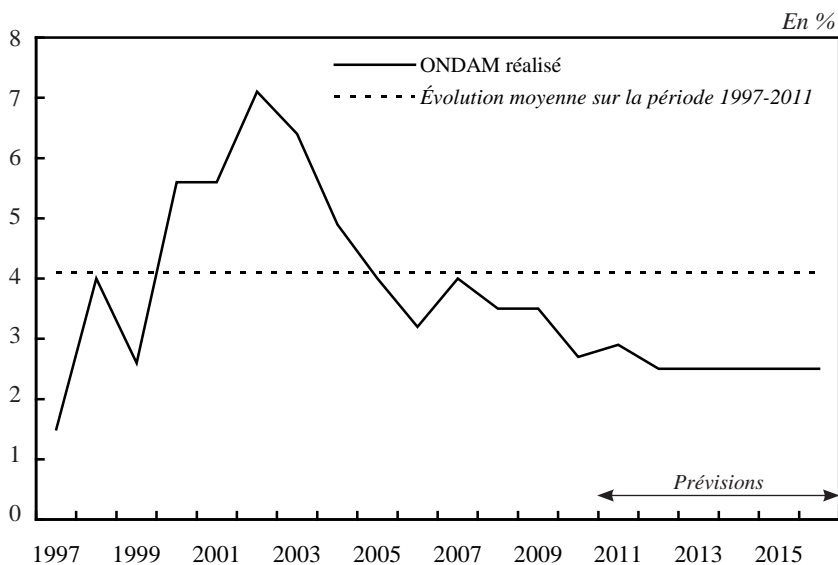
### **1.1. La stratégie pluriannuelle de retour à l'équilibre des finances publiques en 2016**

De 7,1 % du PIB en 2010, le déficit public devrait être ramené à 4,4 % en 2012 et 3 % en 2013 conformément à l'engagement pris par la France vis-à-vis de ses partenaires européens. Cette stratégie pluriannuelle rigoureuse oblige à appréhender et déterminer les efforts nécessaires dans un souci d'efficacité et d'équité.

Dans ce contexte, des objectifs ambitieux de maîtrise des dépenses d'assurance-maladie ont été définis. Les taux de progression des dépenses d'assurance-maladie (ONDAM, Objectif national de dépenses d'assurance-maladie) sont ainsi fixés à 2,5 % par an jusqu'en 2016. L'effort demandé est sans précédent depuis la création de l'ONDAM en 1997 (*cf.* graphique 1).

Pour assurer le respect de ces objectifs ambitieux, les efforts d'économies sont calibrés annuellement à partir d'une évolution tendancielle des dépenses d'assurance-maladie. À partir du niveau de l'ONDAM en début d'année, est calculée une prévision des dépenses sous ONDAM en fin d'année en fonction de leur évolution spontanée au regard de la tendance passée et en l'absence de toute mesure ou événement nouveaux. Sur le champ retenu pour l'ONDAM, les dépenses de santé évolueraient aujourd'hui spontanément de l'ordre de 4 % par an. Pour respecter un objectif de 2,5 % de croissance de l'ONDAM, il est donc nécessaire de réaliser des économies annuelles sur la tendance spontanée représentant environ 1,5 % des dépenses sous ONDAM jusqu'en 2016.

## 1. Taux de progression de l'ONDAM entre 1997 et 2016



*Lecture* : Entre 1997 et 2010, est indiqué l'ONDAM réalisé. À partir de 2011, est indiqué l'ONDAM voté ou prévu. Les évolutions annuelles ne sont pas rigoureusement comparables car le périmètre de l'ONDAM peut varier d'une année sur l'autre.

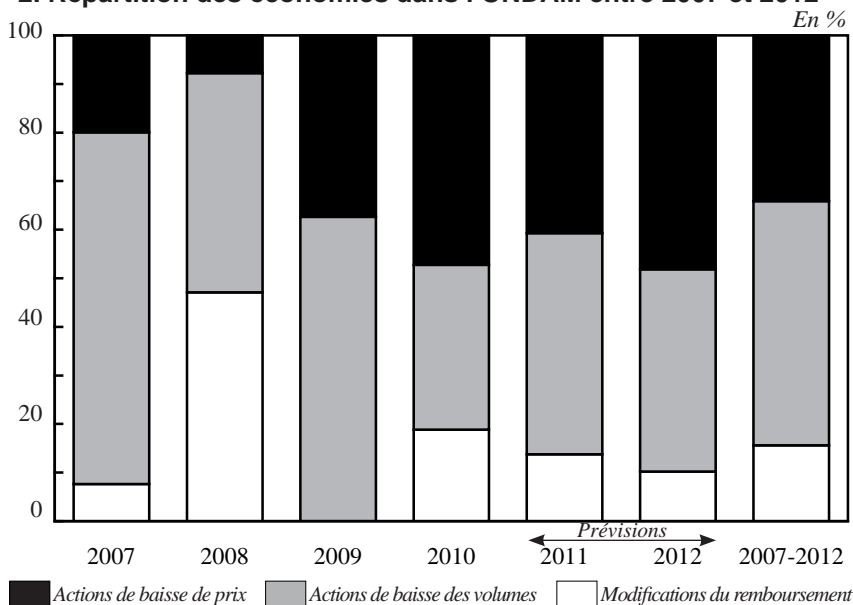
*Source* : DG Trésor.

La dernière loi de financement de la Sécurité sociale (LFSS) documente les nouvelles mesures d'économies permettant de construire l'ONDAM pour 2012. Environ 13 milliards d'euros d'économies sur les dépenses d'assurance-maladie restent à réaliser entre 2013 et 2016 pour sécuriser une évolution de 2,5 % par an de l'ONDAM sur cette période. Ces économies représentent 26 % du total des mesures d'économies en dépenses<sup>(3)</sup> et 55 % des mesures d'économies sur les administrations de Sécurité sociale entre 2013 et 2016 (*cf.* Stratégie pluriannuelle de finances publiques. Mise à jour du rapport économique, social et financier. <http://www.budget.gouv.fr/files/mise-a-jour-rapport-economique-social-financier.pdf>).

Le contexte de poursuite dans les prochaines années de la maîtrise des dépenses publiques présente également l'opportunité de réfléchir dans un cadre pluriannuel aux évolutions possibles du système d'assurance-maladie français.

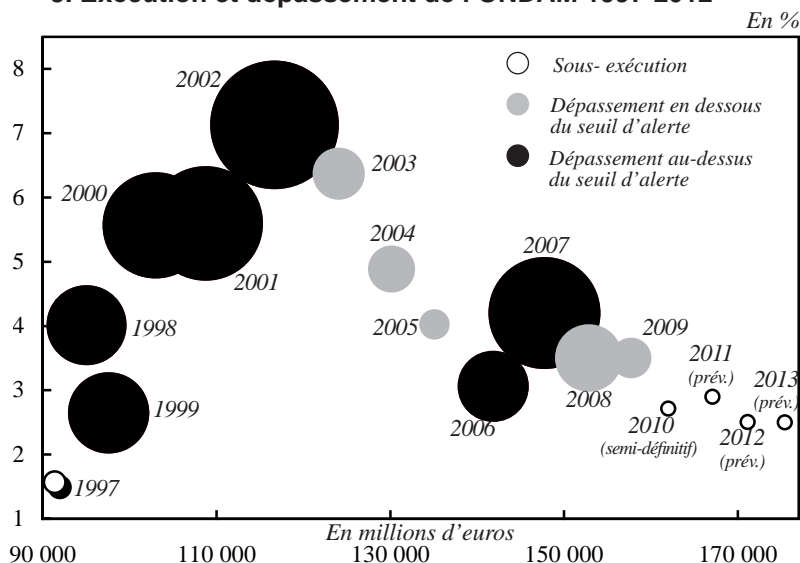
(3) Les autres mesures d'économies en dépense portent sur les dépenses de l'État, les retraites, les collectivités locales et les prestations familles et logement.

## 2. Répartition des économies dans l'ONDAM entre 2007 et 2012



Source : HCAAM, retraitement DG Trésor.

## 3. Exécution et dépassement de l'ONDAM 1997-2012



**Lecture :** En abscisses figure le niveau de dépenses constaté et en ordonnées le taux d'évolution associé ; la taille des bulles représente l'ampleur du dépassement ou surconsommation et de la sous-consommation. Ainsi, en 2007, les dépenses dans le champ de l'ONDAM ont atteint 147,7 milliards d'euros compte tenu d'un dépassement de 2,8 milliards d'euros, soit un taux d'évolution à périmètre constant de 4,2 %.

Source : Rapport Briet sur le pilotage des dépenses d'assurance-maladie (2010). Actualisation et retraitement DG Trésor.

## 1.2. L'ONDAM, vers une élaboration pluriannuelle

Depuis sa création en 1997, l'ONDAM a permis d'infléchir la dynamique des dépenses de santé à travers des mesures d'économies annuelles. Le Haut-Conseil pour l'avenir de l'Assurance-maladie (HCAAM) (2011) identifie trois grandes catégories de mesures d'économies classiquement utilisées pour assurer le respect de l'ONDAM : les baisse de prix et de tarifs (médicaments, radiologie et biologie), les actions de baisse de volumes (fraudes, maîtrise médicalisée comme la diffusion des génériques) et les modifications du remboursement (franchises médicaments, indemnités journalières...). Depuis 2007, les économies réalisées se répartissent ainsi : 35 % sur les prix, 50 % sur les volumes et 15 % sur les modifications de remboursement (cf. graphique 2).

De nombreuses dispositions réglementaires ont récemment été prises pour renforcer la gouvernance de l'ONDAM (création du comité d'alerte et renforcement de son rôle en 2011, délai d'entrée en vigueur différé de six mois des dispositions conventionnelles revalorisant les tarifs des professionnels libéraux en cas de procédure d'alerte en 2008, abaissement progressif du seuil d'alerte depuis 2011) et ont permis de faciliter son respect (cf. graphique 3). Par ailleurs, des pistes complémentaires avaient été envisagées dans le rapport Briet (2010) sur le pilotage des dépenses d'assurance-maladie afin de s'assurer du respect d'ONDAM plus stricts dans les prochaines années. En particulier, la régulation des dépenses devrait s'inscrire dans une logique pluriannuelle car l'impact des différentes mesures d'économies est dynamique et peut donc avoir un effet sur plusieurs années. C'est d'ailleurs ce que souligne le HCAAM (2011). La logique pluriannuelle permettrait en outre de tisser un lien plus fort entre la régulation des dépenses de santé via l'ONDAM et une politique globale de santé publique.

Dans ce contexte et afin de préciser la conduite des politiques publiques en France dans le domaine des finances publiques et des questions de santé, la DG Trésor s'est dotée d'un outil de microsimulation (ISIS, cf. annexe). En rapprochant les données de trois sources complémentaires<sup>(4)</sup>, la maquette décrit le profil complet d'environ 22 000 individus, correspondant à 8 000 ménages, représentatifs de la population générale de France métropolitaine vivant en ménage ordinaire<sup>(5)</sup> en 2006. Elle permet ainsi d'obtenir des éléments détaillés sur la consommation de soins de ville<sup>(6)</sup>, la couverture santé et le reste à charge des individus en fonction de leurs caractéristiques socio-démographiques, de leur activité professionnelle, de leur état de santé et des revenus du ménage. Cet outil permet ainsi d'étudier finement

(4) Enquête Santé et Protection sociale 2006 (IRDES), Échantillon permanent d'assurés sociaux 2006 (CNAMTS) et Enquête sur les contrats les plus souscrits auprès des complémentaires santé (DREES).

(5) C'est-à-dire à l'exclusion des personnes en foyers, institutions, services de long séjour hospitalier et des personnes sans domicile fixe. Cette population représente plus de 96 % de la population de France métropolitaine.

(6) Soins ambulatoires (consultations, auxiliaires, laboratoires) et biens médicaux (officines pharmaceutiques, optique, prothèses).

l'impact financier de diverses mesures portant sur le remboursement des soins des assurés (hausse du ticket modérateur, mise en place de franchise à la base...) notamment sur les assurés et l'assurance-maladie obligatoire. Il vise à renforcer l'information et la transparence sur l'impact des mesures.

Si la répartition future de l'effort reste conforme à ce qui a été observé dans le passé récent, sur les 13 milliards d'euros d'économies à réaliser entre 2013 et 2016, 2 milliards (15 %) pourraient résulter de modifications du système de prise en charge des soins. Au moyen de la maquette ISIS, différentes possibilités d'évolution du remboursement des soins sont simulées et analysées. La maquette ISIS ne couvre que les soins de ville (*cf.* annexe). En effet, les dépenses d'hospitalisation sont largement sous-estimées et mal individualisées dans la maquette, faute de sources disponibles suffisamment détaillées. À titre illustratif et sans préjuger des décisions futures dans la structure des efforts demandés, les différentes évolutions simulées ci-après permettent toutes d'enregistrer une économie globale d'un milliard d'euros sur les remboursements de soins de ville par l'assurance-maladie.

## **2. Quelles évolutions possibles de la prise en charge des assurés pour les soins de ville ? les différents scénarios envisagés**

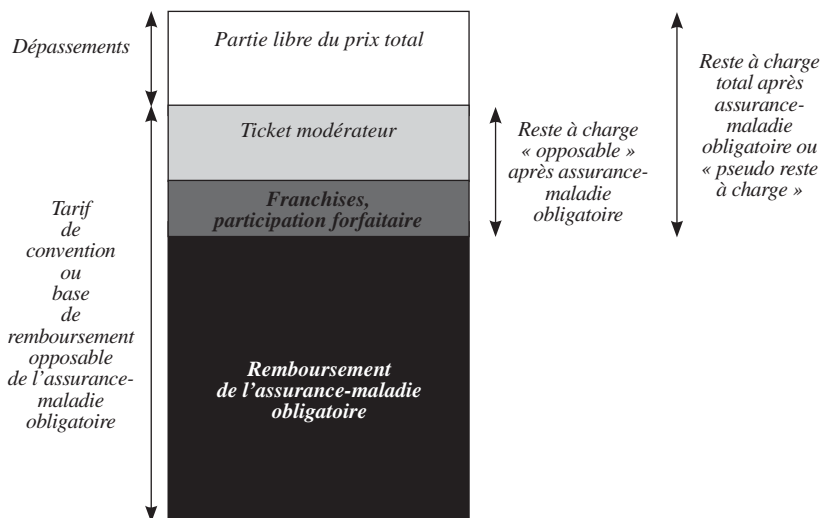
La prise en charge des soins en France s'articule autour de différents instruments de participation financière des patients. Pour les soins de ville, au ticket modérateur en vigueur depuis 1930, se sont ajoutées progressivement les participations financières de 1 euro (en 2005) sur toutes les consultations ou actes réalisés par un médecin, sur les examens radiologiques et sur les analyses de biologie médicale puis, en 2008, les franchises médicales sur les boîtes de médicaments (*cf.* graphique 4). Parallèlement, ont été mis en place différents dispositifs d'exonérations visant à atténuer les conséquences financières de ces mesures pour les populations les plus fragiles (affection de longue durée, ALD, couverture maladie universelle complémentaire, CMU-C<sup>(7)</sup>...).

Les différents scénarios envisagés de réforme de la participation financière vont de l'évolution paramétrique simple des dispositifs actuels jusqu'à une redéfinition complète de la couverture assurantielle obligatoire, en s'inspirant notamment des mécanismes mis en place dans d'autres pays européens. Six scénarios seront simulés ; une hausse de ticket modérateur, deux modalités de mise en place d'une franchise à la base et trois scénarios de mise en place d'une franchise, d'un ticket modérateur unique et d'un plafond de reste à charge dans la lignée des réflexions sur le bouclier sanitaire (Briet et Fragonard, 2007).

(7) Le seuil de revenu annuel pour bénéficier de la CMU complémentaire pour une personne seule est de 7 771 euros au 1<sup>er</sup> juillet 2011.



#### 4. Remboursement des soins, participation financière des patients et reste à charge<sup>(\*)</sup>



Note : (\*) Le reste à charge opposable après assurance-maladie obligatoire correspond à la différence entre le tarif de convention et le remboursement effectivement versé par l'assurance-maladie obligatoire. Le reste à charge total après assurance-maladie obligatoire correspond à la différence entre le tarif de convention auquel on ajoute la partie libre du prix et le remboursement par l'assurance-maladie obligatoire.

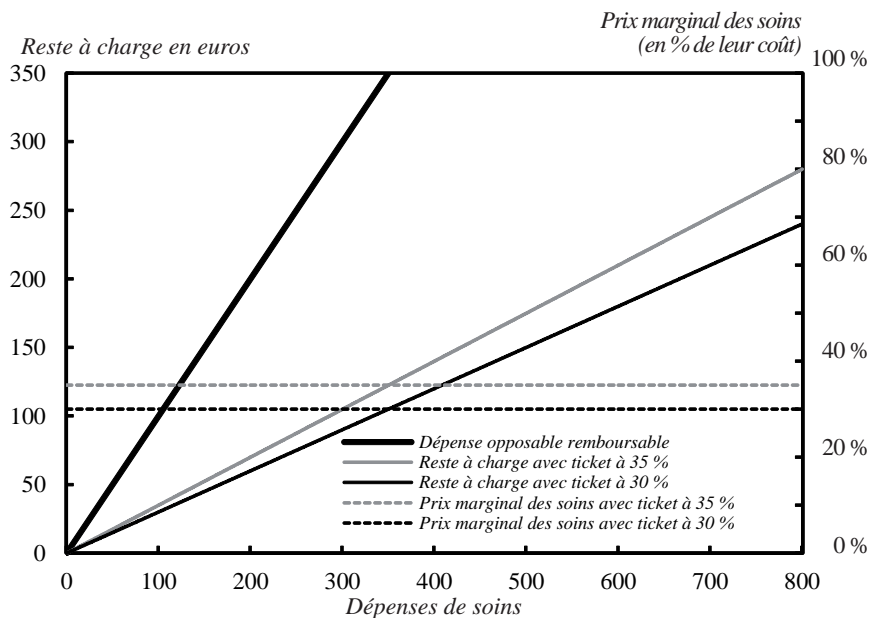
Sources : HCAAM et CNAMTS.

### 2.1. Augmentation du ticket modérateur

Le premier scénario simulé s'appuiera sur le ticket modérateur, principal mode<sup>(8)</sup> de participation financière des patients du système actuel (scénario 1). Pour tous les postes de consommation sur les soins de ville, le ticket modérateur est augmenté uniformément de façon à enregistrer une économie globale d'un milliard d'euros pour l'assurance-maladie. Une telle réforme s'inspire notamment de la loi Veil de 1993 qui décida une hausse générale de 5 points du ticket modérateur. Conformément à la pratique récente, cette hausse s'applique uniquement aux assurés de droit commun. Ainsi, les individus bénéficiant pour certaines dépenses d'une exonération du ticket modérateur la conservent pour ces soins.

(8) Le ticket modérateur représente environ 90 % du reste à charge opposable des assurés.

## 5. Hausse du ticket modérateur



*Lecture* : Pour une dépense de 200 euros, le reste à charge s'élève à 70 euros dans le cas d'un ticket modérateur à 35 %. Le ticket modérateur étant unique, le prix marginal des soins est égal à 35 % quel que soit le niveau de dépense.

Source : DG Trésor.

Après la loi Veil, les hausses du ticket modérateur ont été plus ciblées sur certains soins. Ces hausses ont, par exemple, pris la forme d'une augmentation du ticket modérateur de 5 points sur les médicaments à vignette bleue (en 2010, ticket modérateur passant de 65 à 70 %), ou encore sur les dispositifs médicaux (en 2010, ticket modérateur passant de 35 à 40 %). La modification du taux ne change pas en profondeur la philosophie du système de remboursement actuel. Évolution paramétrique à la marge, la hausse du ticket modérateur est simple à mettre en œuvre.

Pour autant, les hausses de ticket modérateur accentuent les limites inhérentes à ce mécanisme de participation financière. Geoffard (2006) souligne notamment que pour certaines personnes ayant besoin de recourir à des soins coûteux, un taux de ticket modérateur même modéré peut conduire à un reste à charge élevé. Comme le ticket modérateur est proportionnel à la dépense de soins, plus la dépense est élevée plus le reste à charge est élevé. Il exige donc une participation financière d'autant plus importante que l'état de santé est dégradé.

Certains pays européens ont opté pour d'autres instruments de participation financière. La microsimulation permet de quantifier les conséquences financières de la mise en place de tels instruments dans le système de remboursement des soins français.

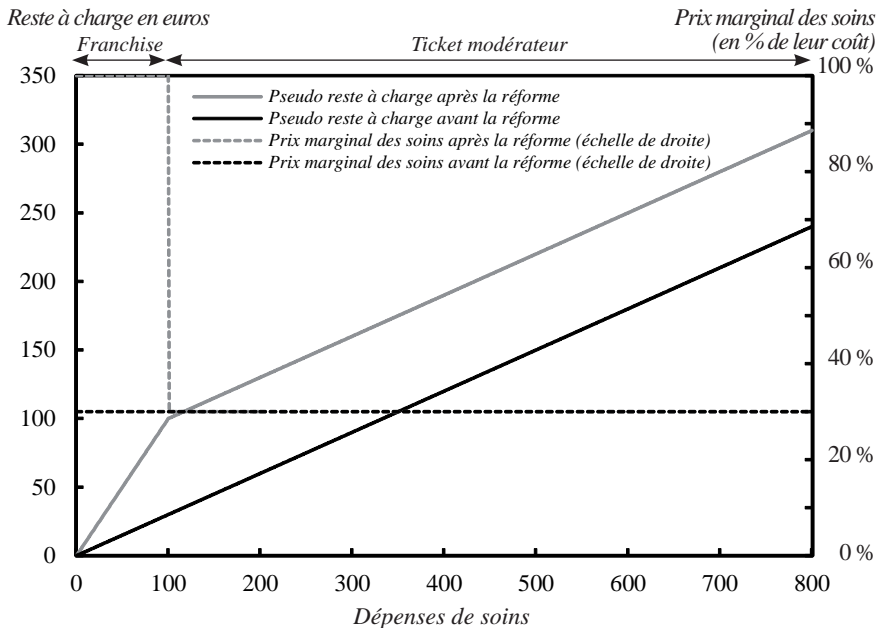
## 2.2. Exemples d'instruments de participations financières européens

### 2.2.1. Instauration d'une franchise à la base (Suisse, Pays-Bas, Suède)

La mise en place d'une franchise décale l'intervention du mécanisme de couverture assurantielle. Le remboursement des dépenses de santé n'intervient qu'à partir du moment où les dépenses cumulées des patients depuis le début de l'année, ou la période de référence choisie, dépassent le montant de la franchise. En deçà de ce montant, les patients supportent la totalité de leurs dépenses. Les franchises à la base sont le plus souvent définies sur des périodes de douze mois (non glissantes).

Certains pays européens ont opté pour ce mécanisme de participation financière et ce, quel que soit le système de prise en charge des soins, public ou privé. Les assurés suisses et néerlandais, qui bénéficient d'un système d'assurance-maladie concurrentiel géré par des caisses d'assurances privées, se voient offrir des contrats avec, pour les premiers, une franchise variant d'environ 180 à 1 500 euros et, pour les seconds, une franchise variant de 0 à 500 euros. Les assurés suédois, couverts par un système national de santé financé par impôt, doivent, pour leur part, s'acquitter d'une franchise annuelle à la base de 100 euros sur leur dépense de médicaments.

### 6. Mise en place d'une franchise à la base



**Lecture :** Avant la réforme, le patient ne doit payer que le ticket modérateur. La réforme consiste à mettre en place une franchise à la base en plus du ticket modérateur. Dans cet exemple, la franchise est fixée à 100 euros. Lorsque la dépense est inférieure à la franchise, le reste à charge est égal à la dépense et le prix marginal des soins est donc égal à 100 %. Lorsque la dépense de soins dépasse la franchise, le ticket modérateur s'applique à la différence entre la dépense de soins et la franchise. Le reste à charge est alors égal au montant de la franchise auquel on ajoute le ticket modérateur. Le prix marginal des soins est alors égal au ticket modérateur (30 % dans cet exemple).

**Source :** DG Trésor.

L'une des principales critiques formulées à l'encontre de la franchise se fonde sur le lien présumé entre niveau de franchise et renoncement aux soins. Une franchise trop élevée peut entraîner des problèmes d'accès aux soins notamment pour les assurés aux revenus modestes. Un recours tardif aux soins peut également être à l'origine de complications médicales, entraînant *in fine* un surcoût de traitement. À l'inverse, la franchise est une incitation financière qui sensibilise le patient au coût total des soins et peut être à l'origine de changements de comportement. De plus, l'effet incitatif joue à plein car la franchise se concentre sur les petites dépenses où l'élasticité de la demande de soins est la plus forte. Elle permet ainsi, théoriquement, un recours aux soins plus efficace (Newhouse, 1993). Parallèlement, l'effort maximum demandé est plafonné au montant de la franchise. Elle ne demande pas, tant que son montant reste modéré, un effort d'autant plus fort que l'état de santé de l'assuré est dégradé.

Deux scénarios introduisant des franchises seront simulés dans le présent complément, à savoir le cas où une telle franchise s'applique à l'ensemble de la dépense de soins (scénario 2) et celui où elle s'applique seulement aux dépenses actuellement non exonérées de participation financière (scénario 3). En s'alignant sur les modalités actuelles de prise en charge des soins, la franchise est supposée être à la charge exclusive du patient, au même titre que la participation forfaitaire ou les franchises médicales (pas de prise en charge par les mutuelles).

### 2.2.2. *Instauration d'un plafond de reste à charge (Allemagne, Suisse, Suède, Belgique)*

Un plafond de dépense est défini généralement sur une base annuelle calendaire (non glissante), en fonction de la dépense remboursable (dépense engagée dans la limite du tarif de référence) ou du ticket modérateur. Au-delà de ce plafond, l'assurance prend en charge la totalité du coût des soins, toujours dans la limite de la dépense remboursable. La participation financière marginale de l'assuré est alors nulle.

Ce dispositif est utilisé sous différentes formes dans de nombreux pays européens. Son niveau peut tout d'abord dépendre du revenu. Lorsque le reste à charge dépasse une certaine part du revenu de l'assuré ou du ménage, l'assurance-maladie prend en charge la totalité de ses dépenses supplémentaires de soins. En Allemagne, le reste à charge ne peut représenter plus de 2 % du revenu annuel du ménage. Ce seuil est de 1 % pour les assurés atteints d'une affection de longue durée ou les personnes âgées dépendantes. Un plafond indépendant du revenu existe également dans certains pays européens. En Suisse, en fonction du contrat et de la franchise choisis, le plafond varie de 600 à 2 000 euros. Le plafond peut, enfin, varier en fonction du type de soins. En Suède, le plafond de reste à charge est fixé à 100 euros pour les consultations, à 200 euros pour les médicaments et 222 euros pour les dispositifs médicaux. Ces plafonds ont pour but de prévenir les difficultés financières d'un assuré confronté à un risque catastrophique de santé, c'est-à-dire un risque à faible probabilité mais qui entraîne de fortes pertes.

La France a opté pour un système différent. Afin d'atténuer les conséquences financières des instruments de participation des patients à leur dépense de soins, différents dispositifs d'exonération et de plafonnement ont été mis en place. Le plus important, en nombre d'assurés bénéficiaires comme en montant de prise en charge, est celui des affections de longue durée (ALD) qui exonère du ticket modérateur les dépenses de soins liées à la maladie exonérante<sup>(9)</sup>, dite ALD. Le nombre de patients exonérés en ALD a augmenté de 4 % en 2010, dans la poursuite de la tendance antérieure (entre 2005 et 2010 le taux de croissance annuel a été compris entre 3,5 et 4,2 %). Environ 9 millions de personnes bénéficiaient de ce dispositif en 2010 soit 15,5 % de la population<sup>(10)</sup>. Les remboursements destinés aux bénéficiaires en affection de longue durée représentaient, en 2008, 62 % du total des remboursements du régime général d'assurance-maladie (cf. Baillet, 2010). La CNAMTS prévoit une croissance dynamique du nombre de bénéficiaires en ALD qui devrait atteindre 17 % des assurés en 2014. À côté de ce dispositif, la Cour des comptes totalise une trentaine de motifs d'exonération pour le seul ticket modérateur (les invalides, les femmes enceintes, les nouveau-nés...).

Ces exonérations présentent deux limites principales. D'une part, du fait de leur foisonnement, elles rendent complexe et peu lisible notre système de remboursement des soins. Cette « illisibilité [apparaît] contradictoire avec l'idée affichée de vouloir responsabiliser les patients » par le biais de la participation financière (Elbaum, 2008). D'autre part, malgré la multitude de motifs d'exonération, les assurés ne sont pas forcément bien couverts contre les risques catastrophiques. Comme le souligne le HCAAM (2011), « si une grande majorité des assurés ont un reste à charge après assurance-maladie obligatoire relativement peu élevé, il apparaît en revanche très important pour une minorité significative des assurés » (cf. tableau 1).

### 1. Reste à charge annuel total moyen après AMO pour les soins de ville et à l'hôpital selon les déciles de reste à charge

*En euros*

	Soins de ville	Hôpital	Ensemble
Premier décile de reste à charge (D1)	18	0	18
Médiane (D5)	219	7	227
Dixième décile de reste à charge (D10)	1 797	286	2 084
Dernier centile de reste à charge (C100)	3 982	946	4 928
Ensemble	454	44	498

*Lecture* : D10 : les 10 % des assurés qui ont le plus fort reste à charge ont un reste à charge annuel total moyen après AMO (assurance maladie obligatoire) de 1797 euros pour les soins de ville et de 286 euros pour les soins hospitaliers ; Hôpital correspond aux activités de médecine, chirurgie, obstétrique (MCO).

*Source* : HCAAM à partir des données de la CNAMTS, ATIH, EGB 2008.

(9) Ce dispositif s'applique à trente affections énumérées (diabète, VIH, cancer...) auxquelles s'ajoutent des affections évolutives ou invalidantes « hors liste » (31°) et des poly-pathologies invalidantes (32°).

(10) Cf. Proposition de l'assurance-maladie sur les charges et produits pour l'année 2012, CNAMTS, 2011. <http://www.annuaire-secu.com/pdf/Propositions-AM-charges-produits2012.pdf>

Ces deux limites ont contribué à développer la réflexion sur une réforme plus ambitieuse de la participation financière de l'assuré en France. Elles ont notamment abouti aux travaux menés sur le bouclier sanitaire (Briet et Fragonard, 2007). Dans la lignée de ces travaux sur le bouclier sanitaire et des modèles de couverture du risque maladie en Europe, nous simulerons trois réformes systémiques dans lesquelles toutes les exonérations actuelles sont supprimées et où sont alors mis en place :

- une franchise, un ticket modérateur unique de 30 % dans un souci de simplification du système de remboursement (comme le suggère le rapport sur le bouclier sanitaire) et un plafond de reste à charge opposable fixe comme en Suisse (scénario 4) ;
- une franchise, un ticket modérateur unique de 30 % et un plafond de reste à charge opposable en proportion du revenu par unité de consommation<sup>(11)</sup> (scénario 5) ;
- une franchise, un ticket modérateur unique de 30 % et un plafond de reste à charge opposable en proportion du revenu mais différencié en fonction des pathologies. Dans ce scénario, à l'image du système d'assurance allemand où le plafond est de 2 % du revenu pour les assurés de droit commun et de 1 % du revenu pour les assurés souffrant d'une affection de longue durée, le plafond de reste à charge est moitié moindre pour les assurés souffrant d'une affection de longue durée (scénario 6).

Les six scénarios retenus seront calibrés, simulés et analysés à l'aide de la maquette ISIS. Cette analyse permet de quantifier leurs avantages et leurs inconvénients au regard de différents critères socio-économiques.

### 3. Résultats des simulations

#### 3.1. Résultats généraux

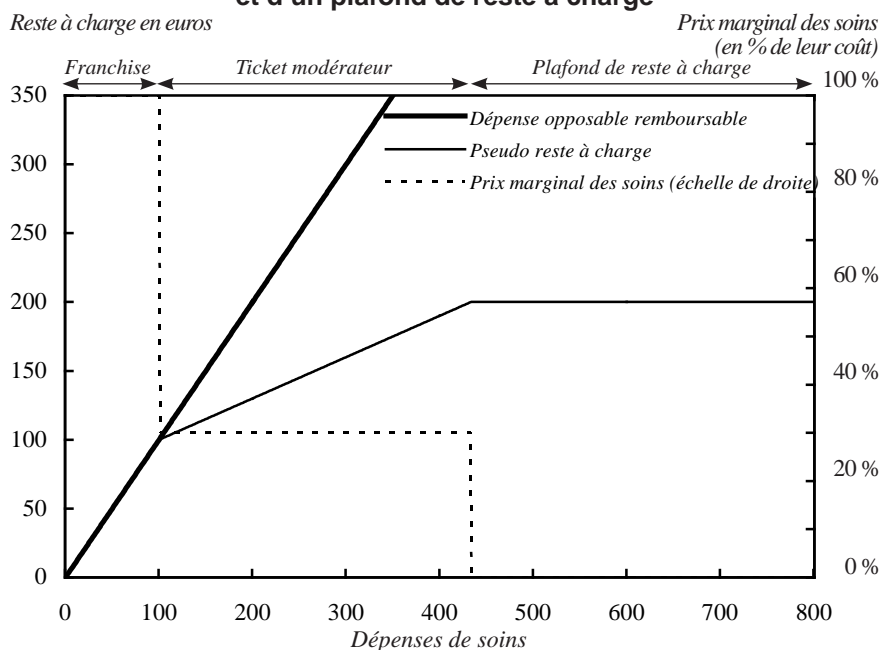
Chacun des six scénarios est calibré de telle façon qu'il permette une économie d'un milliard d'euros sur l'ensemble des soins de ville.

Dans tous les cas, les bénéficiaires de la Couverture maladie universelle complémentaire (CMU-C) restent couverts intégralement et ne sont soumis à aucune modification du système de remboursement, comme l'envisageait le rapport sur le bouclier sanitaire (Briet et Fragonard, 2007) pour préserver cette population fragile. Enfin, la participation financière du patient reste individualisée. Il serait en effet possible de calculer les différentes participations financières au niveau du ménage. Pour autant, ce mode de calcul peut engendrer des comportements opportunistes de surconsommation des soins (Newhouse, 1993) notamment pour les proches d'une personne soumise à d'importantes dépenses de soins.

(11) Le revenu par unité de consommation consiste à rapporter le revenu total du ménage à sa taille, qui, en raison d'économies d'échelle sur de nombreuses dépenses (logement, transports), évolue moins que proportionnellement au nombre de membres le composant.

Pour les trois derniers scénarios qui comportent des changements plus importants, certaines hypothèses ont été posées et pourraient être modulées. Dans un premier temps, on calcule à partir de la base de remboursement (cf. graphique 7) le nouveau remboursement par l'assurance-maladie induit par la mise en place d'une franchise de 40 euros<sup>(12)</sup>, d'un ticket modérateur unique de 30 %<sup>(13)</sup> et la suppression de l'ensemble des exonérations. Une fois ces changements simulés, la baisse des dépenses de l'assurance-maladie est quantifiée. Dans un second temps, le niveau d'un plafond de reste à charge, qui constitue un coût pour l'assurance-maladie, est déterminé de façon à enregistrer une économie finale d'un milliard d'euros.

### 7. Mise en place d'une franchise, d'un ticket modérateur unique et d'un plafond de reste à charge



*Lecture :* Dans cet exemple, la franchise est fixée à 100 euros, le ticket modérateur à 30 % et le plafond de reste à charge à 200 euros. Lorsque la dépense est inférieure à la franchise, le reste à charge est égale à la dépense et le prix marginal des soins est donc égal à 100 %. Lorsque la dépense de soins dépasse la franchise, le ticket modérateur s'applique à la différence entre la dépense de soins et la franchise. Le reste à charge est alors égal au montant de la franchise auquel on ajoute le ticket modérateur. Le prix marginal des soins est alors égal au ticket modérateur (dans cet exemple, 30 %). Lorsque la dépense atteint le plafond, le reste à charge est égal au plafond quelle que soit la dépense et le prix marginal des soins est donc de 0 %.

*Source :* DG Trésor.

(12) Le montant de la franchise a été fixé afin de s'aligner sur le montant du co-paiement demandé en Allemagne pour un recours aux soins réguliers sur une année.

(13) La seule mise en place d'un ticket modérateur unique à 30 % pour les dépenses non exonérées représente une perte pour l'Assurance-maladie. En effet, le ticket modérateur moyen est actuellement de 32 % en moyenne pour ces dépenses sur les soins de ville.

## 2. Récapitulatif des différents scénarios simulés

	Description succincte du scénario	Nom du scénario dans les résultats
Scénario 1	Augmentation générale du ticket modérateur sur l'ensemble des soins de ville	TM
Scénario 2	Mise en place d'une franchise sur l'ensemble des dépenses de soins de ville y compris sur les dépenses exonérées de ticket modérateur	FRENS
Scénario 3	Mise en place d'une franchise sur l'ensemble des dépenses de soins de ville à l'exception des dépenses exonérées de ticket modérateur	FREXO
Scénario 4	Restauration d'une franchise, un ticket modérateur unique et un plafond de reste à charge fixe comme en Suisse	PLFIX
Scénario 5	Restauration d'une franchise, un ticket modérateur unique et un plafond de reste à charge en proportion du revenu	PLREV
Scénario 6	Restauration d'une franchise, un ticket modérateur unique et un plafond de reste à charge en proportion du revenu et moitié moindre pour les assurés souffrant d'une affection de longue durée comme en Allemagne	PLALD

Source : DG Trésor.



Le tableau 3 présente les résultats de ces étalonnages. 1 milliard d'euros d'économies nécessite ainsi une hausse uniforme de 2,3 points du ticket modérateur sur les dépenses non exonérées ou la mise en place d'une franchise sur l'ensemble des dépenses de 19 euros (20 euros dans le cas du scénario 3 FREXO s'appliquant aux seules dépenses non exonérées). Avec une franchise de 40 euros et un ticket modérateur unique de 30 %, le même rendement est obtenu avec un plafond de reste à charge fixe d'environ 560 euros pour le scénario calqué sur le modèle Suisse (PLFIX), 4 % du revenu par unité de consommation<sup>(14)</sup> lorsque le plafond est proportionnel au revenu. Dans le scénario d'un plafond différencié en fonction de l'état de santé, il s'élève à 6,2 % du revenu par unité de consommation (UC) pour les assurés de droit commun et à 3,1 % pour les patients atteints d'une ALD. Le calcul du plafond à partir du revenu par unité de consommation permet d'homogénéiser la prise en charge des ménages de tailles ou de compositions différentes. La prise en compte du revenu par unité de consommation pour le calcul d'un droit serait une nouveauté dans le système actuel de prestations sociales.

### 3. Les différents scénarios pour économiser 1 milliard d'euros sur les remboursements des soins de ville

	Définition des scénarios			
	Hausse du taux	Franchise	Plafond	
	En points	En euros/an	En euros/an	%REV/UC
TM	+ 2,3	—	—	—
FRENS	0	19	—	—
FREXO	0	20	—	—
PLFIX	—	40	560	—
PLREV	—	40	—	4
PLALD	—	40	—	6,2 (patients non-ALD) 3,1 (patients ALD)

*Lecture* : Le montant très proche des franchises dans le cas des scénarios FREXO et FRENS s'explique par un montant de dépenses non exonérées qui dépasse 20 euros pour plus de 85 % des assurés qui bénéficient d'une exonération de ticket modérateur. Ainsi, ces assurés même s'ils sont exonérés de ticket modérateur pour une partie de leur dépense s'acquitteront de la franchise sur leur dépense non exonérée.

*Source* : DG Trésor, Maquette ISIS.

(14) Nous avons choisi l'échelle actuellement la plus utilisée (dite de l'OCDE) qui retient la pondération suivante :

- 1 UC pour le premier adulte du ménage ;
- 0,5 UC pour les autres personnes de 14 ans ou plus ;
- 0,3 UC pour les enfants de moins de 14 ans.

## Effet des changements de comportement suscités par les différentes réformes

Chaque scénario est étalonné afin d'enregistrer une économie pour l'assurance-maladie obligatoire d'un milliard d'euros. Cependant, cette économie pour l'assurance-maladie constitue une diminution des remboursements pour les assurés et donc un renchérissement du coût apparent des soins. Comme l'ont montré de nombreuses études (par exemple, Newhouse, 1993), le changement de prix des soins peut induire une modification du comportement de consommation des assurés. Afin d'appréhender l'impact des différentes mesures étudiées sur la décision de consommation des assurés, les changements de comportement des assurés sont intégrés à la maquette ISIS.

Ces simulations reposent sur différentes hypothèses. Tout d'abord, il est nécessaire de calibrer la réaction des agents confrontés à une modification du prix des soins. Le spectre d'estimation de l'élasticité-prix de la demande de soins reste large autour d'une valeur centrale de  $-0,2$  (Newhouse, 1993) c'est-à-dire que pour une augmentation de 10 % du prix des soins on constate une diminution de la demande de soins de 2 %. Afin de prendre en compte les divergences d'estimation et une possible adaptation partielle des comportements juste après la réforme, deux hypothèses d'élasticité de  $-0,1$  et  $-0,2$  sont retenues. Ensuite, il convient de déterminer l'augmentation du prix des soins. Dans la lignée des travaux d'Eichner (1998) qui utilisent le prix constaté de fin d'année pour l'étude de l'élasticité-prix de la demande de soins aux Pays-Bas, nous considérons que le prix des soins est égal au reste à charge opposable annuel après assurance-maladie obligatoire. L'augmentation du prix des soins est donc égale à la variation du niveau de ce reste à charge. Une dernière hypothèse sur les modalités de couverture par l'assurance-maladie complémentaire est nécessaire. En s'alignant sur les modalités actuelles de prise en charge des soins, le ticket modérateur est considéré comme couvert et mutualisé par les assurances complémentaires alors que la franchise est, au même titre que la participation forfaitaire ou les franchises médicales, à la charge exclusive du patient. Ainsi, pour les assurés qui bénéficient d'une assurance complémentaire, l'augmentation du ticket modérateur n'aura pas d'effet direct sur leur comportement de consommation de soins au contraire d'une franchise non réassurable. En effet, leur reste à charge final hors prime d'assurance ne sera pas modifié par une hausse du ticket modérateur. Pour les scénarios avec franchise et plafond, les changements de prix sont plus complexes à appréhender. Nous considérons donc que, pour les assurés qui bénéficient d'une assurance complémentaire, seule la mise en place de la franchise augmente le prix des soins et influence donc leur comportement. Le plafond de reste à charge se substitue à la prise en charge des organismes complémentaires. Pour les assurés sans couverture complémentaire, l'évolution de leur reste à charge opposable correspondra à l'augmentation du coût des soins.

## 1. Modification du comportement de consommation et d'assurance des assurés sociaux

	Comportement de consommation (pris en compte dans la maquette)	Comportement d'assurance (non pris en compte dans la maquette)
Assuré social affilié à un organisme complémentaire	<ul style="list-style-type: none"> <li>• Ticket modérateur réassurable</li> <li>• Pas de modification du comportement de consommation</li> <li>• Franchise</li> <li>• Diminution de la demande de soins</li> </ul>	<ul style="list-style-type: none"> <li>• Augmentation des primes</li> <li>• Diminution de la demande d'assurance complémentaire.</li> <li>• Diminution de la couverture publique</li> <li>• Augmentation de la demande d'assurance complémentaire</li> </ul>
Assuré social uniquement couvert par l'assurance-maladie obligatoire	<ul style="list-style-type: none"> <li>• Augmentation du ticket modérateur</li> <li>• Diminution de la demande de soins</li> <li>• Franchise</li> <li>• Diminution de la demande de soins</li> </ul>	<ul style="list-style-type: none"> <li>• Augmentation des primes</li> <li>• Désincitation plus importante à s'affilier auprès d'un organisme complémentaire</li> <li>• Diminution de la couverture publique</li> <li>• Augmentation de la demande d'assurance complémentaire</li> </ul>

Source : DG Trésor.

Pour les assurés bénéficiant d'une assurance complémentaire, la hausse des primes relative à un transfert mutualisé de la prise en charge de l'assurance-maladie obligatoire vers l'assurance-maladie complémentaire pourrait entraîner une modification du comportement d'assurance. Cet effet n'est pas pris en compte dans ISIS. Seuls sont décrits les changements de consommation de soins des assurés liés directement à la modification du coût des soins. Les différents effets de comportement sont résumés dans le tableau 1.

La hausse des prix apparents induite par les réformes simulées du système de participation financière induit une baisse de la demande de consommation de soins. Les différentes réformes, dont les paramètres ont été calibrés sur une réduction des dépenses avant effets comportementaux d'un milliard d'euros pour l'assurance-maladie, engendrent finalement des baisses plus fortes, une fois pris en compte les changements de comportement de consommation des assurés. Les résultats des simulations intégrant les changements de comportement des assurés sont résumés dans le tableau 2. Une hausse uniforme de ticket modérateur aurait peu d'impact sur le comportement des assurés. Ceci explique les faibles économies supplémentaires pour l'assurance-maladie publique. Seuls 7 % de la population ne bénéficient pas d'une assurance complémentaire et changeraient donc de comportement en cas de hausse du ticket modérateur. De plus, leur dépense est plus faible que celle des assurés avec une couverture complémentaire (Geoffard, 2011). La franchise a un effet marqué sur le comportement de consommation de soins (*cf.* tableau 2) car,

selon les hypothèses retenues, elle ne serait pas couverte et mutualisée par les assureurs complémentaires. Lorsque les changements de comportement sont pris en compte, la différence de rendement entre le scénario 2 FRENS et le scénario 3 FREXO s'explique par la suppression des exonérations dans le cas du scénario 2 FRENS. Très faible dans le cas des exonérations, le prix apparent bondit avec la mise en place d'une franchise ce qui a pour effet de réduire l'ensemble de la demande de soins (par nature élevée) des personnes souffrant d'une ALD. Le fort rendement des mesures comprenant la mise en place d'un plafond de reste à charge est principalement lié au niveau élevé de la franchise (40 euros).

## 2. Baisse des dépenses d'assurance-maladie en fonction des hypothèses de changements de comportement

*En milliards d'euros*

	Gain en fonction des changements de comportement		
	Élasticité nulle	Élasticité de - 0,1	Élasticité de - 0,2
TM	1	1,03	1,06
FRENS	1	1,50	1,90
FREXO	1	1,20	1,35
PLFIX	1	1,70	2,20
PLREV	1	1,60	2,10
PLALD	1	1,60	2,10

*Lecture :* Dans le cas du scénario 1 TM, la hausse du ticket modérateur s'accompagne d'une baisse d'un milliard d'euros des dépenses d'assurance-maladie sans changement de comportement, elle s'élève à 1,03 milliard d'euros en prenant en compte les changements de comportement avec une élasticité-prix de la demande de soins de - 0,1 et 1,06 milliard d'euros avec une élasticité-prix de - 0,2.

*Source :* DG Trésor, Maquette ISIS.

Ces résultats appellent principalement deux remarques. Tout d'abord, comme il n'est pas possible d'estimer l'élasticité-prix de la demande de soins, une élasticité unique est assignée à l'ensemble des assurés. Pour autant, différentes études empiriques ont montré la diversité des élasticités entre les types de soins et les catégories d'assurés. On constate par exemple dans les simulations de forts changements de comportement de consommation des assurés ALD qui avaient avant la réforme des restes à charge très faibles. L'estimation différenciée de l'élasticité-prix des assurés en ALD aurait permis de constater si les assurés souffrant de maladies chroniques ont une demande de soins élastique alors que leur traitement est souvent vital. Une telle estimation aurait également permis de mettre en avant d'éventuels effets de seuil pour l'élasticité-prix de la demande de soins en fonction du niveau de consommation de soins voire du niveau de revenu. Ensuite, la relation non linéaire des résultats entre une élasticité de - 0,1 et - 0,2 s'explique principalement par le moindre rendement de la franchise du fait de la diminution de la consommation de soins. En effet, en fonction du niveau d'élasticité, un nombre croissant d'assurés qui avaient des remboursements au-dessus de la franchise a, du fait du changement de comportement, une consommation qui devient inférieure à la franchise.

## 3.2. Résultats microéconomiques

Cette partie analyse les principales conséquences des différents scénarios sur la prise en charge des assurés. Dans un souci de clarté, seules les différences significatives entre scénarios sans modification de comportement sont présentées.

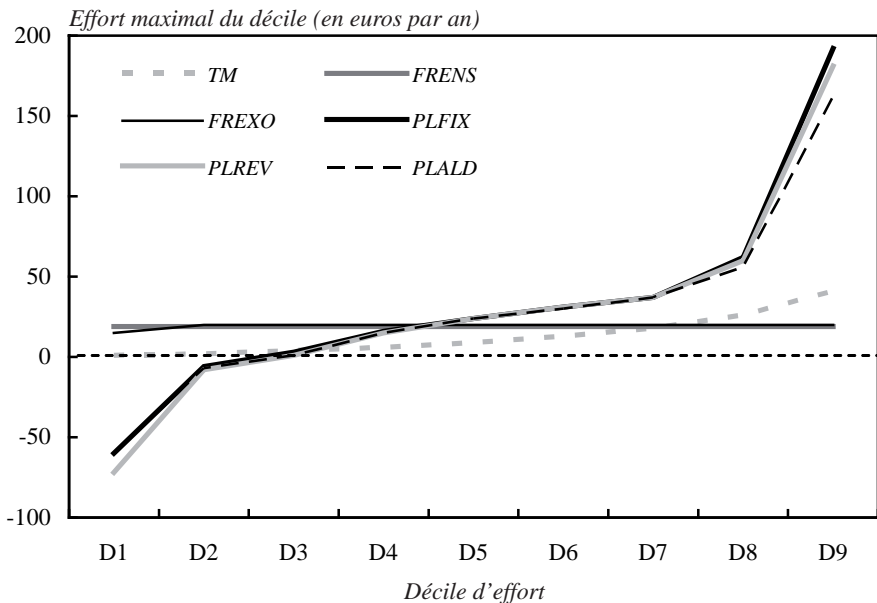
### 3.2.1. Effort des assurés

On définit pour chaque assuré l'effort comme étant la variation du reste à charge par rapport à la situation actuelle et après modification des mécanismes de participation financière qui conduisent ici à une diminution globale des remboursements de l'assurance-maladie d'un milliard d'euros.

#### 3.2.1.1. Une forte disparité des efforts individuels pour chaque scénario sauf dans le cas de la mise en place d'une franchise

L'effort moyen sur l'ensemble de la population concernée par la modification du système de remboursement s'élève à 18 euros, avec cependant des profils très différents selon la réforme (graphique 8).

## 8. Distribution de l'effort dans les six scénarios



*Lecture* : Dans le cas du scénario TM, 10 % de la population (décile d'effort, D9) fournissent un effort supérieur à 41 euros suite à la modification du système de remboursement.

*Source* : DG Trésor, Maquette ISIS.

Une analyse comparée des déciles d'effort permet de constater que les scénarios avec une franchise demandent un effort réparti uniformément sur l'ensemble de la population (environ 19 euros) alors que l'effort dans le scénario de hausse du ticket modérateur est plus concentré sur certains assurés. Ainsi, 50 % des assurés ont une variation du reste à charge inférieure à 9 euros alors que 10 % de la population ont un effort supérieur à 41 euros. Le rapport interdécile (D9/D1) est très élevé (40) pour le scénario TM alors que, dans le cas des scénarios comportant une franchise, ce rapport est unitaire. L'unité de ce ratio signifie que la quasi-totalité des assurés fournissent un effort équivalent en niveau.

Les scénarios dans lesquels seule une franchise est mise en place peuvent donc être assimilés à l'instauration d'un prélèvement forfaitaire sur l'ensemble des consommateurs. En effet, la quasi-totalité des consommateurs bénéficient aujourd'hui d'un remboursement par l'assurance-maladie qui dépasse le montant retenu de la franchise. Au contraire, les autres scénarios sont à l'origine de pertes ou de gains qui dépendent des caractéristiques du patient et notamment de sa consommation de soins.

Les scénarios avec un plafond de reste à charge sont caractérisés par un spectre des efforts encore plus large et présentent la particularité de créer des gagnants. Dans ces scénarios, environ 22 % de la population<sup>(15)</sup> y gagnent, c'est-à-dire que leur reste à charge après la réforme est plus faible que celui observé dans la situation actuelle. Dans le cas du scénario avec plafond de reste à charge fixe (PLFIX), 10 % des assurés gagnent ainsi plus de 60 euros par an (cf. graphique 8). Entre 25 % des ALD pour le scénario PLALD et 19 % pour le scénario PLFIX profiteront de la mise en place de ces nouveaux systèmes de remboursement. Lorsque celui-ci est défini en fonction du revenu (scénario 5 PLREV, scénario 6 PLALD), 35 % des individus appartenant à un ménage du premier quartile et 25 % de ceux du deuxième quartile y gagnent. 35 % (scénario 6 PLALD), 32 % (scénario 5 PLREV) et 27 % (scénario 4 PLFIX) des personnes se déclarant en mauvais ou très mauvais état de santé profitent également de cette réforme du système de remboursement.

L'analyse de l'effort par sous-population, et notamment selon le statut en ALD, permet de juger de l'effort demandé en fonction des caractéristiques individuelles des consommateurs.

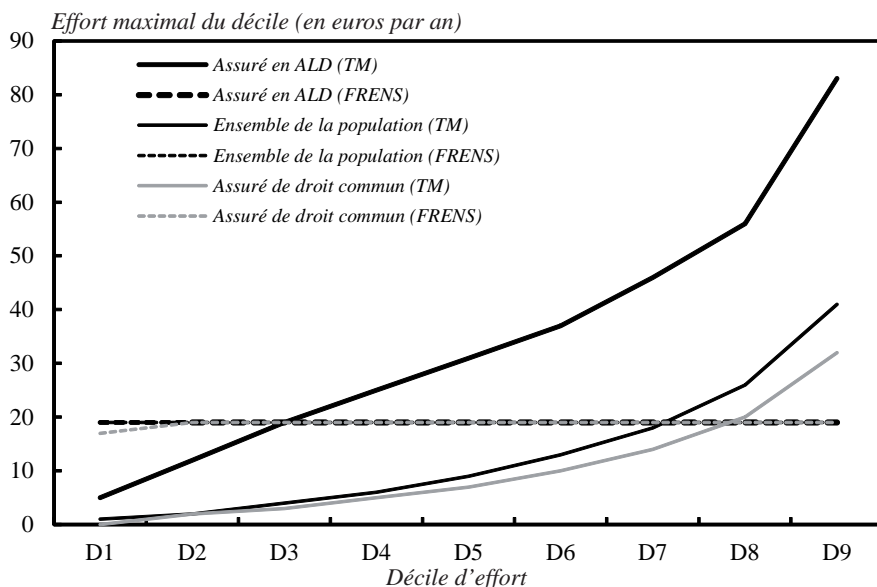
### 3.2.1.2. L'instauration d'une franchise n'engendre pas de discrimination en fonction du statut ALD, au contraire des autres scénarios

L'exonération du ticket modérateur pour cause d'affection de longue durée est un indicateur qui permet de rendre compte à la fois de l'état de santé du patient mais également de son besoin de soins. On compare les différents profils d'effort pour la population bénéficiant de l'exonération au titre d'une ALD selon les différents scénarios.

(15) 22,5 % dans le cas du scénario PLFIX, 23,3 % dans le cas du scénario PLREV et 23,2 % dans le cas du scénario PLALD.

La mise en place d'une franchise n'introduit pas de discrimination dans l'effort en fonction du statut en ALD, au contraire du scénario TM (*cf.* graphique 9). Dans le cas des scénarios de franchise (scénario 2 FRENS, scénario 3 FREXO), la variation moyenne du reste à charge est identique pour les ALD et les non-ALD s'élevant à 18 euros. Dans le scénario 1 TM, la différence d'effort entre ces populations apparaît en revanche significative, l'effort moyen pour les ALD représente plus de trois fois celui des assurés de droit commun (45 euros contre 14). Les pertes individuelles consécutives à une hausse du ticket modérateur peuvent également s'avérer importantes pour les assurés en ALD puisque 10 % d'entre eux voient leur reste à charge augmenter d'au moins 83 euros contre seulement 32 euros pour les 10 % des assurés non-ALD perdant le plus à une telle réforme (*cf.* graphique 9).

### 9. Distribution de l'effort selon le statut en ALD (TM, FRENS)

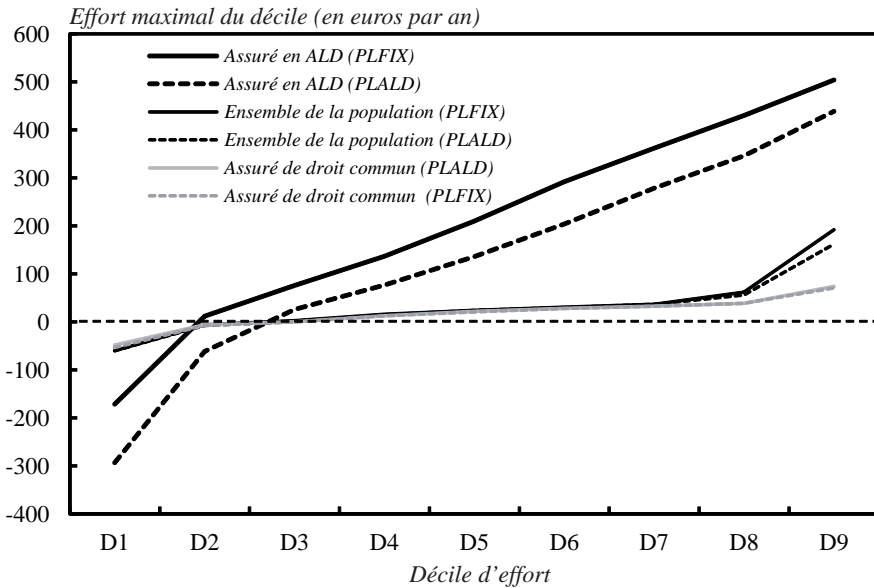


*Lecture :* Dans le cas du scénario TM, 10 % de la population non-ALD fournissent un effort supérieur à 32 euros suite à la modification du système de remboursement.

*Source :* DG Trésor, Maquette ISIS.

Dans le cas des scénarios avec plafond de reste à charge comme dans le scénario de hausse du ticket modérateur, l'effort demandé aux assurés en ALD est significativement supérieur à celui des autres assurés (*cf.* graphique 10). La suppression des exonérations explique ce supplément d'effort pour les assurés souffrant d'une ALD. La modulation du plafond en fonction du statut en ALD dans le scénario 6 PLALD permet de diminuer l'effort qui porte sur les ALD par rapport aux autres scénarios dans lesquels le plafond est identique pour l'ensemble de la population (*cf.* graphique 10). Parallèlement

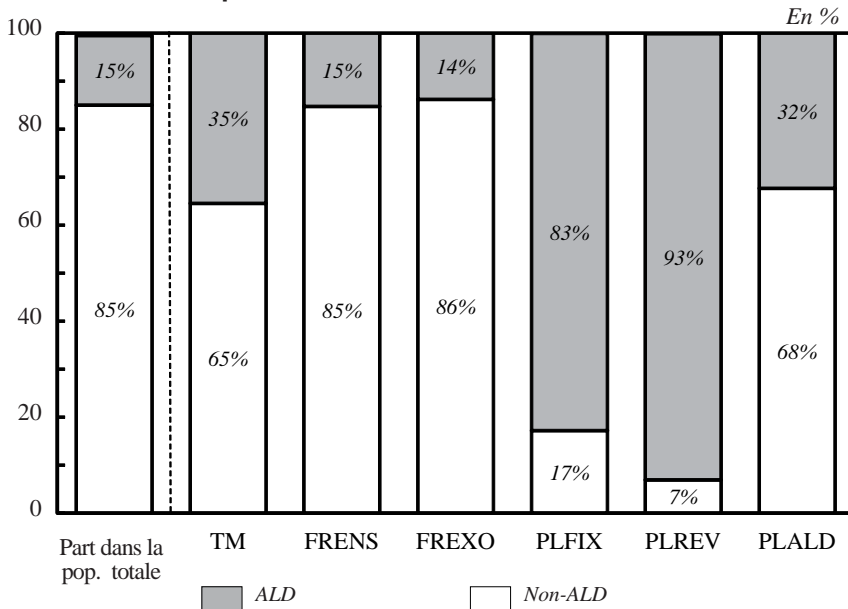
## 10. Distribution de l'effort selon le statut en ALD (PLFIX, PLALD)



Lecture : Dans le cas du scénario PLALD, 10 % de la population en ALD fournissent un effort supérieur à 439 euros suite à la modification du système de remboursement.

Source : DG Trésor, Maquette ISIS.

## 11. Répartition de l'effort selon le statut ALD



Lecture : Dans le cas du scénario PLALD, 32 % de l'effort total est supporté par les populations en ALD.

Source : DG Trésor, Maquette ISIS.



dans ce scénario, parmi les gagnants, le gain des ALD est plus important. En moyenne, les ALD voient leur reste à charge diminuer de 660 euros dans le cas du scénario PLFIX contre 130 euros pour les assurés non-ALD.

Ces différentes analyses de l'effort des assurés mettent en avant l'hétérogénéité des conséquences des différentes réformes envisagées selon les caractéristiques de l'individu. Une analyse selon la part de l'effort supportée par sous population permet de rendre compte des effets redistributifs entre assurés comme des populations contributrices et receveuses.

### 3.2.2. Répartition de l'effort

La concentration de l'effort met en regard la part d'une sous-population dans la population totale avec la part que cette sous-population représente dans l'effort total demandé, soit un milliard d'euros. Cette comparaison sera effectuée au regard de trois indicateurs : le statut en ALD, le revenu et l'âge.

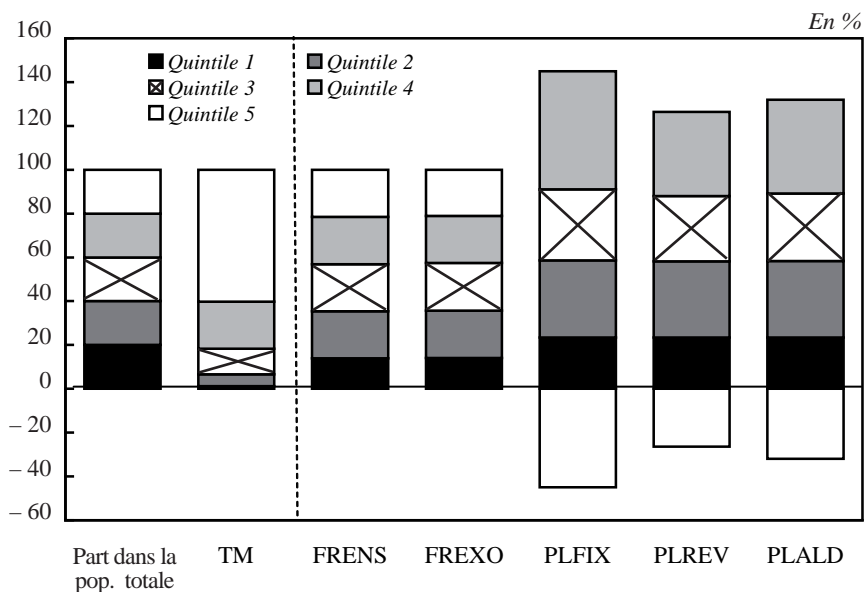
#### 3.2.2.1. La suppression des dispositifs d'exonération pénaliserait fortement les assurés bénéficiant du régime des ALD

Les assurés bénéficiant du régime des ALD représentent 15 % de la population. Dans le cas de l'augmentation du ticket modérateur, 35 % de l'effort total est supporté par les assurés en ALD contre 65 % pour les assurés de droit commun (*cf.* graphique 11). Au contraire, dans le cas des scénarios de franchise, la part de l'effort et la part dans la population totale est la même. Ainsi, 15 % de l'effort est supporté par les assurés en ALD et 85 % par les assurés de droit commun (*cf.* graphique 11).

Dans les scénarios de suppression intégrale des exonérations (scénario 4 PLFIX, 5 PLREV, 6 PLALD), l'effort repose principalement sur les populations en ALD. Ainsi dans le cas des scénarios PLFIX et PLREV, les ALD supportent respectivement 83 et 93 % de l'effort total. Le financement plus important dans le cas du scénario 5 PLREV par les ALD s'explique notamment par un effort conséquent des assurés en ALD ayant des revenus élevés. Ces derniers ne sont en effet plus remboursés à 100 % avant le plafond et n'atteignent que tardivement ce plafond même avec des dépenses importantes. Parallèlement, la modulation du plafond en fonction de la présence d'une maladie chronique permet de limiter considérablement cet effet. Ainsi, l'effort total supporté par les ALD est ramené à 32 % dans le cas du scénario 6 PLALD (*cf.* graphique 11).

Pour autant, ce constat peut être relativisé par l'analyse de la concentration de l'effort en fonction de l'importance de la consommation. Les assurés sont classés par quintile de dépense de soins de ville (*cf.* graphique 12). Les scénarios comprenant un plafond de reste à charge introduisent une redistribution des assurés qui font partie des quatre premiers quintiles de dépenses vers les assurés qui appartiennent au cinquième quintile de dépense. Ainsi, la mise en place d'un plafond de reste à charge permet effectivement une meilleure couverture des assurés qui font face à des dépenses importantes.

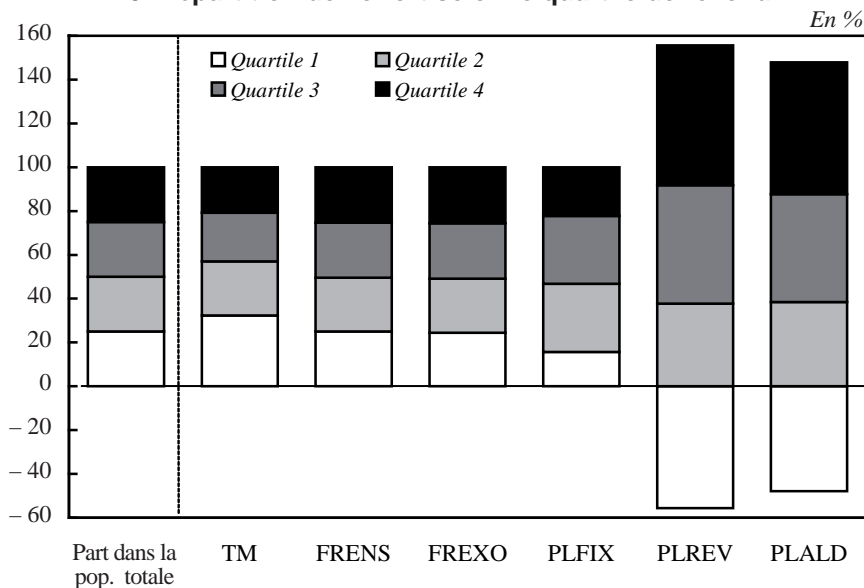
## 12. Répartition de l'effort selon le quintile de dépense de soins de ville



*Lecture* : Dans le scénario TM, les assurés qui font partie des 20 % des plus gros consommateurs (quintile 5) supportent 60 % de l'effort total.

*Source* : DG Trésor, Maquette ISIS.

## 13. Répartition de l'effort selon le quartile de revenu



*Source* : DG Trésor, Maquette ISIS.

### 3.2.2.2. L'introduction d'un critère de revenu instaurerait une redistribution vers les assurés aux revenus les plus modestes

Une analyse comparée de la part de l'effort pour chaque quartile de revenu permet de constater que le scénario TM demande un effort plus important pour les assurés du premier quartile. En effet, 32 % de l'effort total est supporté par le premier quartile de revenu contre 21 % pour le quatrième quartile (cf. graphique 13). Les assurés en ALD et en état de santé moyen ou mauvais sont en effet surreprésentés dans le premier quartile de revenu. 42 % des assurés en ALD et 56 % des personnes se déclarant en mauvais état de santé appartiennent à ce quartile.

Les scénarios introduisant une prise en compte du revenu dans le calcul du plafond (scénario 5 PLREV et scénario 6 PLALD) permettent d'instaurer une redistribution de la couverture assurantielle vers les assurés aux revenus les plus modestes. Ainsi, le premier quartile de revenu est globalement gagnant suite à ces modifications du système de remboursement (cf. graphique 13). À l'inverse, les assurés du dernier quartile sont largement mis à contribution puisqu'ils fournissent, respectivement, 64 et 60 % de l'effort total (cf. graphique 13).

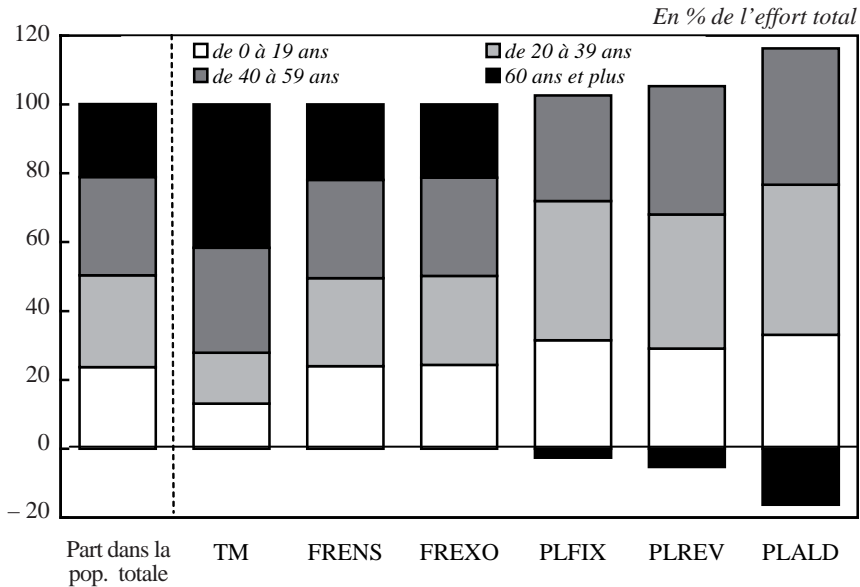
### 3.2.2.3. L'introduction de plafonds de reste à charge induit une redistribution vers les assurés les plus âgés

Dans le cas du scénario TM, les assurés de plus de 60 ans fournissent 42 % de l'effort total alors qu'ils ne représentent que 21 % de la population (cf. graphique 14). Les dépenses de soins sont en effet fortement croissantes avec l'âge du fait d'un état de santé qui se dégrade progressivement. La dépense moyenne annuelle d'un assuré dont l'âge est compris entre 20 et 39 ans s'élève ainsi à 620 euros contre 2 400 en moyenne pour les assurés de plus de 60 ans.

Dans les scénarios avec un plafond de reste à charge, l'analyse de l'effort par âge montre également une redistribution des assurés dont l'âge est inférieur à 60 ans vers les assurés dont l'âge est supérieur à 60 ans (cf. graphique 14). Le plafond de reste à charge permet ainsi aux assurés les plus âgés d'être mieux couverts pour leur dépense de santé. Parmi les plus de 60 ans, ce sont les 80 ans et plus qui bénéficient de l'instauration d'un plafond de reste à charge.

Ces analyses permettent de juger de la répartition de l'effort selon chaque caractéristique retenue de la population. Elles ne permettent pas, en revanche, de décrire l'effort dans son ensemble ou « toutes choses égales par ailleurs ». Une régression par les moindres carrés ordinaires sur l'effort le permet, en contrôlant l'effet de différentes variables et en isolant l'effet propre de chaque déterminant de l'effort.

## 14. Répartition de l'effort selon l'âge



Source : DG Trésor, Maquette ISIS.

### 3.2.3. Les déterminants de l'effort : une analyse « toutes choses égales par ailleurs »

Pour quatre des six scénarios, une régression par les moindres carrés ordinaires est menée sur le montant de l'effort en s'inspirant des travaux de l'Institut de recherche et documentation en économie de la santé (IRDES) lors de l'évaluation du bouclier sanitaire (Debrand et Sorasith, 2010). Parmi variables explicatives, outre des variables relatives à l'état de santé, au revenu et l'âge, sont ajoutées des variables telles que le statut professionnel ou le niveau d'étude (cf. tableau 4). Cette analyse n'est pas menée sur les scénarios 2 FRENS et 3 FREXO de mise en place d'une franchise car la variabilité de l'effort est alors très faible. La quasi-totalité des consommateurs s'acquitte de la franchise. En revanche, une telle analyse pourrait devenir opportune dans un scénario alternatif de mise en place d'une franchise plus élevée et donc non saturée par une part plus importante de la population.

Le résultat des régressions confirme que le scénario TM demande un effort d'autant plus important que l'état de santé est dégradé. Les personnes se déclarant en très bon ou en bon état de santé ont un effort moins important que les assurés se déclarant en état de santé moyen. À l'inverse, les personnes se déclarant en mauvais ou très mauvais état de santé ont un effort plus important que ceux en état de santé moyen.

#### 4. Régression sur l'effort selon les six scénarios

	TM	PLFIX	PLREV	PLALD
État de santé (réf. : moyen)				
• très bon	- 6,662 <sup>(***)</sup>	23,675 <sup>(***)</sup>	19,888 <sup>(***)</sup>	14,251 <sup>(**)</sup>
• bon	- 5,814 <sup>(***)</sup>	20,637 <sup>(***)</sup>	18,240 <sup>(***)</sup>	16,085 <sup>(***)</sup>
• mauvais	9,483 <sup>(***)</sup>	64,452 <sup>(***)</sup>	58,731 <sup>(***)</sup>	57,375 <sup>(***)</sup>
• très mauvais	15,684 <sup>(***)</sup>	101,498 <sup>(***)</sup>	82,311 <sup>(***)</sup>	84,630 <sup>(***)</sup>
Au moins un motif d'exonération	15,903 <sup>(***)</sup>	146,984 <sup>(***)</sup>	165,675 <sup>(***)</sup>	79,526 <sup>(***)</sup>
Revenu (réf. : Q4)				
• Q1	- 0,861	- 7,313	- 97,403 <sup>(***)</sup>	- 78,839 <sup>(***)</sup>
• Q2	- 0,438	3,394	- 23,758 <sup>(***)</sup>	- 17,703 <sup>(***)</sup>
• Q3	0,059	2,088	- 10,545	- 9,391
Homme	- 2,896 <sup>(***)</sup>	24,539 <sup>(***)</sup>	27,356 <sup>(***)</sup>	23,637 <sup>(***)</sup>
Âge				
• âge	- 0,003	4,906 <sup>(***)</sup>	3,931 <sup>(***)</sup>	4,432 <sup>(***)</sup>
• âge <sup>2</sup>	0,003 <sup>(***)</sup>	- 0,084 <sup>(***)</sup>	- 0,073 <sup>(***)</sup>	- 0,076 <sup>(***)</sup>
Couverture complémentaire	4,071 <sup>(***)</sup>	- 22,670 <sup>(*)</sup>	- 29,087 <sup>(***)</sup>	- 26,626 <sup>(***)</sup>
Statut (réf. : actifs)				
• étudiants	2,071 <sup>(*)</sup>	24,751 <sup>(**)</sup>	21,971 <sup>(*)</sup>	23,371 <sup>(***)</sup>
• retraités	0,163	9,719	10,010	7,005
• chômeurs	- 0,067	58,979 <sup>(***)</sup>	59,808 <sup>(***)</sup>	64,465
• inactifs	0,773	2,297	0,081	2,427
• NSP	- 12,569 <sup>(***)</sup>	- 37,873 <sup>(**)</sup>	- 52,619	- 12,500
Niveau d'étude (réf. : 3 <sup>e</sup> cycle)				
• deuxième cycle	0,331	7,162	7,563	7,144
• premier cycle	1,634 <sup>(**)</sup>	6,565	7,374	9,627
• primaire	2,306 <sup>(**)</sup>	8,883	- 1,134	3,368
• sans diplôme	7,140 <sup>(***)</sup>	26,355	3,239	13,600
Nombre de personnes dans le ménage	- 0,433 <sup>(**)</sup>	- 4,739	- 13,704 <sup>(***)</sup>	- 11,979 <sup>(***)</sup>
Constante	8,57 <sup>(***)</sup>	- 38,49 <sup>(**)</sup>	44,55 <sup>(**)</sup>	30,08 <sup>(*)</sup>

*Lecture* : Pour le scénario PLFIX, avoir un an de plus nécessite en moyenne un effort supplémentaire de 4,906 euros. Pour le même scénario, les hommes fournissent en moyenne un effort supplémentaire de 24,539 euros par rapport aux femmes.

*Notes* : (\*\*\*) Significatif à 1 % ; (\*\*) Significatif à 5 % ; (\*) Significatif à 10 %.

*Source* : Maquette ISIS, DG Trésor.

Parallèlement, les scénarios introduisant le revenu dans le calcul de la participation financière du patient (scénario 5 PLREV et scénario 6 PLALD) sollicitent un effort d'autant moins important que le revenu est faible. Dans ces deux scénarios, les individus dont le revenu est inférieur à la médiane fournissent un effort plus faible que les assurés les plus aisés. La différence d'effort demandé au troisième quartile par rapport au quatrième quartile n'est par ailleurs pas significative.

Enfin, le scénario TM demande un effort plus important aux très jeunes et aux très âgés. Les trois derniers scénarios sollicitent davantage les assurés au milieu de la pyramide des âges. En effet, pour le scénario TM, l'effort

varie en fonction de l'âge selon une courbe en U. Pour les scénarios avec plafond de reste à charge, la courbe d'effort décrit un U inversé.

Ces différentes analyses montrent la diversité de l'effort en fonction des caractéristiques socio-économiques. La régression par les moindres carrés ordinaires sur l'effort a permis de plus de contrôler l'effet de différentes variables et d'isoler l'effet propre de chaque déterminant de l'effort. Si l'ensemble de ces analyses repose sur l'analyse de la variation du reste à charge avant intervention des assurances-maladie complémentaires, du fait notamment des limites de la maquette ISIS, il semble utile d'appréhender même globalement l'impact de ces mesures sur le secteur de l'assurance complémentaire.

### *3.2.4. Impact sur les organismes complémentaires santé*

Les résultats précédents décrivaient l'évolution de la prise en charge de l'assurance-maladie obligatoire. Pour autant, l'évolution de la couverture de l'assurance-maladie obligatoire pourrait induire des changements sur le marché de l'assurance complémentaire. Après l'évaluation de l'effet mécanique sur les primes d'assurance d'un transfert de prise en charge d'un milliard d'euros, les implications de la mise en place d'un plafond de reste à charge sur la place des assurances complémentaires dans le système d'assurance-maladie sont ensuite présentées.

#### 3.2.4.1. Un faible impact à la hausse sur les primes d'assurance

Les hypothèses retenues en termes de modalités de prise en charge des soins reprenant les pratiques actuelles, le ticket modérateur est considéré comme intégralement couvert par l'assurance complémentaire. La hausse du ticket modérateur implique alors un transfert de financement vers les financeurs privés dont les organismes complémentaires. Ce transfert n'est pas total. 6 % de la population ne bénéficient pas d'une assurance complémentaire (Perronnin, Pierre et Rochereau, 2011). À partir de la maquette ISIS, on détermine le montant qui est transféré vers les assureurs complémentaires. Sur le milliard d'euros d'économie réalisée par l'assurance-maladie, ce sont 940 millions d'euros qui seraient ainsi transférés aux organismes complémentaires.

Parallèlement, le ratio de remboursement versé sur primes perçues<sup>(16)</sup> s'élève en moyenne sur la période 2005-2010 à 77 %. On fait l'hypothèse que ce ratio est constant et on en déduit alors la hausse des primes consécutive à une hausse des remboursements de 940 millions d'euros. En moyenne, sur les contrats individuels, les primes annuelles devraient augmenter de 11 euros du fait du transfert de charge soit une augmentation de 2,4 % de la prime moyenne. Pour les contrats collectifs, la hausse serait de 10 euros (cf. tableau 5).

(16) Cf. Fonds CMU.

## 5. Primes moyennes et augmentations prévisibles des primes

	Contrats individuels		Contrats collectifs	
	Primes moyennes annuelles	Augmentation des primes	Primes moyennes annuelles	Augmentation des primes
Classe A	634	15	428	10
Classe B	484	12	415	10
Classe C	458	11	428	10
Classe D	348	8	421	10
Ensemble	468	11	420	10

*Lecture* : La classe A, qualifiée de « haut de gamme », regroupe des contrats offrant des garanties élevées. Par rapport aux contrats « standard » C, les contrats B se définissent comme apportant une meilleure couverture en optique et en dentaire. La classe D est qualifiée d'« entrée de gamme ». Les garanties offertes par les contrats de la classe D vont rarement au-delà du tarif de la Sécurité sociale. Pour plus de détails sur les garanties de ces contrats voir Arnould et Vidal (2008). La prime indiquée pour les contrats collectifs correspond uniquement à la part versée par le salarié.

*Sources* : Enquête DREES sur les contrats modaux (2006) et calculs DG Trésor.

Si les franchises n'étaient pas couvertes par les assureurs complémentaires, leur mise en place ne devrait pas avoir d'impact sur les primes. Si elles étaient couvertes, leur impact sur les primes serait comparable à celui d'une hausse du ticket modérateur.

Pour les scénarios introduisant un plafond de reste à charge, des changements dans la régulation du mode de fixation des primes pourraient intervenir. Actuellement dans le cadre des contrats « solidaires », les différents organismes complémentaires ne subordonnent pas l'accès aux contrats à un questionnaire de santé. Les organismes complémentaires fixent les prix des contrats seulement selon des critères d'âge, de sexe et de lieu de résidence. Ces contrats solidaires sont encouragés par les pouvoirs publics, qui prévoient des dispositifs fiscaux avantageux pour les organismes ayant adopté cette formule. Pour autant, des modifications des paramètres pris en compte pour la fixation du prix du contrat pourraient intervenir suite à la mise en place d'un plafond de reste à charge. Par exemple, si le plafond était calculé en fonction du revenu, il serait envisageable d'introduire le revenu dans le calcul de la prime d'assurance complémentaire. En effet, dans ce cas, la couverture complémentaire serait d'autant plus faible que le revenu est faible. La prise en charge intégrale par l'assurance-maladie publique se déclencherait pour des niveaux de dépenses faibles pour les assurés aux revenus modestes.

### 3.2.4.2. Un risque de désaffiliation mais qui semble limité dans le cas de la mise en place d'un plafond de reste à charge

La mise en place d'un plafond de reste à charge pourrait inciter les assurés à abandonner leur couverture complémentaire. Les assurés sociaux

en bonne santé pourraient notamment décider de ne plus se couvrir auprès d'un assureur complémentaire pour deux raisons. Tout d'abord, ils savent que leur état de santé est satisfaisant et ne se sentent donc pas exposés à des dépenses de soins importantes. La couverture publique est alors suffisante dans ce cas. Parallèlement, l'existence d'un plafond leur permet de connaître leur risque maximum. L'utilité d'une assurance complémentaire pour se prémunir contre la survenance d'un risque catastrophique n'est alors plus nécessaire. Le risque de sélection des mauvais risques pourrait, à terme, remettre en question la viabilité du secteur de l'assurance complémentaire.

Pour autant, selon Briet et Fragonard (2007), deux raisons essentielles devraient limiter la portée du risque de désaffiliation. Tout d'abord, la décision de s'affilier à une assurance complémentaire dépend certes de la prise en charge du ticket modérateur mais aussi et surtout de la prise en charge des dépenses peu remboursées par la Sécurité sociale (optique, dentaire) ou encore la prise en charge de la partie libre des tarifs des soins (dépassements). Ensuite, les prestations d'assurance supplémentaire, c'est-à-dire sur les dépenses en dehors du panier de soins remboursables, peuvent également constituer un motif d'affiliation.



## Annexe

# Présentation de la maquette ISIS

**Bérengère Mesqui**

*DARES*

Le bureau Santé et comptes sociaux de la Direction générale du Trésor a développé, en collaboration avec la DREES, une maquette de microsimulation des dépenses de soins de santé et de leur prise en charge par les organismes obligatoire et complémentaires d'assurance-maladie.

## 1. Les données utilisées

La maquette ISIS<sup>(1)</sup> est construite à partir des données de l'Enquête santé et protection sociale (ESPS) 2006 de l'IRDES<sup>(2)</sup> appariées avec l'échantillon permanent des assurés sociaux (EPAS)<sup>(3)</sup>, qui regroupe des assurés de l'ensemble des régimes de sécurité sociale et est géré par la CNAMTS (*cf.* graphique 1).

L'enquête ESPS fournit, pour un échantillon de 22 460 individus<sup>(4)</sup>, des données sur l'état de santé, la couverture maladie et le recours aux soins en fonction des caractéristiques sociales (âge, sexe, catégorie socioprofessionnelle, occupation, éducation...), ainsi que des informations sur la composition et le revenu des ménages.

---

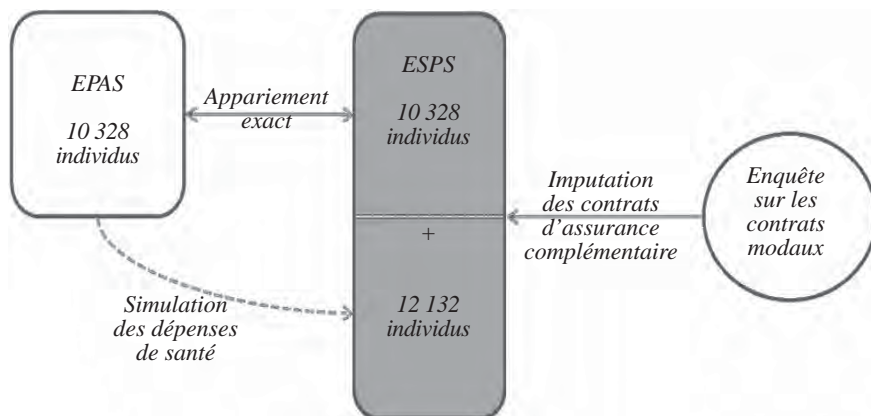
(1) Instrument de simulation individualisée des dépenses de santé.

(2) Pour plus d'information sur ESPS, voir Allonier, Dourgnon et Rochereau (2006).

(3) L'EPAS s'appelle maintenant EGB (échantillon général des bénéficiaires).

(4) Parmi ces individus, certains ont plusieurs contrats d'assurance complémentaire. Au total, on a donc 23 000 couples individu-contrat.

## 1. Les différentes sources de données



Source : Auteur.

Pour environ la moitié de ces individus (10 328), l'EPAS décrit les dépenses totales de santé, ainsi que les remboursements de l'assurance-maladie obligatoire (AMO). Ces dépenses sont ventilées en 15 postes<sup>(5)</sup>, dont 4 pour les praticiens, 4 pour les auxiliaires médicaux, 6 pour les prestations sanitaires diverses et 1 pour les dépenses hospitalières. Pour chacun de ces postes de dépenses, on connaît le montant total dépensé dans l'année, les remboursements de l'AMO, le montant des tickets modérateurs, ainsi que les dépassements payés par le patient.

Les informations sur les conditions de remboursement des complémentaires santé sont tirées de l'enquête de la DREES sur les contrats les plus souscrits auprès des organismes complémentaires. Cette enquête fournit, pour 957 contrats, les taux de remboursement d'une consultation chez le spécialiste et les montants forfaitaires remboursés pour les prothèses dentaires et pour les lunettes.

Finalement, pour les 22 460 individus de l'enquête ESPS, la maquette ISIS fournit des informations socio-démographiques, des données sur la perception de l'état de santé, les revenus du ménage, déclarés ou simulés, les dépenses de santé et les remboursements AMO par grand type, observés ou simulés et les remboursements AMC simulés.

(5) Omnipraticiens, spécialistes, dentistes, sages-femmes, infirmières, kinésithérapeutes, autres auxiliaires médicaux, actes de biologie, pharmacie, prothèses, orthèses, matériel, transports, optique, autre pharmacie, hôpital.

## 2. Imputation des revenus des ménages

Dans l'enquête ESPS, le revenu total des ménages est renseigné pour environ deux tiers des ménages enquêtés. Pour les autres ménages, il a été imputé par une méthode de résidus simulés. Afin que l'imputation soit la plus précise possible, les ménages ont été divisés en trois groupes :

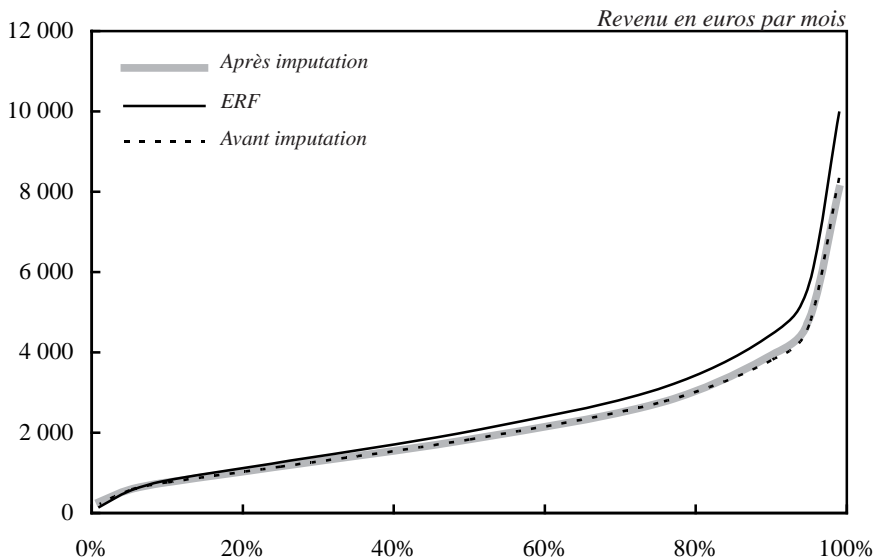
- les ménages dans lesquels tous les adultes sont actifs ;
- les ménages dans lesquels tous les adultes sont inactifs ;
- les autres ménages.

Dans chacune de ces classes, on effectue une régression linéaire sur le logarithme des revenus des ménages répondants. Les variables explicatives retenues sont les suivantes : l'âge des personnes du ménage, leur occupation principale, leur niveau d'étude, leur catégorie socioprofessionnelle, leur statut d'emploi (salarié, fonctionnaire ou non salarié), le sexe du chef de famille, la présence dans le ménage d'au moins un bénéficiaire de la CMU ainsi que celle d'un bénéficiaire du RMI.

Le revenu imputé aux ménages non répondants est alors égal à la somme de la valeur prédite par la régression et d'un résidu tiré aléatoirement dans la distribution observée des résidus pour les répondants.

On vérifie que l'imputation n'affecte pas sensiblement la distribution des revenus des ménages (cf. graphique 2).

### 2. Distribution des revenus des ménages avant et après imputation



Source : Calcul de l'auteur.

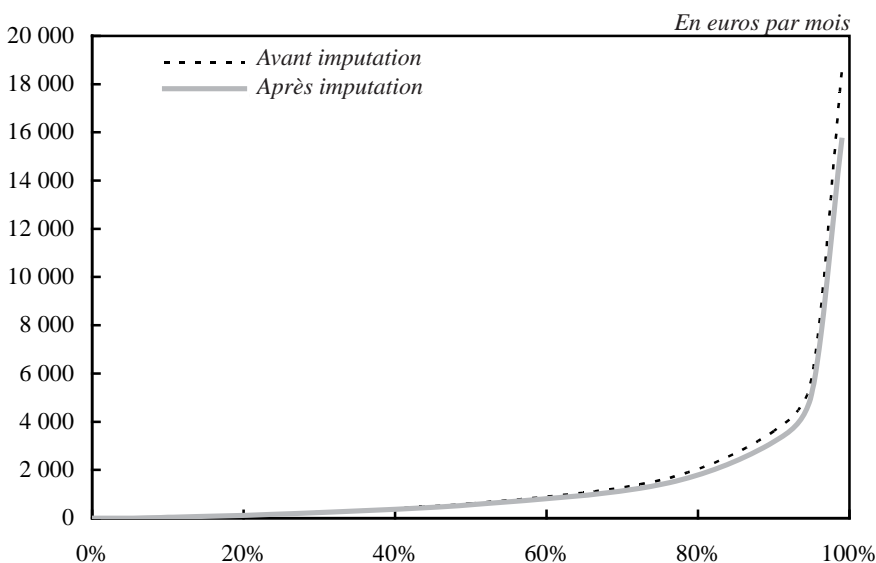
### 3. Imputation des dépenses de santé

Les données sur les dépenses de santé des ménages et les remboursements de l'AMO sont disponibles dans l'EPAS pour la moitié environ des individus de l'enquête (10 328). On utilise une méthode de *hotdeck* pour imputer les dépenses de santé des individus pour lesquels cette information n'est pas disponible. Cette méthode d'imputation consiste à attribuer aléatoirement à un individu les dépenses de santé d'un individu de mêmes caractéristiques et dont on connaît les dépenses. Elle permet d'obtenir des estimateurs sans biais et de préserver la distribution de la variable d'intérêt.

Les différents postes de dépenses de santé, hors optique et dentaire, étant fortement corrélés, on impute le profil complet de ces dépenses et des remboursements AMO à partir d'un même individu. Les dépenses et remboursements d'optique et des soins dentaires sont imputés de manière séparée. Comme précédemment, la distribution des variables d'intérêt est peu affectée par l'imputation (*cf.* graphiques 3 et 4). La méthode de *hotdeck* permet ici de conserver une partie des liens entre les différents postes de dépense de santé.

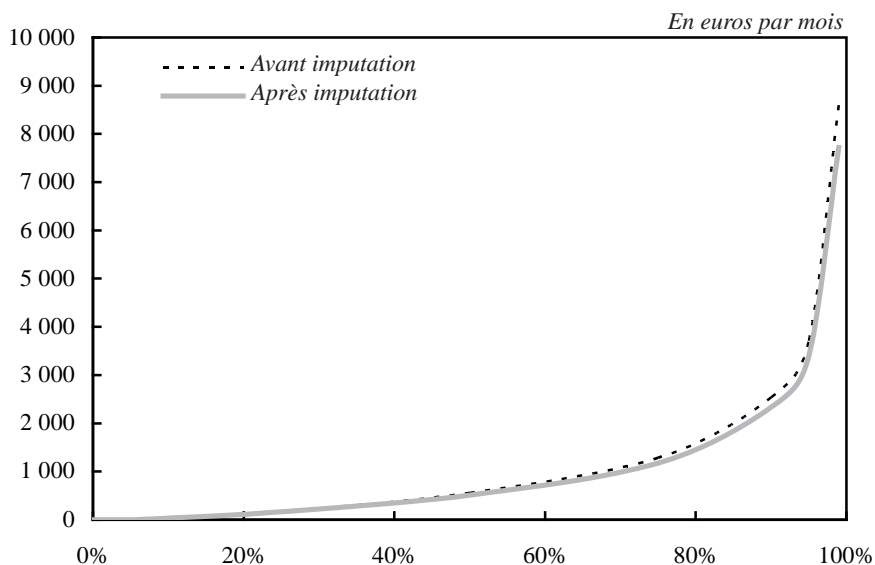
Les caractéristiques individuelles utilisées pour imputer les dépenses sont les suivantes : le sexe et l'âge de l'individu, son niveau d'études, sa perception de son état de santé, la classe de niveau de vie de son ménage (trois classes ont été définies), son nombre de visites chez le généraliste, son régime de Sécurité sociale et le type d'organisme lui fournissant sa couverture complémentaire.

#### 3. Distribution des dépenses médicales totales



Source : Calcul de l'auteur.

#### 4. Distribution des dépenses en médecine ambulatoire



Source : Calcul de l'auteur.

#### 4. Remboursements de l'assurance-maladie complémentaire

L'appariement ESPS-EPAS ne donne aucune information sur les montants remboursés par les organismes complémentaires. Pour les simuler, on utilise l'enquête de la DREES sur les contrats les plus souscrits auprès des complémentaires santé. À chaque individu de l'enquête ESPS on attribue à l'aide d'un tirage aléatoire (proportionnel au nombre d'assurés pour chaque contrat) un contrat de l'enquête DREES, selon que son contrat complémentaire est collectif ou individuel, obtenu ou non par son entreprise, et fourni par une mutuelle, un assureur ou un institut de prévoyance. Les remboursements de l'organisme complémentaire sont ensuite calculés en fonction des dépenses de santé de l'individu et des caractéristiques du contrat qui lui a été attribué. Les informations retenues pour les contrats comprennent des taux de remboursement par type de dépenses.

### La maquette ISIS est globalement représentative en niveau comme en structure de la consommation de soins en 2006

En %

	Maquette ISIS				Comptes de la santé en 2006			
	AMO	AMC	Ménages	Dépenses totales <sup>(*)</sup>	AMO	AMC	Ménages	Dépenses totales <sup>(*)</sup>
Soins ambulatoires	66	22	12	38 591	65	22	13	4 2495
• médecins	74	17	9	19 109	72	20	7	19 823
• auxiliaires	80	18	2	6 700	80	12	8	9 457
• dentistes	38	33	30	9 005	35	38	26	9 080
• laboratoires d'analyse	72	24	4	3 777	76	23	1	4 135
Transports	85	13	3	3 209	95	4	2	3 083
Pharmacie	77	21	2	26 819	67	20	12	31 942
Autres biens médicaux	36	30	34	8 815	41	31	28	8 888
Total soins de ville	66	21	13	77 433	67	19	14	86 408

Note : (\*) En millions d'euros.

Source : Calculs de l'auteur.

## En conclusion

Les points forts de la maquette ISIS sont :

- elle reconstruit une population représentative de la population française caractérisée par un grand nombre de variables socio-économiques et les dépenses de soins par postes de consommation (22 500 individus et 8 100 ménages) ;
- elle décrit la situation récente (2006) de notre système d'assurance-maladie ;
- elle permet l'étude détaillée au niveau des individus comme des ménages des effets des mesures affectant la prise en charge des assurés ;
- la maquette permet d'étudier le reste à charge moyen individuel après intervention de l'organisme complémentaire.

Les limites de la maquette ISIS liées aux données disponibles sont :

- le montant de franchise médicale acquittée par l'assuré n'est pas renseigné et reste délicat à simuler ;
- les dépenses de pharmacie ne sont pas détaillées selon leur taux de remboursement ;
- les dépenses d'hospitalisation ne sont pas correctement collectées dans la base de données de l'échantillon permanent des assurés sociaux (EPAS). Ces dépenses sont sous-estimées et mal individualisées dans cette base.

## Références bibliographiques

- Allonier C., P. Dourgnon et T. Rochereau (2006) : « L'Enquête Santé protection sociale 2004, un outil d'analyse pluridisciplinaire de l'accès aux soins, de la couverture maladie et de l'état de santé des Français. », *Questions d'Économie de la Santé IRDES*, n° 105, janvier.
- Baillet A. (2010) : « Une estimation du coût du dispositif des affections de longue durée en soins de ville pour le régime général d'assurance-maladie », *DREES, Comptes nationaux de la Santé*.
- Briet R. (2010) : *Rapport du groupe de travail sur le pilotage des dépenses d'assurance-maladie*, Présidence de la République, La Documentation française.
- Briet R. et B. Fragonard (2007) : *Mission 'Bouclier sanitaire'*, ministère de la Santé, de la Jeunesse et des Sports, La Documentation française.
- Debrand T. et C. Sorasith (2010) : « Bouclier sanitaire : choisir entre égalité et équité ? Une analyse à partir du modèle ARAMMIS », *Document de Travail de l'IRDES*, n° 32.
- Geoffard P-Y. et G. de Lagasnerie (2011) : « Réformer le système de remboursement pour les soins de ville : une analyse par microsimulation », *Économie et Statistique*, à paraître.
- Haut-Conseil pour l'avenir de l'Assurance-maladie (HCAAM) (2011) : *L'assurance-maladie face à la crise. Mieux évaluer la dépense publique d'assurance-maladie : l'ONDAM et la mesure de l'accessibilité financière des soins*, Rapport annuel. Disponible sur [http://www.securite-sociale.fr/IMG/pdf/rapport\\_annuel\\_hcaam\\_version\\_definitive\\_2011-2.pdf](http://www.securite-sociale.fr/IMG/pdf/rapport_annuel_hcaam_version_definitive_2011-2.pdf)
- Jones A.M., E. Zucchelli et N. Rice (2010) : « The Evaluation of Health Policies Through Microsimulation Methods », *University of York HEDG Working Paper*, n° 10/03.
- Newhouse J.P. (1993) : *Free for All? Lessons from the RAND Health Insurance Experiment*, Harvard University Press.
- Perronnin M., A. Pierre et T. Rochereau (2011) : « La complémentaire santé en France en 2008 : une large diffusion mais des inégalités d'accès », *IRDES, Questions d'Économie de la Santé*, n° 161.



## Complément C

# Évaluation économique de la vie humaine : fondements théoriques, mesure empirique et implications en matière de choix social

**Thomas Weitzenblum**

*Université Lille 2 et EQUIPPE*

**Jérôme Wittwer**

*LEDa, Université Paris Dauphine*

### Introduction

La théorie économique en général, et la microéconomie en particulier, analysent les choix efficaces des agents en les fondant sur la valeur relative des différents biens considérés et sur l'existence d'une contrainte globale de ressources. Cette démarche est-elle applicable, telle quelle, dans le cas de la santé ? S'il est délicat de répondre dès à présent par l'affirmative, considérer à l'inverse que la vie humaine ne peut donner lieu à aucun arbitrage, car « la vie n'a pas de prix », n'est pas non plus raisonnable. Au niveau macroéconomique, il est clair que ce type de raisonnement mènerait en effet à des dépenses de santé arbitrairement grandes, qui entreraient en conflit avec d'autres dépenses, tout aussi indispensables – nourriture, énergie entre autres. Il n'est donc pas question de consacrer l'intégralité des ressources d'une économie à son secteur de santé.

De fait, les choix collectifs en dépenses de santé, illustrés par exemple par la part du PIB consacrée à la santé, correspondent à un certain arbitrage, qu'il soit implicite ou explicite. Dès lors, en quoi les outils de l'économiste pourraient-ils éclairer les choix d'une société en matière de dépenses de santé ? Certains pays ont-ils dès à présent mis en place des arbitrages entre les dépenses de santé reposant sur des critères économiques clairement identifiés ?

Pour répondre à ces questions, il convient de se doter d'une méthode permettant la valorisation de la bonne santé. *A minima*, cette valorisation devra être relative, c'est-à-dire pouvoir ordonner et comparer différents états de santé entre eux. L'étape la plus aboutie de cette démarche – mais aussi celle qui suscite les questionnements, sinon les controverses, les plus nombreux – consiste à valoriser en termes monétaires la vie et la bonne santé. Évaluer en unités monétaires une année de vie peut légitimement heurter les principes des uns, et il conviendra de conserver à l'esprit le fait que ces valorisations, dans le meilleur des cas, demeureront quantitativement imprécises. Pourtant, cette étape apparaît comme un passage obligé pour mener à bien une démarche économique des choix en matière de santé, car dès lors qu'un arbitrage avec d'autres dépenses est effectué, la valeur de la santé relativement à d'autres biens de consommation se pose inévitablement.

Cette valorisation, bien qu'elle ait pour vocation d'être utilisée dans un processus de choix collectif, doit cependant être construite à partir d'une approche individuelle, pour des raisons évidentes. C'est pourquoi, la valorisation économique de la vie humaine trouve ses fondements dans la description des comportements microéconomiques à l'égard de la bonne santé et de la vie. Si, au niveau individuel, le problème de la valeur absolue de la vie semble insoluble – quelle somme une personne à l'article de la mort serait-elle prête à dépenser pour éviter de décéder immédiatement ? –, il est toutefois possible de construire une valorisation de la vie qui contourne cette difficulté, en comparant des risques de décès, plutôt qu'en comparant de manière certaine la vie et la mort. La méthode employée présentera au moins ce premier intérêt, qui est de représenter la valeur de la santé ou la vie, non pas directement, mais indirectement en quantifiant les dépenses que l'individu est prêt à consacrer pour réduire la probabilité d'occurrence de l'accident.

L'objet du présent complément est de donner un aperçu de la méthodologie employée par les économistes pour déterminer la valeur économique de la vie, puis de la façon dont cette valorisation est, ou pourrait être, prise en compte dans le processus de choix collectif, qui lui-même mérite d'être clarifié. Nous présentons d'abord les outils dont les économistes disposent, et qui sont d'ordre théorique comme empirique, pour évaluer la vie humaine, puis, la question du choix macroéconomique et effectif des dépenses de santé est posée. À cette occasion, certains choix effectivement entérinés dans les grands pays industrialisés, et qui reposent au moins partiellement sur des raisonnements économiques, sont mis en lumière.

# 1. Évaluation économique de la vie humaine

Dans un premier temps, nous nous penchons sur les méthodes employées par les économistes pour évaluer économiquement la vie humaine. Ces approches sont de deux ordres : théoriques et empiriques. L'approche théorique vise à intégrer le « bien santé » dans le comportement micro-économique de référence. L'approche empirique, certainement la plus immédiatement mobilisable, quantifie la valeur de la vie humaine à partir de comportements observés ou d'enquêtes réalisées. La valeur de la vie se déduit par exemple de l'observation du lien statistique entre revenu du travail et risque d'accident mortel : cette valeur est donc appréhendée comme le résultat d'arbitrages individuels. La question de l'intégration de cette valorisation individuelle dans un choix macroéconomique est abordée dans la seconde partie du présent complément.

## 1.1. La valeur de la vie : approches théoriques

Il peut *a priori* sembler présomptueux pour l'économiste de modéliser les comportements des agents mettant en jeu leur santé, et la manière dont ils l'évaluent – notamment en comparaison des biens économiques plus usuels, comme les biens de consommation courante. Les analyses empiriques sont d'ailleurs elles-mêmes bien loin de donner lieu à un consensus. Pourtant, l'analyse théorique présente de nombreux attraits, que l'on peut d'ailleurs situer, pour certains en amont, et pour d'autres en aval, de l'analyse empirique :

- en amont, la modélisation de l'effet des comportements individuels, en termes de demande de santé, sur les probabilités de survie ainsi que sur la qualité de vie, force à se poser la question de l'allure de la fonction d'utilité, donc à s'interroger sur la valorisation des périodes en vie par rapport à la mort. Les bases seront donc posées, qui pourront être ultérieurement reprises, ou à défaut, reliées, aux différents modèles économétriques (*cf.* partie empirique) ;
- en aval, une fois le modèle théorique construit, il peut être calibré – ce qui nécessite en soi le plus grand soin – puis simulé, pour évaluer l'effet de tel ou tel choc. La quantification obtenue inclut alors la prise en compte de l'ajustement des comportements micro-économiques. On pourra, par exemple, tenter de quantifier l'évolution future de la part des dépenses de santé dans le PIB. D'aucuns jugeraient qu'il s'agit, dans l'esprit, de l'étape la plus aboutie de l'analyse. Ces mêmes personnes devront toutefois admettre que cette démarche est très hasardeuse, car s'y rencontrent de fortes hypothèses théoriques pour la plupart discutables, et des quantifications parfois peu sûres.

### 1.1.1. *Espérance de vie et santé dans l'analyse microéconomique*

Pour l'économiste, le cadre de référence microéconomique est la maximisation de l'utilité intertemporelle. Ce paradigme repose, entre autres, sur l'hypothèse de rationalité des agents. La santé peut y être appréhendée de diverses manières, non exclusives les unes des autres. Les deux principales sont :

- la santé est un déterminant de l'espérance de vie ;
- la santé peut affecter la qualité de la vie.

L'effet via l'espérance de vie pose la question de la valorisation d'une année de vie en plus, en termes de pure satisfaction de l'agent, ainsi qu'en termes plus aisément quantifiables comme le surcroît de salaire ou de fertilité (l'agent ayant vu sa durée de vie prolongée d'une année est susceptible de procréer durant cette année de vie supplémentaire).

La prise en compte de la qualité de la vie invite à considérer la santé comme un bien qui affecte le bien-être des individus. Il s'agit donc d'un argument de la fonction d'utilité instantanée et non pas seulement de l'allongement de l'horizon de vie. Dès lors, se pose la question du degré de substituabilité de la santé avec le bien de consommation usuel. Doit-on considérer que ces deux biens sont complémentaires (on profite d'autant mieux de la consommation de biens courants que l'on est en bonne santé : un téléviseur avec très haute définition d'image n'est utile que si l'on a une bonne vue) ou substituables (lorsqu'on a le plus grand mal à se déplacer, une aide – comme des fauteuils roulants motorisés – peut accroître substantiellement la satisfaction, bien plus qu'un vélo faciliterait la vie d'un bien portant) ? À l'évidence, il n'existe pas de réponse définitive, car les cas de figure sont très nombreux, et correspondent pour certains au cas de la complémentarité, pour d'autres à celui de la substituabilité. À tout le moins, la question vaut d'être posée, et invite dès à présent à réfléchir, non pas uniquement au concept de valeur marginale de la santé, mais également aux interactions entre la valeur marginale de la consommation standard et celle du « bien santé ».

Par ailleurs, le type de fonction d'utilité intertemporelle retenu – par exemple, une fonction séparablement additive dans le temps – et sa paramétrisation ne seront pas du tout neutres quant aux conclusions qui en seront tirées *in fine* : dans quelle mesure la quantité de vie et la qualité de vie sont-elles substituables ? Si ces deux grandeurs présentent un très fort degré de substituabilité, tout se passe comme si un agent était indifférent, par exemple, entre une espérance de vie de dix ans et un niveau de consommation donné, d'une part, et, d'autre part, une espérance de vie de cinq ans seulement pourvu que la consommation soit doublée. Dans ce cas, la valeur d'une année de vie supplémentaire peut-être nulle, car le temps n'est pas valorisé en tant que tel, et car une vie accrue nécessite de répartir un volume de consommation donné sur un plus grand nombre de périodes. À l'inverse, en cas de parfaite complémentarité, la valeur d'une année de vie supplémentaire peut devenir arbitrairement grande : ce cas limite est l'illustration analytique de l'idée selon laquelle « la vie n'a pas de prix ». Ceci témoigne bien de la nécessité, lorsqu'on pose les jalons de la modélisation théorique de la santé et de ses liens avec les comportements économiques, d'aborder les caractéristiques essentielles de la santé, qui pourront être intégrées dans l'analyse.

### 1.1.2. La prise en compte de la santé au sein du paradigme microéconomique

Historiquement, la prise en compte des variations de la mortalité sur le bien-être intertemporel fut très précoce. Usher (1973), par exemple, dans un cadre très épuré, calcule la valeur d'une variation marginale du taux de mortalité à un âge donné, mesurée en unités de consommation certaine à une date initiale. Arthur (1981) met en lumière les différents canaux par lesquels la baisse de la mortalité affecte le bien-être : surcroît de « pure » satisfaction, surcroît d'offre de travail, surcroît de fertilité, mais d'un autre côté surcroît de consommation à financer. En revanche, les agents ne sont pas maîtres de l'évolution de leur état de santé, qui est supposée exogène et appréhendée par des tables de mortalité, et la qualité de vie n'est pas un argument de la satisfaction des agents<sup>(1)</sup>. Les modèles développés par la suite dépasseront ces limites<sup>(2)</sup>.

Le comportement d'un agent économique peut être représenté par la maximisation de son espérance d'utilité intertemporelle<sup>(3)</sup> :

$$U(a) = \sum_{t=a}^T \frac{S_{t,a}}{(1+\delta)^{t-a}} u(c_t, H_t)$$

où  $a$ ,  $c_t$ ,  $H_t$ ,  $u(c_t, H_t)$ ,  $T$ ,  $\delta$  et  $S_{t,a}$  sont respectivement l'âge de l'agent, la consommation à l'âge  $t$ , l'état de santé à l'âge  $t$ , la fonction d'utilité instantanée, l'âge maximal possible, le taux de préférence pour le présent et le taux de survie à l'âge  $t$  conditionnellement au fait d'être en vie à l'âge  $a$ . La valeur théorique de la vie peut alors être appréhendée comme le taux marginal de substitution (TMS) entre la mortalité et l'utilité marginale de la consommation à l'âge courant  $a$  :

$$TMS(a) = - \frac{\frac{\partial U(a)}{\partial \pi(a)}}{\frac{\partial U(a)}{\partial c}}$$

(1) C'est également le cas pour l'analyse développée par Rosen (1988), qui combine modèle théorique et application économétrique.

(2) Murphy et Topel (2006) modélisent la qualité de vie et son effet sur l'utilité courante tout en considérant que le profil de la qualité de vie au cours du cycle de vie s'impose à l'agent sans que celui-ci ne puisse la modifier par ses comportements. Ehrlich et Yin (2004) construisent un modèle à des fins de simulations numériques, où la dépense de santé réduit le taux de mortalité, sans toutefois affecter la qualité de vie, qui est absente de la modélisation. Enfin, Hall et Jones (2007) décrivent explicitement le comportement de dépenses et de production de santé par les agents économiques, qui affecte à la fois la qualité de vie et l'espérance de vie.

(3) L'approche adoptée suppose l'existence d'un marché parfait d'annuités : dans ce cas, les agents lèvent toute incertitude concernant leur profil de consommation. L'unique risque porte sur leur mortalité.

où  $\pi(a)$  est la probabilité de décès à l'âge  $a$ . Ce taux représente la valeur que l'agent attribue à la réduction du risque de mortalité, évaluée en termes de biens de consommation. Plus ce taux est élevé, plus précieuse est la réduction de la mortalité aux yeux de l'agent, donc plus grande est la valeur de la vie. Pour simplifier la présentation, on se situera donc au début de la dernière période de vie possible  $T$ , (donc  $\pi(T+1) = 1$ ) et on ne tiendra pas compte de l'effet de la qualité de la santé  $H_t$ . L'utilité intertemporelle se résume donc à :

$$U(T) = (1 - \pi(T))u(c_T)$$

Une variation infinitésimale  $d\pi(T)$  est envisagée. La contrainte budgétaire de l'agent, en raison de l'annuité qui lui est versée s'il demeure en vie à la dernière période, s'écrit :

$$A_{T-1} \frac{1+r}{1-\pi(T)} + w_T = c_T$$

où  $w_T$  désigne le revenu du travail à la date  $T$ .

La valeur marginale de la vie, soit  $dU/d\pi$ , vaut alors :

$$\begin{aligned} \frac{\partial U}{\partial \pi(T)} &= \frac{\partial}{\partial \pi(T)} \left[ (1 - \pi(T))u \left( A_{T-1} \times \frac{1+r}{1-\pi(T)} + w_T \right) \right] \\ &= -u(c_T) + \frac{(1+r)A_{T-1}u'(c_T)}{1-\pi(T)} = -u(c_T) - (w_T - c_T)u'(c_T) \end{aligned}$$

Exprimée par unité de consommation, on obtient la valeur d'une année de vie (car la mort est certaine à l'issue de la période  $T$ ) :

$$-\frac{\frac{\partial U}{\partial \pi}}{\frac{\partial U}{\partial c}} = \frac{u(c_T)}{u'(c_T)}(T) + (w_T - c_T)$$

Cette formule se comprend aisément : ce que vaut, en termes de biens de consommation, le fait de vivre une période de plus, est la somme de l'utilité pure tirée de la vie (ramenée à des unités de consommation en normalisant cette utilité par l'utilité marginale) et du revenu que l'on reçoit durant cette période supplémentaire, à laquelle on retire la consommation effective. Si la consommation correspond exactement au revenu, alors une année de vie supplémentaire s'écrit simplement  $u(c_T)/u'(c_T)$ . On notera que lorsque la qualité de vie  $H_T$  apparaît sous forme multiplicative, elle n'a aucune influence sur la valeur d'une année de vie en plus. En d'autres termes, si le fait d'avoir les facultés motrices réduites, abaisse de moitié la qualité de vie ( $H_T = 1/2$ ), l'utilité tirée d'une année d'existence supplémentaire est également réduite de moitié. Toutefois, il en va de même pour l'utilité marginale, en sorte que la valeur d'une année de vie, exprimée en biens de consommation, demeure inchangée.

La formule précédente se généralise lorsque le calcul porte sur un âge situé en amont dans le cycle de vie : la valeur de la vie  $VV$  est alors la somme escomptée des valeurs des différentes années de vie :

$$VV(a) = \sum_{t=a}^T \frac{S_{t,a}}{(1+\delta)^{t-a}} \left[ \frac{u(c_t)}{u'(c_t)} + (w_t - c_t) \right]$$

La valeur de la vie apparaît donc comme la somme escomptée de différents termes, chacun relatif à une période – par convention, une année. Chacun de ces termes correspond donc à la valeur d’une année de vie, comme établi précédemment.

Les différentes approches théoriques peuvent également donner lieu à diverses quantifications : après avoir calibré les modèles, différentes simulations sont effectuées. Entre autres, on a pu mesurer le gain monétaire de l’allongement de l’espérance de vie en tenant compte de son effet cumulatif (la baisse de la mortalité étant effective non seulement pour les générations courantes, mais également pour toutes les générations futures qui profiteront de ce progrès), reproduire l’évolution croissante au cours du temps de la part des dépenses de santé dans les dépenses totales (Hall et Jones, 2007), ou encore mesurer l’effet des efforts individuels de santé sur l’allongement de l’espérance de vie (Ehrlich et Yin, 2004).

## 1.2. La valeur de la vie : approche empirique

L’approche théorique présente d’incontestables mérites, comme, par exemple, la nécessité de poser la question de la spécificité du « bien santé » et son articulation avec les autres biens dont les agents tirent de la satisfaction. La démarche ne saurait toutefois s’arrêter à ce stade, car l’objectif est bien de tenter de se doter de méthodes d’évaluation économique de la santé et de la vie. Cette étape de quantification, délicate s’il en est, requiert l’usage de données, de méthodes statistiques et économétriques dont certaines sont complexes, et d’une bonne dose de précaution.

Il est possible, à l’instar de Zweifel (2009), de distinguer les méthodes d’évaluation suivant :

- qu’elles ne mettent en jeu que des grandeurs objectives et uni-dimensionnelles, comme l’espérance de vie, ou encore des variables décrivant l’état de santé des patients (pression artérielle, taux de cholestérol, etc.) ;
- qu’elles visent à définir une mesure de la qualité de vie, agrégeant les multiples dimensions associées à la santé (espérance de vie, morbidité, vue, ouïe, degré des douleurs ressenties, etc.) en un unique scalaire. On parlera alors de QALYs (*quality-adjusted life years*), sans nécessairement associer une valeur monétaire à cette qualité ;

- qu'elles poursuivent la quantification jusqu'au point de proposer une valorisation monétaire du bien santé (sur le principe de l'analyse coût-bénéfice).

La première méthode ne correspond pas, en tant que telle, à une évaluation économique de la santé. Elle se limite à comparer, sur une échelle uni-dimensionnelle, les performances de divers traitements. Si un traitement coûte plus cher et produit moins d'effets favorables qu'un ou plusieurs autres traitements, alors il sera dominé et devra être abandonné.

Les deux autres méthodes présentent un lien plus direct avec la question de la valorisation de la santé, c'est pourquoi nous limiterons notre présentation à celles-ci. La définition des QALYs est une étape importante, tentant de rendre comparables divers états de santé. Bien que la quantification monétaire ne soit pas inhérente à cette méthode, elle en est pourtant un prolongement naturel.

### 1.2.1. La construction des QALYs

L'acronyme QALY signifie *quality-adjusted life years*, soit encore la durée de vie pondérée par la qualité de la vie. L'idée sous-jacente à la construction de cet indice est de rendre comparables – non seulement en termes ordinal, mais également cardinal – différents états de santé. Cela correspond à une projection des différentes dimensions caractérisant l'état de santé d'une personne sur une échelle uni-dimensionnelle. Une fois admis le principe de cette projection, l'échelle peut être normalisée de manière à ce que l'indice de qualité se situe dans l'intervalle [0 ; 1]. À 0 est associée la qualité de vie la plus basse – en d'autres termes, la mort – et à 1 est associée la parfaite santé.

Les QALYs ont historiquement été voués à être intégrés dans des approches dites de coût-utilité (*cost-utility analysis*) : il s'agit de se doter d'une évaluation cardinale de l'utilité de la santé<sup>(4)</sup>. Un avantage substantiel à cette approche est qu'elle évite la pure quantification monétaire de la valeur de la vie humaine : elle se limite, dans son principe de construction, à comparer entre elles des situations présentant des degrés de handicap/maladie différents. Pour autant, elle n'est pas sans soulever, sinon des objections, du moins de sérieux questionnements.

Dans la pratique, la détermination des poids (QALYs situés entre 0 et 1) est opérée par des enquêtes et les modalités suivant lesquelles elles sont menées vont considérablement influencer sur les résultats obtenus (*cf.* Neumann *et al.*, 1997).

En premier lieu, on peut se poser la question de savoir qui devrait déterminer ces poids : faut-il s'en remettre aux patients, atteints des maladies/handicap dont on cherche à évaluer le coût en termes de qualité de vie ? Ou bien convient-il d'interroger toute la population, y compris ceux qui

(4) Il s'agit d'un cas particulier des analyses dites de coût-efficacité.



ne sont pas atteints ? Ou devrait-on enfin se limiter aux professionnels de santé, qui ont un regard expert sur les conditions de vie des malades ? Le résultat risque de fortement différer selon le choix effectué, avec certains biais systématiquement observés. Par exemple, lorsqu'on demande à des bien-portants d'évaluer le coût subjectif à perdre l'usage de leurs membres inférieurs et à se déplacer en fauteuil roulant, ils associeront une fraction  $\alpha$  à cet état ( $\alpha = 0,75$ , par exemple, signifiant que la qualité de vie en fauteuil roulant vaut les trois quarts de la qualité de vie en parfaite santé). En revanche, lorsque la même question est posée aux personnes subissant effectivement ce handicap, ils évaluent leur qualité de vie, presque systématiquement (voir, par exemple, Cutler et Richardson, 1997) à une fraction sensiblement supérieure à  $\alpha$ . Cela semble finalement très compréhensible : des bien-portants redoutent fortement le fait de perdre l'usage de leurs membres inférieurs, mais si cela devait se produire, ils s'y accoutumeraient. Si la différence entre ces deux valeurs se comprend aisément, la question du « bon » poids à retenir demeure entière : doit-on considérer le QALY subjectif des bien-portants, avec dans l'idée qu'on maximise l'utilité de toute la population avant la réalisation des dommages – à l'instar du choix social sous le voile de l'ignorance –, c'est-à-dire *ex ante* ? Ou bien doit-on tenir compte de l'évaluation *ex post* ? Cette question semble loin d'être définitivement tranchée.

En deuxième lieu, il existe différentes méthodes pour déterminer les QALYs. On peut proposer la méthode de la loterie standard, du taux de substitution temporel, ou simplement de l'évaluation subjective. Dans le premier cas, on propose à la personne sondée de choisir entre se trouver de manière permanente dans un état (maladie, handicap) donné, et accepter la loterie suivante : avec la probabilité  $\pi$ , la personne continue de vivre en parfaite santé, et avec la probabilité  $1 - \pi$ , elle décède instantanément. En tâtonnant sur la probabilité  $\pi$ , il est possible d'aboutir à une situation où l'agent est exactement indifférent entre être durablement affecté par la maladie/le handicap considéré, et jouer à la loterie risquée. Si, par exemple,  $\pi = 0,8$ , alors cela signifie que la personne évalue le handicap à 80 % de la vie en parfaite santé : le handicap coûterait autant à l'agent que le fait de risquer de décéder immédiatement avec une probabilité de 20 %. Dans le deuxième cas, on propose à la personne sondée d'échanger une année de vie avec le handicap/la maladie contre une fraction  $\pi$  d'une année en parfaite santé. On ajuste  $\pi$  jusqu'à aboutir à l'indifférence de la personne interrogée. Si une personne sondée évalue à  $\pi = 0,6$  la perte de la vue, cela signifie que vivre une année supplémentaire en étant aveugle lui est équivalent à vivre 60 % d'une année avec une bonne vue. Selon la troisième méthode, on demande directement au sondé ce que vaut, selon sa propre estimation subjective, la vie avec le handicap, en proportion de la vie en parfaite santé.

Une fois qu'à chaque état de santé est associé un poids compris entre 0 et 1, il est aisé de calculer le gain en QALY de tel ou tel traitement, permettant d'être débarrassé temporairement ou définitivement d'une pathologie, ou

bien de surmonter partiellement un handicap (cas des prothèses articulaires par exemple). Ainsi, si la pose d'une prothèse de hanche élève le QALY par année de vie de 0,7 à 0,85, le gain vaut  $0,85 - 0,7 = 0,15$  par année de vie. Le gain total en QALY s'obtient en multipliant ce résultat par l'espérance de vie de la personne considérée.

Ces différentes approches reposent sur l'idée selon laquelle la qualité et la quantité de vie sont comparables, puisqu'on y propose d'échanger des situations qui diffèrent aussi bien par la qualité que la durée de vie. Or, cela n'a rien d'évident, et certaines personnes pourraient préférer systématiquement l'option qui maximise la durée de vie restante, quel que soit l'état de santé. Par ailleurs, les fondements théoriques sur lesquels repose la construction des QALYs sont ceux de l'utilité espérée. En particulier la neutralité par rapport à la durée de vie est nécessaire. En d'autres termes, pour que l'axiomatique soit applicable, il faut que les agents soient indifférents entre connaître avec certitude leur durée de vie restante (ce qui implique connaître avec exactitude la date de leur mort) et disposer d'une espérance de vie de même durée (mais alors avec une date de mort aléatoire). Or, la plupart des sondés préfèrent nettement qu'on leur offre  $X$  années d'espérance de vie, avec une date de décès incertaine, plutôt que de disposer d'un compte-à-rebours précis de  $X$  années.

Ainsi, la construction et l'utilisation des QALYs ne sont pas immédiates. Pourtant, ils sont largement utilisés, dans certains cas par les autorités de santé, comme nous le verrons en section 2.

Une fois les QALYs déterminés, le processus de choix collectif peut prendre de nombreuses formes, dont la description sort du cadre de cette présentation. Comme les QALYs ne définissent que des valeurs relatives d'état de santé, ils ne sont pas mesurables en unités monétaires. Pour les utiliser, il faut disposer en amont d'un budget pour les dépenses globales de santé. Étant donné ce budget, les pouvoirs publics peuvent souhaiter maximiser le volume total de QALYs produits, c'est-à-dire les gains en qualité de vie engendrés par les dépenses de santé consenties. En effet, les QALYs possèdent cette propriété qu'étant parfaitement homogènes entre eux, il est *a priori* légitime de les additionner les uns aux autres. La question de la distribution des gains en QALY se pose : le fait de maximiser le volume total de QALYs gagnés risque fort de laisser de côté les malades atteints de pathologies peu fréquentes, qui nécessiteraient des dépenses conséquentes, et dont les bénéfices seraient répartis sur un faible nombre de personnes. Le choix collectif peut également intégrer de l'aversion aux inégalités, dont la forme extrême est le critère du maximin. Selon ce principe, la répartition des dépenses de santé est choisie de manière à maximiser le plus faible gain en QALY ressenti dans la population. Cette seconde étape correspond donc à l'utilisation des QALYs dans le processus de choix social (ou collectif), qui sera plus largement abordé dans la suite de ce document (*cf.* section 2).

### 1.2.2. La mesure de la valeur monétaire de la vie

L'étape ultime de cette démarche est celle de la quantification, en termes monétaires, de la valeur de la vie humaine. Celle-ci suscite des controverses fondées sur des jugements éthiques qui sortent du cadre de l'analyse économique, et dont nous ne traiterons pas. Évidemment, il ne s'agit pas d'associer à telle ou telle personne précisément une valeur monétaire, pour ensuite comparer cette valeur au coût du soin qui lui est nécessaire.

Dénommée *Value of a statistical life* (VSL) en anglais, il s'agit bien de la valorisation en termes monétaires de la valeur de la vie humaine en général. Une mesure peut-être plus aisée à manipuler est la valeur d'une année de vie. En effet, la valeur de la vie statistique représente la valeur d'une vie moyennée sur une population donnée. Or, l'espérance de vie d'une personne très âgée, proche de l'échéance finale, est faible, en sorte qu'on peut concevoir que la valeur de la vie associée soit plus faible que celle d'un adolescent. Évidemment, cela suppose que l'on dépasse le principe selon lequel « la vie n'a pas de prix », faute de quoi aucune comparaison et aucune quantification, ne seraient possibles.

On définit alors la « valeur de la vie statistique »<sup>(5)</sup>, car cette valeur est déterminée en se référant à diverses situations, caractérisées par des niveaux de risque de mortalité différents. Il n'est en effet pas possible de calculer la valeur absolue de la vie, comme réponse par exemple à la question : « quelle somme seriez-vous prêt(e) à donner pour éviter une mort certaine et imminente ? ». À cette question, un grand nombre d'individus répondraient qu'ils abandonneraient volontiers toutes leurs possessions pour éviter cette issue tragique. Bien qu'il soit concevable que la valeur de la vie dépende en partie de caractéristiques économiques comme la productivité ou le revenu des agents, une corrélation si forte entre le patrimoine et la valeur de la vie suggère que la question n'est pas adéquate. Plutôt, on demandera aux personnes interrogées, ou on inférera à partir de leurs comportements, quelle somme ils seraient prêts à consacrer pour une réduction du risque de mortalité de  $x$  %. Les comportements de demande de biens de sécurité, comme les protections dans les automobiles (airbags, assistance au freinage) en sont un exemple évident : ces protections ont un prix aisément identifiable, et permettent d'abaisser la probabilité d'être gravement blessé. Le consentement à acheter à un prix plus élevé un bien immobilier davantage éloigné d'une source de pollution dangereuse (proximité avec un site productif polluant, une centrale nucléaire) en est un autre. Ces comportements sont donc relativement courants, et ne choquent pas, car on échange une loterie risquée contre une loterie moins risquée, et non pas la vie contre la mort.

Ainsi, si l'on peut quantifier à  $P$  le consentement à acheter un bien qui réduit d'un point la probabilité de décéder, la valeur de la vie statistique se calculera simplement comme  $100P$ . Clairement, cette valeur est locale et porte, le plus souvent, sur des probabilités très faibles : la probabilité de décéder à la suite d'un accident automobile sur une année est basse, donc

(5) Et non pas valeur statistique de la vie, comme on pourrait le penser de prime abord.

l'effet de la présence d'une protection supplémentaire consistera à l'abaisser, soit une variation du risque très faible. Rien n'assure que la valeur du surcroît de 1 % de chance de survie soit indépendante du taux de mortalité considéré dans la situation de référence<sup>(6)</sup>.

Toutefois, on remarquera que si l'unique façon d'appréhender au niveau microéconomique la valeur de la vie est statistique – ou plutôt, probabiliste –, ce concept s'accommode très bien du passage au niveau macroéconomique. Les choix de société, notamment la part du PIB globalement consacrée aux dépenses de santé, ont des effets qui peuvent présenter des régularités statistiques. On peut, par exemple, calculer de manière assez fiable, l'effet du développement d'actes comme les pontages, ou la pose de stents, sur le taux de mortalité de cohortes précisément dénombrées (*cf.* Cutler et McClellan, 2001) et même sur l'accroissement de l'espérance de vie de ces cohortes. Dès lors, on peut mesurer l'effet d'une dépense supplémentaire de *X* milliards d'euros, sur la probabilité de décès d'une sous-population donnée. La présence d'externalité peut certes créer une divergence entre la valeur privée de la vie statistique et la valeur sociale, mais les principes gouvernant sa quantification demeurent très proches.

Cette relation statistique entre dépense globale et risque de mortalité n'est bien sûr pas propre à la santé, et s'applique à d'autres domaines, au premier rang desquels on trouve la sécurité routière. La construction d'infrastructures plus sûres, ainsi que l'abaissement des vitesses maximales autorisées, permettent de réduire la mortalité routière, mais engendrent également des coûts. C'est pourquoi le concept de valeur statistique de la vie humaine est également utilisé dans ce cas, et a même donné lieu à une démarche de définition d'un référentiel destiné à limiter l'hétérogénéité des évaluations de la valeur statistique de la vie (*cf.* Baumstark et *al.*, 2008).

Sans nullement prétendre à l'exhaustivité, ni à une présentation avec un haut degré de détail<sup>(7)</sup>, nous rappelons brièvement le type de démarche empirique adoptée pour déterminer cette valeur de la vie statistique. Précisément, on peut distinguer deux types de démarche : celles fondées sur les préférences révélées, et celles qui reposent sur des enquêtes.

#### 1.2.2.1. La mesure de la valeur de la vie statistique à partir des préférences révélées

On se fonde ici sur le comportement effectif des agents, dans certaines situations où leurs comportements font apparaître un arbitrage entre le degré de risque et un flux monétaire (reçu ou bien payé). Parmi celles-ci, on trouve en premier lieu les équations hédoniques de salaire, qui estiment le salaire en fonction de différentes caractéristiques de l'emploi considéré, dont le degré

(6) Pour l'illustrer, nous pouvons revenir au cas d'une personne à l'article de la mort, à qui l'on proposerait d'échanger la morte certaine (probabilité de décès = 100 %) avec la loterie telle qu'on décède avec une probabilité de 99 %, et l'on survit avec une probabilité de 1 %. Dans ce cas, la valorisation de la possibilité de ne pas décéder avec certitude devrait être très élevée.

(7) Nous renvoyons les lecteurs cherchant davantage de précisions aux revues de la littérature de Viscusi et Aldy (2003) ou de Viscusi (1993).

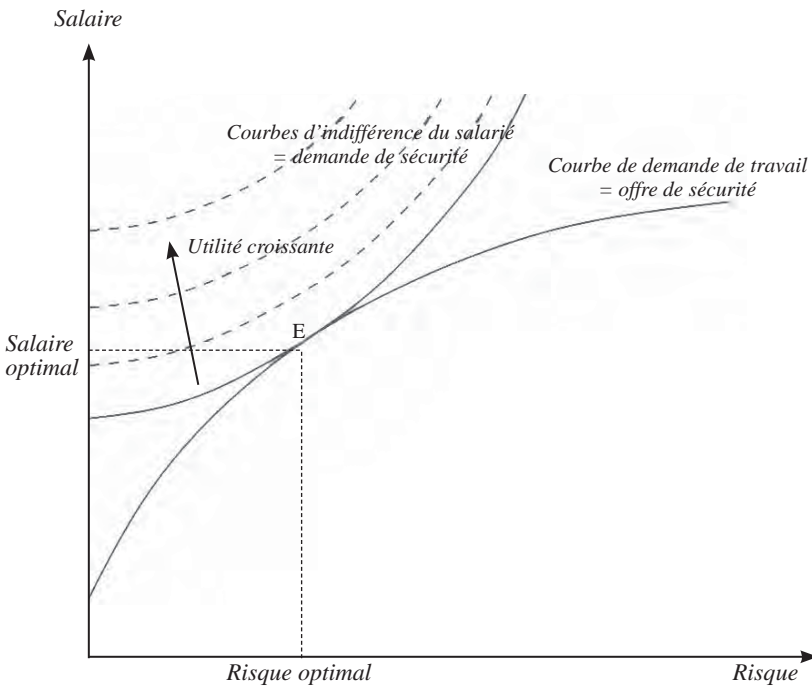
d'exposition au risque, qu'il soit mortel ou non. Ces études se concentrent, pour des raisons évidentes, sur des secteurs d'activité où le risque associé à l'emploi est non négligeable (industrie chimique, par exemple). La forme générale de l'équation est :

$$w_i = \alpha + \sum_{m=1}^M \psi_m x_{im} + \gamma_0 p_i + \gamma_1 q_i + \gamma_2 RR_i + u_i$$

où  $w_i$ ,  $x_{im}$ ,  $p_i$ ,  $q_i$ ,  $RR_i$  et  $u_i$  désignent respectivement le salaire à expliquer de l'agent  $i$ , ses caractéristiques personnelles ainsi que les caractéristiques de l'emploi occupé, le taux de mortalité au travail, la probabilité d'accident non mortel au travail, le revenu de remplacement en cas d'accident du travail et le terme d'erreur.  $\alpha$ ,  $\psi_m$ ,  $\gamma_0$ ,  $\gamma_1$  et  $\gamma_2$  désignent des constantes.

Le coefficient  $\gamma_0$  associé à la variable explicative  $p_i$  mesure le prix exigé par les travailleurs pour accepter une unité de risque supplémentaire. Cette formulation suppose que les travailleurs expriment des préférences au regard de la sécurité et du salaire, et qu'un arbitrage entre ces deux attributs de l'emploi est possible. La figure ci-dessous représente une courbe d'indifférence du travailleur.

### Représentation de l'offre et de la demande de sécurité



Source : Auteurs.

En particulier, le coût associé à un accroissement du risque est lui-même croissant, d'où la convexité de la courbe. Du côté de la demande de travail (qui est de manière équivalente, la courbe d'offre de sécurité), on considère que les firmes sont également capables de réduire le risque associé à un emploi, en contrepartie de quoi elles répercutent sur le salaire versé tout ou partie du coût associé à cette sécurité accrue. La courbe de demande de travail est elle aussi croissante avec le risque, mais elle est concave : en partant d'une situation caractérisée par un risque élevé, le coût à réduire à la marge le risque est limité, en sorte que la baisse du salaire proposé est faible. À mesure que le risque s'abaisse, le coût marginal de réduction du risque s'élève, et la baisse du salaire proposé s'en trouve accrue. Le choix optimal du travailleur se trouve au point de tangence entre la courbe de demande de travail et la courbe d'indifférence. Ce point de tangence dépend à la fois des caractéristiques propres à l'emploi proposé, ainsi qu'aux préférences du travailleur. Le prix du risque est la pente au point de tangence.

Parmi les écueils à éviter lors de l'estimation du prix associé au risque de mortalité, on trouve :

- la difficulté de séparer le risque de mortalité du risque d'accident ; ces deux risques sont susceptibles de présenter une forte corrélation et de rendre ardue leur prise en compte simultanée. Si l'on omet de distinguer les deux termes, on risque d'accorder un poids excessif à la valeur mesurée de la vie, qui mêlerait alors l'effet d'un accroissement du taux de survie et d'une baisse du risque d'accident simple (non mortel). Le surcroît de salaire exigé pour une hausse marginale du risque d'accident non mortel apparaît dans l'équation ci-dessus au travers du coefficient  $\gamma_1$  ;
- la difficulté de s'assurer que les travailleurs mesurent correctement les risques associés à leur activité. On a pu mettre en évidence un biais psychologique, selon lequel les agents avaient tendance à surestimer les risques très faibles, et à sous-estimer les risques plus forts. Cela implique que les agents sous-estiment fréquemment tout accroissement marginal du risque auquel ils sont soumis ;
- la difficulté de disposer de données suffisamment précises sur le risque associé à un poste précis. Dans de nombreuses études, seul le secteur d'activité est connu, ce qui représente une perte considérable d'information ;
- la méthode des préférences révélées peut également être appliquée, entre autres, à l'achat de détecteurs de fumée dans des maisons en bois, à l'usage de la ceinture de sécurité dans les automobiles, ou encore à l'utilisation des couloirs souterrains pour traverser des axes à circulation dense.

Finalement, ces diverses études aboutissent à des valorisations de la vie humaine qui diffèrent sensiblement entre elles, mais qui demeurent cependant dans des ordres de grandeur semblables. Viscusi (1993), qui présente les résultats des principales études, suggère que l'intervalle de 3 à 7 millions de dollars pour la valeur d'une vie humaine, semble émerger. Ramené en

valeur d'une année de vie (et escomptée sur la durée de vie restante espérée), cela revient à considérer qu'une année de vie vaut approximativement 100 000 dollars. Le tableau 1 présente les résultats de plusieurs estimations de la valeur de la vie statistique, dont certaines sont fondées sur des arbitrages sans lien avec le marché du travail.

### 1. Valeur de la vie statistique pour différents risques

	Pays	Type de risque étudié	Valeur statistique de la vie (en millions de dollars)
Thaler et Rosen (1975)	États-Unis	Marché du travail	1,0
Leigh et Folsom (1984)	États-Unis	Marché du travail	1,01-13,3
Moore et Viscusi (1990)	États-Unis	Marché du travail	20,8
Siebert et Wei (1994)	Royaume-Uni	Marché du travail	9,4-11,5
Miller, Mulvey et Norris (1997)	Australie	Marché du travail	11,3-19,1
Blomquist (1979)	États-Unis	Accidents automobiles/ usage de la ceinture de sécurité	1,0
Garbacz (1989)	États-Unis	Achat de détecteur de fumée	2,56
Jenkins, Owen et Wiggins (2001)	États-Unis	Achat de casque de protection pour cyclistes	1,4-4,3

Source : Viscusi et Aldy (2003).

#### 1.2.2.2. La méthode de l'évaluation contingente

Le principe de l'évaluation contingente repose sur des enquêtes au cours desquelles le répondant se voit demander d'évaluer en termes monétaires le fait de disposer de tel ou tel bien. Couramment pratiquée pour valoriser les biens environnementaux, l'évaluation contingente donne également lieu à des applications dans le cas du « bien santé ». On peut distinguer différentes méthodes :

- les techniques dites ouvertes (*open-ended techniques*) consistent à demander directement aux individus le montant maximal qu'ils seraient prêts à dépenser pour obtenir ou conserver un bien, un état particulier (santé, biens environnementaux). Parfois, on propose des montants aux personnes interrogées, et on l'accroît tant qu'ils se montrent prêts à déboursier pareille somme, jusqu'à atteindre le montant maximal supporté ;
- les techniques dites fermées (*closed-ended techniques*) reviennent à demander de répondre par oui ou par non à la question : accepteriez-vous de payer tel prix pour obtenir le bien en question ? On ne cherche alors pas à trouver le prix maximal pour chaque agent, mais

on balaye une grille de prix prédéfinie. La répartition des oui-non permet de construire une courbe de demande, en reliant le prix avec la quantité demandée (soit la proportion de réponses favorables).

Cette méthode permet en théorie de calculer la valeur d'un état de santé en tant que tel, et non pas uniquement la valeur de la variation du risque de se retrouver dans un état de santé (on demande au sondé de comparer deux situations très différentes : la première, où il demeure en bonne santé, et la seconde où il est victime de telle ou telle affection). La plupart des études, toutefois, posent la question, plus aisée à se représenter pour les personnes sondées, de la valeur qu'ils accordent à la baisse du risque de voir la santé se dégrader : il s'agit donc de l'évaluation de l'effet d'une variation marginale, ou faible, du risque de morbidité, et non pas la comparaison de deux états de santé. À cet égard, cette méthode pose le problème de la prise en compte subjective des probabilités faibles. Quand on demande à une personne la somme qu'elle serait prête à déboursier pour réduire son risque d'accident automobile de 0,001 %, soit une baisse du risque de 1 pour 100 000, il est fréquent que la personne sondée annonce un montant qui diffère à peine de celui qu'elle déclarerait, si la réduction du risque s'élevait à 0,1 %. En effet, les individus ont tendance à surpondérer les probabilités très faibles. Ainsi, les différentes études aboutissent à des résultats dont les ordres de grandeur peuvent différer significativement.

### 1.2.3. QALYs et valeur économique de la vie humaine

La distinction entre les QALYs, qui comparent entre eux différents états de santé sans forcément leur attribuer une valeur monétaire, et la valeur de la vie statistique, peut assez naturellement être dépassée. En effet, une fois que la valeur d'une année de vie humaine est déterminée, on peut considérer qu'elle correspond à la valeur d'un QALY. Un traitement de santé élevant la qualité de vie pendant une année de 0.1 QALY, se verra donc attribuer une valeur économique correspondant à 10 % d'un QALY (soit 10 000 euros si l'on se réfère à la valeur communément admise d'une année de vie en bonne santé). Ainsi, il est possible de définir la valeur de chaque traitement, dans l'optique de la comparer à son coût. Notons que l'application de cette règle – comparaison du coût objectif des soins au gain en QALY évalué en unités monétaires – n'est pas en soi compatible avec la contrainte d'une enveloppe globale de dépenses de soins donnée, car alors le pilotage du système serait surdéterminé<sup>(8)</sup>. Enfin, comme cela a déjà été précisé précédemment, et comme cela a dû apparaître clairement dans cette présentation, cette évaluation est fondée sur l'évaluation individuelle de la vie humaine. La propriété de fongibilité des QALYs (le fait qu'on peut les additionner les uns avec les autres, même s'ils correspondent à différents agents) a uniquement été mentionnée ; elle est conceptuellement possible, mais cela ne justifie nullement qu'on applique ce principe sans autre précaution. Cela relève déjà d'une démarche de choix social sur laquelle nous allons maintenant nous pencher.

(8) Dans la pratique, la pure application de la tarification à l'activité (T2A) peut donner lieu à des dépassements conséquents des dépenses par rapport aux prévisions. Des mécanismes d'ajustement doivent alors être implémentés (respect de l'ONDAM).



## 2. Valeur de la vie humaine et choix social en matière de dépenses de santé

La valeur d'une année de vie (en bonne santé) peut être utilisée comme métrique pour guider la régulation publique du système de santé en traduisant en surplus monétaire les bénéfices des traitements médicaux. C'est sur cette idée que nous souhaitons nous attarder maintenant.

### 2.1. Coût des soins, valeur de la vie et choix public

Le système de santé français garantit la socialisation du coût des traitements médicaux dans le périmètre d'un panier de soins préalablement circonscrits. Cette délimitation du champ d'action de l'assurance maladie relève d'un choix public engageant l'ensemble des assurés sociaux<sup>(9)</sup>. Pour un traitement donné, il serait naturel, pour un économiste tout au moins, de faire reposer cette décision sur une confrontation de l'efficacité du traitement, en termes d'amélioration du bien-être, à son coût, comme un individu confronterait l'utilité qu'il retire d'une consommation à son coût<sup>(10)</sup>.

En France, le processus institutionnel de décision qui conduit à la prise en charge d'un traitement par l'assurance-maladie ne découle pas directement d'un tel arbitrage. Accordons-nous un détour pour mieux comprendre ce processus sans pour autant entrer trop avant dans les détails institutionnels. Au regard d'éléments cliniques, le service médical rendu (SMR) est évalué par la HAS (Haute autorité de santé) et le remboursement accordé si le SMR est jugé suffisant<sup>(11)</sup>. Si le constat est fait d'une amélioration du SMR (ASMR) alors une augmentation du prix (par rapport au traitement de référence) est possible<sup>(12)</sup>. La négociation sur les prix, conduite par le CEPS (Comité économique des produits de santé) et qui suit la décision du remboursement, n'est pas indépendante de l'appréciation de l'ASMR par la HAS ce qui introduit de fait un arbitrage entre coût et efficacité. Mais cette négociation est théoriquement indépendante de celle de la décision du remboursement. Elle n'est pas fondée sur un rapprochement systématique entre le coût du traitement et le gain social exprimé en années de vie sauvées.

Une décision récente de la Haute autorité de santé montre pourtant que la décision du remboursement d'un traitement n'est peut-être pas toujours étrangère à ce type d'arbitrage. En 2010, une molécule anticancéreuse,

(9) Il faut noter que le système français a la particularité de proposer des taux de remboursement qui varient en fonction du traitement. Nous ne nous attarderons pas ici sur cette spécificité en évoquant le remboursement des traitements sans plus de précision.

(10) Notons enfin que, par souci de clarté et de simplicité, nous avons volontairement ignoré les arguments de nature prospective qui pourrait justifier le remboursement d'un traitement coûteux en considérant que sa diffusion dans la pratique médicale est la condition, à terme, de développement de thérapies efficaces et moins coûteuses.

(11) Le taux de remboursement peut être modulé en fonction du SMR.

(12) L'absence d'ASMR exige en revanche une baisse du prix.

l'erlotinib, prescrit pour les cancers broncho-pulmonaires a vu son AMM (autorisation de mise sur le marché) être étendue pour le cancer du pancréas métastatique. La prise en charge par l'assurance-maladie a été refusée pour cette indication<sup>(13)</sup> alors que l'efficacité de la molécule a été reconnue : elle prolongerait de deux à trois mois l'espérance de vie des patients<sup>(14)</sup>. On peut s'interroger sur cette décision et se demander si le jugement n'a réellement porté que sur l'appréciation de l'ASMR et non pas sur un rapprochement de l'ASMR et du coût très élevé du traitement. La décision n'aurait-elle pas été différente si le coût avait été moins moindre<sup>(15)</sup> ?

Cette décision illustre la difficulté, en France notamment, à considérer explicitement le coût des soins dès lors qu'il est question de prolonger la vie des patients. Il est pourtant nécessaire de différencier l'injonction faite à tous, et aux médecins en particulier, de mettre tous les moyens à disposition pour secourir un patient, du choix public décidant des moyens dont disposent les médecins. Les évaluations coût-utilité reposant sur une évaluation monétaire des années de vie gagnées peuvent aider à éclairer ce choix comme nous allons en discuter maintenant.

## **2.2. La valeur monétaire des années de vie gagnées comme critère de choix**

Pour mesurer l'opportunité du remboursement d'un traitement médical, il est possible, comme pour tout investissement public, de faire appel aux méthodes usuelles de l'évaluation des choix publics en confrontant le coût de ce traitement au bénéfice social attendu. Ce dernier peut être apprécié par le nombre de QALYs (années de vie en bonne santé)<sup>(16)</sup> gagnés et traduit en bénéfice monétaire en multipliant ce nombre par une estimation annuelle de la valeur de la vie. Il s'agit en définitive d'appliquer le critère du surplus global après avoir évalué le bénéfice du traitement en termes de QALYs gagnés, ce qui revient à confronter le coût par QALY d'un traitement particulier à la valeur monétaire d'une année de vie en bonne santé. Ce type de calcul peut également être mis en œuvre à l'échelle des dépenses de santé

(13) Dans les faits, il est plus que probable qu'aucun patient atteint d'un cancer du pancréas avancé n'a jamais payé ce médicament de sa poche car l'ordonnance que fait le médecin spécialiste ne mentionne aucune indication de diagnostic.

(14) Au prix d'effets indésirables fréquents et sévères.

(15) Il est à noter que le Conseil d'Etat (décision du 12 mai 2010) a validé l'appréciation de la HAS sur le service médical rendu par cette molécule.

(16) Comme détaillée plus haut, les QALYs permettent de mesurer, sur une échelle unique, gain en espérance de vie et gain en qualité de vie. D'autres métriques peuvent être mobilisées, notamment les échelles de qualité de vie utilisées en santé publique (le SF36, par exemple, voir McHorney et al., 1993). Par souci de clarté et de simplicité nous utilisons la métrique des QALYs comme référence. Cela ne réduit pas la portée générale des commentaires.

Ce détour par l'évaluation des QALYs n'est pas une nécessité théorique mais une commodité pratique qui permet de normaliser et de faciliter l'évaluation des bénéfices retirés des dépenses de santé.

dans leur globalité pour s'interroger sur l'efficacité du système de santé dans son ensemble<sup>(17)</sup>.

Revenons ainsi un instant à la question du remboursement de l'anticancéreux évoquée plus haut. Admettons que la décision de ne pas les rembourser relève implicitement d'une décision en termes de coût par QALY. Elle peut se justifier au regard du coût très élevé des soins et du faible gain en espérance de vie. Il est cependant nécessaire de manier avec précaution la métrique des QALYs. Permettre à une personne brutalement atteinte d'un cancer du pancréas de vivre quelques mois supplémentaires, ce n'est pas seulement augmenter son espérance de vie c'est aussi lui permettre d'organiser sa fin de vie. Dis autrement, l'hypothèse de linéarité qui préside à la valorisation monétaire des gains en QALYs<sup>(18)</sup> est fort contestable comme l'ont argumenté Philippon et *al.* (2010). Il est bien sûr possible de raffiner la démarche en s'appuyant sur des estimations monétaires individuelles des gains en espérance de vie mais au prix d'une complexité sensiblement accrue de la méthode.

Interrogeons-nous maintenant plus globalement sur la pertinence de l'application d'une analyse en termes de surplus global au cas de l'évaluation d'un traitement médical.

On sait, depuis longtemps, que le critère du surplus global ne permet pas de construire un choix social cohérent dans le sens où il ne permet pas de construire un classement univoque entre les différents investissements en santé<sup>(19)</sup>. Cette critique porte sur la méthode en général du surplus global, sa portée n'est pas plus large dans le cadre de l'évaluation des traitements médicaux que dans tout autre cadre.

Une autre critique faite à la méthode du surplus global est relative à l'absence de prise en compte de l'inégale distribution des niveaux de bien-être individuels. Cette critique est tout aussi général mais elle se trouve être particulièrement dérangeante dans le contexte sanitaire. L'évaluation monétaire des gains en santé d'un traitement donné se définit usuellement, nous l'avons dit, comme le produit des QALYs gagnés par la valeur d'une année de vie en bonne santé<sup>(20)</sup>. En conséquence, les dépenses permettant des petits gains en qualité de vie (ou en espérance de vie) pour une très vaste

---

(17) Voir, par exemple, Cutler et *al.* (2006) pour une évaluation sur les Etats-Unis.

(18) Le calcul du surplus global, comme produit du nombre de QALY gagnés et de la valeur monétaire d'un QALY, présuppose en effet que le valeur d'un QALY soit constante et donc indépendante du nombre de QALY gagnés (et indépendante des caractéristiques de l'individu en bénéficiant).

(19) Voir, par exemple, Boadway et Bruce, 1984.

(20) On voit ici que la médiation du calcul par l'évaluation du nombre de QALYs gagnés nous éloigne quelque peu de la méthode standard du surplus global (défini par la somme des surplus individuels). Précisément appliqué, cette méthode imposerait d'évaluer monétairement, pour chaque individu, le gain en espérance de vie en bonne santé du traitement considéré et de faire la somme des évaluations individuelles.

population seront mécaniquement privilégiées par rapport à des dépenses conduisant à des gains importants pour une partie très restreinte de la population. Une expérience célèbre est celle de l'Oregon, État américain qui, dans le cadre d'une politique de rationnement des soins, dans les années quatre-vingt-dix, a établi une liste de priorité des traitements en fonction du coût par QALY<sup>(21)</sup>. Ce mode de gestion du rationnement des soins, conduit mécaniquement à privilégier les traitements de pathologies mineures touchant une très large population aux dépens des traitements de pathologies lourdes et à prévalence réduite. Ce qui mène à des refus de soins pour des patients dans un grand besoin.

L'évaluation monétaire moyenne des QALYs gagnés, comme la méthode du surplus global en général, ne rend pas compte de la priorité que l'on peut légitimement donner aux individus les plus défavorisés<sup>(22)</sup>.

Bien entendu, la théorie du choix social n'interdit pas d'intégrer cette priorité à l'action publique. Il suffit pour cela de mobiliser des fonctions de bien-être social autorisant une pondération privilégiant les situations individuelles les plus défavorables. Ceci suppose cependant d'ordonner les niveaux de bien-être individuel afin d'affecter les poids en fonction d'un ordre de priorité. Ceci relève d'un autre type de méthode que celle mise en œuvre pour l'évaluation monétaire des QALYs. La valeur monétaire des QALYs est en effet une mesure de la *variation* globale du surplus (de *l'amélioration* du bien-être) qui ne nous renseigne en rien sur les *niveaux* d'utilité ou de bien-être individuels.

Des travaux récents, fondés sur une nouvelle lecture normative de la notion de revenu équivalent, proposent une mesure des niveaux d'utilité dans un cadre où l'état de santé caractérise, avec le revenu, les situations individuelles (Fleurbaey et Schokkaert, 2009). L'idée est d'indicer l'utilité d'un individu par le revenu, appelé revenu équivalent, qui laisserait son utilité inchangée s'il pouvait bénéficier d'un bon état santé (le revenu équivalent est donc égal au revenu courant si l'individu en question est effectivement en bonne santé). La distribution des revenus équivalents ainsi obtenue permet de prendre en considération les aspects distributifs des politiques de santé. Une étude empirique pilote sur données françaises, utilisant la méthode d'évaluation contingente pour estimer les revenus équivalents, est aujourd'hui très avancée (Fleurbaey et al., 2012). Les questions méthodologiques et conceptuelles soulevées par une telle démarche demeurent encore nombreuses. Elle constitue cependant une méthode alternative intéressante qui permet notamment de mettre en évidence les manquements, du point de vue notamment de la prise en considération des aspects distributifs des dépenses de santé, des méthodes coût-utilité reposant sur l'évaluation monétaire des QALYs.

(21) Voir Stevenson, 1991, pour une présentation critique de l'expérience.

(22) Notons ici que les raffinements de la méthode du surplus global visant à tenir compte des inégalités en pondérant les surplus individuels par les revenus est inopérante quand il s'agit de prendre en considération des inégalités liées aux états de santé.

Cette critique peut être élargie à la question des inégalités de revenu. Considérons que l'efficacité d'un nouveau traitement soit telle que la valeur monétaire des QALYs gagnés dépasse le coût du traitement et qu'il soit donc décidé d'autoriser son remboursement. Dans un contexte où les dépenses de santé sont financées par cotisations sociales (ou prélèvements proportionnels au revenu), il est possible que la population pauvre, qui valoriserait moins les avantages de ce traitement que la moyenne de la population<sup>(23)</sup>, pâtisse de son financement par l'assurance-maladie. On voit bien ici que la valeur monétaire des QALYs est un critère de choix public contestable du point de vue distributif dans la mesure où il peut conduire à une dégradation du bien-être des individus les plus défavorisés. Notons que la tension entre un objectif d'égal accès aux soins (qui suppose une large socialisation des dépenses de santé) et une évaluation minutieuse des dépenses de santé est accentuée par une démarche *welfariste* attentive aux effets distributifs des dépenses de santé. En effet, l'hypothèse d'un faible consentement à payer des ménages pauvres pour des améliorations marginales de l'espérance de vie ou de la qualité de vie doit conduire à des évaluations plus exigeantes des dépenses de santé. Un argument paradoxal en faveur d'une évaluation par le critère de la valeur monétaire des QALYs est qu'il conduit au maintien du financement public de traitements qu'une démarche *welfariste* pourrait rejeter. Défendre cet argument présuppose cependant d'attribuer une valeur en soi à l'égal accès financier aux soins et donc de se placer en dehors de la logique *welfariste* qui préside à la méthode du revenu équivalent.

Il est utile dans un dernier temps de s'interroger sur la réalité des choix publics en matière sanitaire en s'intéressant à la part de la richesse nationale que chaque pays consacre aux dépenses de santé.

### **2.3. La part des dépenses de santé dans le PIB : le reflet de la valeur de la vie ?**

La part des dépenses de santé dans le produit intérieur brut est assez sensiblement variable au sein des pays développés comme l'illustre le tableau 2.

Dans une certaine mesure, l'importance relative des dépenses de santé rend compte de la valeur donnée par la société à l'espérance de vie (en bonne santé). Dans une certaine mesure seulement puisque, comme nous l'avons vu, des considérations distributives peuvent guider le niveau des dépenses de santé dès lors que l'on adopte un cadre d'analyse *welfariste*. Rappelons aussi que le niveau de socialisation des dépenses est variable entre pays et surtout que les prix des biens de santé sont différents d'un pays à l'autre ce qui constitue deux autres explications majeures aux divergences constatées sur le tableau 2<sup>(24)</sup>. Notons également qu'il est raisonnable

(23) On peut en effet penser qu'une augmentation marginale de l'espérance de vie – ou de la qualité de vie – soit faiblement valorisée par les populations les plus démunies.

(24) On sait, par exemple, que la situation exceptionnelle des États-Unis tient pour bonne partie aux niveaux très élevés des prix des biens de santé.

de supposer que la valeur monétaire de la vie croît avec le revenu, ce qui supposerait que la part des dépenses de santé dans la richesse d'un pays soit fonction croissante de cette dernière (voir Hall et Jones, 2007).

## 2. Dépenses de santé (2009)

*En pourcentage du PIB*

France	Royaume-Uni	Allemagne	États-Unis	Japon <sup>(*)</sup>	Suède
11,8	9,8	11,5	17,4	8,5	10,0

Note : (\*) Données 2008.

Source : OCDE.

On peut aussi comprendre ces différences entre pays comme le reflet de divergences sur la place donnée, dans le processus de décision des autorités de santé, à la confrontation *explicite* des gains en utilité au coût des traitements médicaux. Le cas du Royaume-Uni, dont les dépenses sont, en termes relatifs, parmi les plus faibles des pays développés, est emblématique sur ce point. Le NICE (National Institute for Health and Clinical Excellence)<sup>(25)</sup>, fonde explicitement la prise en charge par le NHS (National Health Service) des traitements au regard du coût par QALY sans pour autant avancer explicitement de valeur plafond que l'on pourrait interpréter comme une évaluation de la valeur de la vie<sup>(26)</sup>. À l'opposé les dépenses de santé en France se situent parmi les plus élevées des pays de l'OCDE. Or, si la HAS (Haute autorité de santé) fait aujourd'hui officiellement référence aux méthodes d'évaluation économique<sup>(27)</sup>, l'évaluation coût-utilité ne tient pas une place aussi explicite, ou tout du moins aussi effective, dans le processus de décision du remboursement des traitements médicaux que dans le NHS. Ainsi, ce n'est pas tant que la France placerait la valeur de la vie au-dessus du Royaume-Uni que l'idée même de l'arbitrage coût-utilité, ou pour le moins son explicitation et sa formalisation, qui est aujourd'hui plus difficilement admise.

Si l'on regrettera que sur les questions de santé, un raisonnement en termes de choix social ne puisse éclairer plus explicitement la décision publique en France, on ne peut cacher qu'un arbitrage coût-utilité, qui se réfère explicitement à la valeur de la vie, est une menace pour la soutenabilité politique du système d'assurance maladie tel que nous le connaissons en France. Il peut en effet conduire à ne pas rembourser des soins qui prolongent la vie et qui ne seraient, de ce fait, financièrement accessibles qu'aux ménages

(25) Voir Cuyler et Rawlins, 2004.

(26) Le jugement du NICE est en outre modulé en fonction de la pathologie, de l'existence de traitements alternatifs et plus généralement de considérations éthiques (Cuyler et Rawlins, 2004 et Baumstark et al., 2008).

(27) La CEESP (Commission d'évaluation économique et de santé publique) concrétise la place prise aujourd'hui par l'évaluation économique au sein de la HAS.

les plus aisés. C'est aussi mettre les médecins dans des situations potentiellement extrêmement délicates, voire insoutenables, en leur demandant de refuser des soins à certains patients alors que les traitements sont disponibles. La menace que fait peser sur le système de santé français l'émergence de thérapies efficaces mais financièrement prohibitives est réelle.

Ce danger doit naturellement être apprécié à sa juste mesure quand il s'agit de décider du remboursement d'un traitement et cela dépasse bien entendu le strict périmètre d'application des calculs coûts-utilités dont nous venons de discuter la pertinence. Pourtant, si le principe de donner une valeur à la vie peut apparaître choquant, toute décision publique décidant ou non de la socialisation de dépenses de santé peut être comprise comme le résultat de la confrontation entre un coût et un bénéfice social. La question est simplement celle de l'explicitation de cet arbitrage pour en faire un élément affiché du choix public.

## Références bibliographiques

- Arthur W.B. (1981) : « The Economics of Risk to Life », *American Economic Review*, vol. 71, n° 1, mars.
- Baumstark L., M-O. Carrère et L. Rochaix (2008) : « Mesures de la valeur de la vie humaine. Usages et enjeux comparés dans les domaines de la santé et des transports », *Les Tribunes de la Santé*, n° 21.
- Blomquist G. (1979) : « Value of Life Saving: Implications of Consumption Activity », *Journal of Political Economy*, vol. 87, n° 3, pp. 540-558.
- Boadway R. et N. Bruce (1984) : *Welfare Economics*, Blackwell, Oxford.
- Cutler D.M. et M. McClellan (2001) : « Is Technological Change in Medicine Worth It? », *Health affairs*, vol. 20, n° 5, septembre/octobre.
- Culyer J. et D. Rawlins (2004) : « National Institute for Clinical Excellence and its Value Judgments », *British Medical Journal*, n° 329, pp. 224-227.
- Cutler D.M. et E. Richardson (1997) : « Measuring the Health of the US Population », *Brookings Papers on Economic Activity. Microeconomics*, vol. 1997, pp. 217-282.
- Cutler D.M. et E. Richardson (1998) : « The Value of Health: 1970-1990 », *AEA Papers and Proceedings*, n° 97-100.

- Cutler D.M., A.B. Rosen et S. Vijan (2006) : « The Value of Medical Spending in the United States, 1960-2000 », *New England Journal of Medicine*, n° 355, pp. 920-92.
- Erhlich I. et Y. Yin (2004) : « Explaining Diversities in Age-Specific Life Expectancies and Values of Life Saving: A Numerical Analysis », *NBER Working Paper*, n° 10759.
- Fleurbaey M. et E. Schokkaert (2009) : « Unfair Inequalities in Health and Health Care », *Journal of Health Economics*, vol. 28 n° 1, pp. 73-90.
- Fleurbaey M., S. Luchini, E. Schokkaert et C. Van de Voorde (2012) : « Évaluation des politiques de santé : pour une prise en compte équitable des intérêts des populations », *Économie et Statistique*, à paraître.
- Garbacz C. (1989) : « Smoke Detector Effectiveness and the Value of Saving a Life », *Economics Letters*, n° 31, pp. 281-286.
- Garen J. (1988) : « Compensating Wage Differentials and the Endogeneity of Job Riskiness », *The Review of Economics and Statistics*, vol. 60, n° 1, pp. 9-16, février.
- Hall R.E. et C.I. Jones (2007) : « The Value of Life and the Rise in Health Spending », *The Quarterly Journal of Economics*, vol. 122, n° 1, pp. 39-72, février.
- Jenkins, R.R., N. Owens et L.B. Wiggins (2001) : « Valuing Reduced Risks to Children: The Case of Bicycle Safety Helmets », *Contemporary Economic Policy*, vol. 19, n° 4, pp.397-408.
- Leigh J.P. et R.N. Folsom (1984) : « Estimates of the Value of Accident Avoidance at the Job Depend on the Concavity of the Equalizing Differences Curve », *Quarterly Review of Economics and Business*, vol. 24, n° 1, pp. 56-66.
- McHorney C.A., J.E. Ware et A.E. Raczek (1993) : « The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and Clinical Tests of Validity in Measuring Physical and Mental Health Constructs », *Med Care*, n° 31, pp. 247-263.
- Miller P., C. Mulvey et K. Norris (1997) : « Compensating Differentials for Risk of Death in Australia », *Economic Record*, vol. 73, n° 223, pp. 363-372.
- Moore M.J. et W.K. Viscusi (1990) : « Discounting Environmental Health Risks: New Evidence and Policy Implications », *Journal of Environmental Economics and Management* n° 18, pp. S51-S62.
- Murphy K.M. et R.H. Topel (2006) : « The Value of Health and Longevity », *Journal of Political Economy*, vol. 114, n° 5, pp. 871-904.
- Neumann P.J., D.E. Zinner et J.C. Wright (1997) : « The Value of Health and Longevity », *Medical Decision Making*, vol. 17, n° 4, pp. 402-408, octobre/décembre.



- Nord E., J.L. Pinto, J. Richardson, P. Menzel et P. Ubel (1999) : « Incorporating Societal Concerns for Fairness in Numerical Valuations of Health Programs », *Health Economics*, n° 8, pp. 25-39.
- Philipson T., G. Becker, D. Goldman et K. Murphy (2010) : « Terminal Care and the Value of Life Near Its End », *Milton Friedman Institute Working Paper*, n° 2010-05.
- Rosen (1988) : « The Value of Changes in Life Expectancy », *Journal of Risk and Uncertainty*, vol. 1, n° 3, pp. 285-304.
- Siebert W.S. et X. Wei (1994) : « Compensating Wage Differentials for Workplace Accidents: Evidence for Union and Nonunion Workers in the UK », *Journal of Risk and Uncertainty*, vol. 9, n° 1, pp. 61-76.
- Stevenson R. (1991) : « The Oregon Formula: Health Economists Dream or Stalinist Nightmare? », *Archives of Disease in Childhood*, n° 66, pp. 990-993.
- Thaler R. et S. Rosen. (1975) : « The Value of Saving a Life: Evidence from the Labor Market » in *Household Production and Consumption*, Terleckyj (ed.), Columbia University Press, pp. 265-300.
- Usher D. (1973) : « An Imputation to the Measure of Economic Growth for Changes in Life Expectancy » in *The Measurement of Economic and Social Performance*, Moss (ed.), NBER Books, pp. 193-232.
- Viscusi W.K. (1993) : « The Value of Risks to Life and Health », *Journal of Economic Literature*, vol. 31, n° 4, décembre.
- Viscusi W.K. et J.E. Aldy (2003) : « The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World », *NBER Working Paper*, n° 9487.
- Zweifel P. (2009) : « Economic Valuation of Life and Health » in *Health Economics*, Zweifel, Breyer et Kifmann (eds), Oxford University Press, 2<sup>e</sup> éd.



## Résumé

Dans ce rapport, Gilles Saint-Paul présente ses réflexions sur l'organisation du système de santé français. Il propose des pistes de réforme, en posant sur ce secteur aux nombreuses spécificités le regard d'un économiste extérieur au domaine de la santé. Avec une prise de recul revendiquée, l'auteur analyse en particulier deux points. D'une part, il se pose la question de savoir ce qui peut justifier que l'assurance et une partie de l'offre de soins soient publiques, et non pas privées. D'autre part, l'auteur se penche sur la tarification à l'activité (T2A). Celle-ci consiste à se doter de prix pour les actes médicaux effectués à l'hôpital. Elle a pour effet d'inciter les hôpitaux à aligner leurs coûts sur les tarifs imposés, mais la question de la fixation des tarifs se pose avec acuité. Enfin, l'auteur réaffirme le bien-fondé de mesures accentuant l'internalisation des coûts chez les différents acteurs (patients, médecins, mutuelles). En particulier, l'auteur préconise l'adoption d'un système de franchise/plafond indexé sur les revenus de l'assuré.

### **1. L'intervention publique dans le secteur de la santé**

Dans un premier temps, Gilles Saint-Paul pose la question de la justification du monopole public de l'Assurance-maladie. L'auteur rappelle que le marché de l'assurance des soins de santé n'est à l'évidence pas un marché comme les autres, notamment en raisons des asymétries d'information qui le caractérisent : aléa moral et sélection adverse.

#### **1.1. Les justifications de l'intervention publique dans le marché de l'assurance-santé**

L'aléa moral justifie l'existence du ticket modérateur, car sinon, rien ne viendrait limiter les dépenses des patients, couvertes à 100 %. La sélection adverse est, elle, susceptible de mener, sinon à la disparition pure et simple du marché de l'assurance-santé, du moins à l'écramage des risques forts, tenus de s'acquitter de primes d'assurance élevées. Cette propriété bien connue en économie de l'assurance, et appliquée au remboursement des soins de santé, empêcherait la redistribution horizontale entre biens portants et malades, qui est pourtant un des premiers fondements de l'assurance-maladie.

Le marché de la santé est par ailleurs incomplet, du fait qu'un certain nombre de maladies ou prédispositions à des maladies, étant héréditaires,

sont présentes dès la naissance, et ne peuvent donc pas donner lieu à une assurance *ex ante*.

En outre, le caractère obligatoire de l'affiliation à l'assurance-maladie est fondé, entre autres, sur une possible myopie des agents, qui pourraient ne pas juger utile de s'assurer contre un risque fort, mais dont la probabilité d'occurrence serait faible. L'État endosse alors un rôle paternaliste en décidant, à la place des agents, de tous les assurer.

Enfin, l'auteur se penche sur les propriétés de l'assurance-maladie comme outil de redistribution en nature. Les raisonnements usuels invitent à préférer la redistribution sous forme monétaire à la redistribution en nature : la redistribution monétaire laisse aux agents le libre choix de l'utilisation des sommes reçues, ce qui est impossible avec une fourniture d'un bien ou service en particulier. Toutefois, l'auteur rappelle que des arguments plus fins découlant du lien entre revenu et demande de soins sont susceptibles de nuancer ce propos.

## 1.2. Les objections à l'intervention publique

À chacun de ces arguments, l'auteur réagit en avançant que la fourniture publique de l'assurance-maladie est certes une façon de le prendre en considération, mais pas nécessairement la seule.

L'existence de franchises est tout à fait envisageable avec un système d'assurance-santé privée. De plus, l'interdiction de procéder à de l'écrémage, potentiellement causé par la sélection adverse, pourrait être imposée à des assureurs privés comme à l'assurance-maladie publique. De la même façon, les assureurs privés pourraient être contraints d'accepter tous les clients, y compris ceux atteints de pathologies héritées ou apparues dès le plus jeune âge. Concernant le paternalisme, l'auteur suggère qu'une information sur les risques suffisamment diffusée, ainsi qu'une inscription d'office dans un régime d'assurance, pourraient résoudre le problème de la myopie des agents.

Au final, l'auteur en déduit qu'aucun des arguments étudiés ne justifie à lui seul que la fourniture privée d'assurance-santé soit inefficace. L'assurance-maladie pâtit, en outre, du système d'assurances complémentaires, que l'auteur souhaiterait voir remplacé par des assurances supplémentaires. Pour chaque bien de santé, un unique assureur serait concerné, ce qui limiterait les problèmes liés à l'absence de coordination entre assureurs multiples.

## 2. Analyse critique de la T2A

Dans un second temps, l'auteur analyse en détail les implications de la mise en place de la T2A sur le marché de la fourniture de soins. La T2A, qui s'applique pour l'heure aux hôpitaux publics et à certaines cliniques privées sous contrat, consiste à définir une grille des diverses pathologies,

les groupes homogènes de malades (GHM), et un tarif de remboursement unique pour chaque GHM. L'auteur reconnaît l'amélioration de la T2A par rapport aux pratiques précédentes, qui donnaient lieu au remboursement des frais encourus par les hôpitaux sur la base de leurs seuls coûts réels. Toutefois, Gilles Saint-Paul considère que certains points méritent d'être discutés :

- les tarifs sont-ils correctement fixés ?
- les GHM sont-ils réellement homogènes ?
- le signal fourni par le tarif donne-t-il réellement lieu aux réallocations souhaitées ?

## 2.1. Définition du tarif et équilibre du marché des soins

Tout en admettant les particularités du secteur de la santé, Gilles Saint-Paul se propose, à l'instar de l'analyse microéconomique de l'équilibre de marché, d'étudier les implications de la T2A comme mécanisme affectant l'offre et la demande de santé.

La T2A représente un premier pas indéniable vers une reconnaissance de l'importance d'un signal-prix sur le marché de la fourniture de soins. Toutefois, le tarif est asymétrique, au sens où il ne véhicule pas, ou mal, la valeur accordée par les consommateurs aux soins reçus, et se limite à un signal-coût à l'endroit des producteurs.

*Du côté de la demande*, il est en effet difficile pour les autorités publiques de mesurer la valeur marginale accordée par les patients aux soins qui leurs sont prodigués. Elles peuvent avoir recours aux évaluations contingentes, de manière à quantifier la valeur du « bien santé ». Toutefois, ces méthodes aboutissent à une valorisation moyenne, et non pas marginale, des soins. Il est alors possible que les pouvoirs publics surestiment la valeur de la santé, et proposent en conséquence des tarifs trop élevés. Ces derniers, à leur tour, inciteront les hôpitaux à accroître leur offre plus que nécessaire. Ces évaluations ne sont pas non plus capables de mesurer les externalités positives associées aux soins (au premier rang desquelles se trouvent les vaccinations) qui participent à la valeur sociale de la demande de soins. Une solution pour inférer la valeur des soins ressentie par les patients serait de mettre en place des phénomènes de rationnement par files d'attente. Les tarifs d'un GHM seraient d'autant plus revalorisés que la pénurie associée serait relativement forte.

*Du côté de l'offre*, la raison d'être de la T2A paraît plus claire : ce système impose un remboursement uniforme sur la base du coût moyen observé par enquête sur un échantillon représentatif d'établissements de santé. La T2A incite donc les hôpitaux à accroître leur productivité, et à réallouer leurs ressources à cet effet. Selon la nature et la structure des coûts supportés par les firmes, les conséquences d'un tarif administré, égal au coût moyen, seront variées.

Si on considère que les coûts unitaires sont constants et indépendants du volume de production, le tarif de T2A forcera les établissements les moins

productifs à restructurer leurs activités, voire à fermer. Si le coût unitaire s'élève avec l'échelle de production (difficulté de dupliquer une même activité), le tarif T2A pourrait avoir des conséquences dramatiques, car alors les hôpitaux ne rentreraient plus dans leurs frais. Le cas inverse où le coût unitaire décroît avec la production (amortissement de coûts fixes) donne, lui, classiquement lieu au monopole naturel. Enfin, dans le cas où, pour chaque établissement, existe une taille optimale minimisant le coût de production, l'instauration de la T2A aurait pour effet de sélectionner les hôpitaux les plus rentables. La question du volume de production se poserait alors, car rien ne garantirait que les établissements dont l'activité serait rentable au tarif imposé offriraient la quantité optimale de soins.

La T2A est également un outil puissant de redéploiement de l'offre de soins. Par exemple, certains soins en ambulatoire ont vu leur tarif s'élever par rapport au tarif des soins en hospitalisation équivalents. Ces ajustements tarifaires traduisent la volonté des pouvoirs publics de développer la médecine ambulatoire, pour répondre à une modification de la demande de soins. L'auteur se demande cependant comment, de manière opérationnelle, les autorités de santé pourraient quantifier la hausse du tarif nécessaire pour infléchir les comportements dans le sens souhaité, et sur quelle durée cette déformation temporaire des tarifs devrait être opérée.

Pour garantir la maîtrise budgétaire globale des dépenses de santé, les tarifs de la T2A sont ajustés de manière uniforme à la baisse, dès que l'objectif de l'ONDAM est dépassé. Or, l'uniformité de la baisse n'exerce aucun effet incitatif sur les comportements individuels : il conviendrait de réduire les tarifs des professionnels qui sur-prescrivent, et non pas ceux de l'ensemble des acteurs. En l'état, l'articulation T2A/ONDAM donne lieu au problème du passager clandestin. L'auteur suggère de davantage fonder la régulation de l'activité sur les incitations individuelles, quitte à ne plus imposer un contrôle global des dépenses, car celui-ci introduit une forme de sur-détermination.

## **2.2. La question de l'homogénéité des GHM**

La définition des GHM (groupes homogènes de malades) pose l'épineuse question du degré de précision de la cartographie des différentes pathologies. En particulier, ces dernières se distinguent non seulement par leurs natures propres (affections respiratoires, digestives, etc.), mais également, à pathologie donnée, par leurs degrés de gravité. L'uniformité du remboursement, quelle que soit la gravité de la maladie, pourrait engendrer une forme de sélection adverse. Les comparaisons internationales suggèrent toutefois que la France a déjà choisi un niveau de détail particulièrement élevé (plus de 2 000 GHM, contre 1 200 en Allemagne).

Par ailleurs, la qualité des soins présente un certain degré d'hétérogénéité entre établissements de santé. Il est possible que l'instauration d'un tarif unique T2A ne mène pas à une homogénéisation de la qualité offerte, mais

plutôt à des phénomènes de files d'attente différenciées : les établissements de haute qualité seraient sujets à rationnement, et l'accès aux autres hôpitaux serait plus aisé. En tout état de cause, le lien entre le tarif T2A et le niveau de qualité jugé acceptable n'est pas évident et gagnerait à être appréhendé avec précaution.

### 2.3. La question des mécanismes de réallocation

La T2A est conçue comme un signal-prix adressé aux fournisseurs de soins. Ce signal devrait, dans l'idéal, les mener à adapter leur offre en fonction du niveau de leurs coûts de production relativement aux tarifs pratiqués. Or, il n'est pas certain que les personnels de santé répondent systématiquement à ces incitations. Pour encourager ces évolutions, des mesures récentes rendent possible l'instauration de primes d'intéressement, mais leurs effets ne sont pas encore perceptibles. Par ailleurs, la logique de l'alignement des coûts de production voudrait que certains établissements cessent leurs activités, mais dans la pratique, de fortes oppositions retarderaient, sinon empêcheraient, ces fermetures.

## 3. Quelques perspectives de politique économique

Le système français d'assurance-santé tente de faire face à deux contraintes antagonistes : assurer un accès égalitaire aux soins d'une part, et enrayer la montée des dépenses d'autre part. À cet égard, les revalorisations des tickets modérateurs suggèrent qu'une dose de responsabilisation individuelle, même progressive, devrait avoir des vertus en termes budgétaires, sans forcément engendrer des effets anti-redistributifs majeurs.

L'auteur est favorable à la mise en place d'un système de franchise/plafond, par exemple annuels, et dont les montants pourraient être indexés sur les revenus des assurés. De cette façon, les pouvoirs publics choisiraient la dose de redistribution verticale opérée par l'assurance-maladie.

Gilles Saint-Paul propose par ailleurs une réforme ambitieuse, consistant en l'instauration de comptes-santé individuels. Ces comptes seraient initialement crédités d'un montant à définir. Puis, les recours aux dépenses de santé viendraient réduire le solde détenu. Les assurés n'auraient jamais l'obligation de financer un solde négatif, mais pourraient en revanche demander la conversion de soldes excédentaires, dépassant un certain seuil, en versements monétaires. Avec un pareil compte santé, les efforts des assurés en matière de prévention, dépistage, soins précoces, leur permettraient de conserver, ou au moins de retrouver après un certain temps, un solde positif. Cela créerait des incitations financières à adopter de telles pratiques. *A contrario*, un système de plafond annuel de reste à charge produirait moins d'incitations à la modération des dépenses et aux comportements précautionneux, car l'assuré ne serait plus sollicité pour le financement de ses soins, une fois le plafond dépassé.

Gilles Saint-Paul considère que les incitations pourraient également être accentuées à l'endroit des professionnels de santé. La T2A est un indéniable premier pas, mais elle fournit le signal prix, sans pour autant offrir de courroie de transmission entre tarif et incitations à la réallocation. Accélérer la mise en place des procédures d'intéressement dans les hôpitaux publics, ou susciter le développement des établissements privés, lucratifs comme non lucratifs, permettrait de surmonter cet obstacle. Concernant la médecine de ville, l'extension de la T2A à ce secteur paraît souhaitable. Il convient néanmoins de reconnaître que les réorganisations causées par la T2A risquent d'être moins efficaces et plus douloureuses pour des cabinets de petite taille, qu'elles ne le sont pour de vastes hôpitaux. La tendance actuelle du regroupement de médecins libéraux dans des structures privées de taille conséquente laisse cependant penser que la T2A gagnerait à être étendue à la médecine de ville.

#### 4. Commentaires

Selon Franck von Lenep, l'argumentaire du rapport de Gilles Saint-Paul tient principalement en ce que le recours au signal-prix est trop peu fréquent dans l'organisation du système de santé français. Or, Franck von Lenep n'est pas convaincu qu'une utilisation accrue du signal-prix soit toujours souhaitable.

Ainsi, du côté du comportement des assurés, Franck von Lenep doute de la pertinence de l'aléa moral *ex ante*. Quant à l'aléa moral *ex post*, s'il existe – en remboursant intégralement les dépenses de santé, l'incitation à sur-consommer est indéniable –, il semble quantitativement faible. De plus, en élevant la part payée par l'assuré, toutes les dépenses sont réduites, et pas uniquement celles considérées comme inutiles, ce qui invite à la plus grande prudence. Une réduction excessive des dépenses de santé – dont celles consacrées à la prévention – peut d'ailleurs s'avérer contre-productive.

Concernant le comportement des offreurs de soins, Franck von Lenep reconnaît la finesse de l'analyse de Gilles Saint-Paul sur les implications de la T2A. Cette démarche peut cependant être dépassée. Des gains d'efficacité, dans la médecine de ville comme à l'hôpital, sont possibles (par exemple, en matière de prescription médicamenteuse). Le levier tarifaire existe bel et bien : à l'instar de l'Allemagne, les tarifs de T2A devraient être dégressifs à l'échelle de la fourniture de soins d'un établissement, et non pas à l'échelle nationale, comme c'est le cas avec l'ONDAM. Toutefois, au-delà de ces incitations par les prix, c'est toute l'organisation du système de soins qu'il convient de redessiner. La définition du panier de soins et des recommandations de bonne pratique, ainsi qu'une meilleure coordination entre médecine hospitalière et médecine de ville, sont des étapes essentielles, qui gagneraient à être également prises en considération.

Brigitte Dormont rejoint Gilles Saint-Paul sur l'utilité de l'analyse économique, pour une clarification des objectifs du système de santé français,



ainsi que pour une évaluation de ses performances. Selon Brigitte Dormont, l'analyse économique de Gilles Saint-Paul est peut-être, en certaines occasions, poussée à un niveau qui la rendrait difficilement applicable. Pour autant, elle reconnaît que des vérités essentielles apparaissent dans ce rapport : la nécessité de séparer les champs d'intervention de l'Assurance-maladie et des assurances complémentaires, l'impossibilité pour la tarification à l'activité de répondre simultanément à trois objectifs, ainsi que la difficile prise en considération de la qualité des soins dans la définition des tarifs T2A.

Brigitte Dormont regrette cependant que ce rapport n'ait pas dressé un état des lieux du fonctionnement du système de soins et n'ait pas montré en quoi la situation a évolué depuis le précédent rapport du CAE de Michel Mougeot.

Concernant le fonctionnement de l'hôpital, Brigitte Dormont rappelle que la grille de la T2A est si précise que le système se rapproche considérablement d'une tarification à l'acte susceptible d'engendrer de la demande induite. Il faudrait par ailleurs considérer les deux types d'aléa moral (sur l'offre et sur la demande) et comparer leurs effets sur la demande de soins.

Brigitte Dormont se tourne ensuite vers la question de l'assurance-maladie, sur laquelle porte l'essentiel de ses commentaires.

La question principale n'est pas le caractère public ou privé de l'assurance-maladie, mais son statut (obligatoire ou facultatif), et le nombre d'assureurs (un seul ou plusieurs en concurrence). L'argument paternaliste, avancé par Gilles Saint-Paul, est certes pertinent, mais au-delà d'une imparfaite prise en compte des risques par les individus, il est une seconde raison qui justifie le caractère obligatoire de l'assurance-maladie : *a posteriori*, les soins vitaux seront nécessairement pris en charge par la collectivité. Il n'est pas envisageable de laisser des individus atteints de graves affections sans soin, au motif qu'ils n'auraient pas choisi de s'assurer.

Comme le souligne Gilles Saint-Paul, l'articulation régime obligatoire/ assurances complémentaires est une particularité française ; ailleurs, les assurances facultatives sont supplémentaires et remboursent des dépenses autres que celles prises en charge par le régime obligatoire. La situation française est bien source d'inefficacité, mais il faudrait aller plus loin dans l'analyse. Premièrement, pour les individus n'ayant pas accès à un contrat collectif d'entreprise, les primes sont dégressives. Elles peuvent être très élevées pour les ménages les plus modestes, leur interdisant de fait l'accès. Deuxièmement, les complémentaires contournent l'interdiction de fonder les tarifs sur l'état de santé, en proposant une variété de contrats différents. Elles segmentent le marché en incitant chaque catégorie de risque à opter pour le contrat qui lui est destiné. Ceci engendre une dégradation de la mutualisation des risques, et des primes élevées pour certains individus. Troisièmement, le fait que la couverture par le régime obligatoire diminue pour certains postes (certains médicaments) est compensé par la plupart

des mutuelles, ce qui s'avère contre-productif. Quatrièmement, le reste à charge peut être très élevé pour une petite minorité d'individus, pas forcément couverts par une complémentaire ou par le régime des affections de longue durée.

L'évolution du système d'assurance-maladie est une exigence, selon Brigitte Dormont. Elle peut prendre la forme d'une régulation des assurances complémentaires, avec interdiction de la sélection et compensation des risques. On peut concevoir aussi bien un unique assureur, qu'une concurrence régulée. Elle peut également prendre la forme d'un bouclier sanitaire, caractérisé par un plafond de reste à charge en cas de dépenses répétées, mais financé par une franchise pesant sur tous.

L'obligation d'assurance, réaffirmée par Brigitte Dormont, doit porter sur un panier de soins de base. La définition de ce panier, et son évolution avec le progrès technique médical, se poseront inévitablement, nécessitant un arbitrage avec d'autres dépenses. Finalement, le système français accorde aujourd'hui, pour la sélection des traitements, davantage de place aux critères médico-économiques, au moyen notamment des QALYs qui permettent de mesurer les différents états de santé des individus.

## Summary

# Considerations Concerning the Organisation of Health Systems

In this report, Gilles Saint-Paul gives his thoughts on the organisation of the health system in France. As an economist working outside the health sector, he proposes a number of possible reforms for a sector of the economy with very specific characteristics. With a deliberately external view, he focuses his attention on two areas. First, the author asks what is it that can justify that health insurance and a part of health provision should be public and not private. Secondly, he explores the system of hospital financing by activity (T2A – for ‘tarification à l’activité’). This is a system whereby each act performed in a hospital has an associated tariff. This has the effect of an incentive for the hospitals to adjust their costs to the level of the tariffs imposed from the outside, but naturally raises the burning question of how these tariffs get fixed. Finally, the author reaffirms the validity of measures that help all the various stakeholders in the system (patients, doctors, health insurance firms) to take on board the costs. In particular the author recommends adopting a system of fixed-sum and ceiling index-linked on individual income.

## 1. State involvement in the health sector

To begin with, Gilles Saint-Paul asks the question concerning the justification of a public monopoly in health insurance. The author explains that the market for health-care insurance is clearly not just like any other market, in particular because it is subject to two key characteristics not found elsewhere – moral hazard and adverse selection.

### 1.1. Justification for state intervention in the health insurance sector

Moral hazard justifies the use of the ‘ticket modérateur’ (the part of the cost not covered by the state insurance scheme), otherwise there would be no effective limit to health expenditure, it being insured at 100%. Adverse

selection would lead, if not exactly to the extinction of the health insurance market, to creaming off the high risk cases and demanding correspondingly high premiums. This well-known characteristic of the insurance market applied to reimbursing the costs of health care would prevent the horizontal redistribution between the healthy and the sick, which is one of the fundamental principles of health insurance.

The health care market is also incomplete in that a certain number of afflictions or predispositions to illness are hereditary and, being present from birth, cannot give rise to insurance *ex ante*.

In addition, the mandatory nature of health insurance is based in part on the notion that people are very short-sighted and sometimes do not see the point of insuring against serious illness if the chance of contracting such illness is very low. The state adopts a paternalistic attitude and decides to insure everyone.

Finally, the author looks at health insurance as a means of redistribution in terms of services or goods. The usual arguments favour monetary redistribution, which allows the individual the choice of how to use the sums provided, something which is impossible with specific services or goods. Nevertheless, the author recognises that a more detailed analysis of the link between income and health-care demand could modify this sort of argument.

## **1.2. Objections to these justifications**

The author responds to each of these arguments by saying that, whilst a state-operated health insurance system is certainly one way of handling the situation, it is not necessarily the only way.

The existence of a fixed sum payable by the individual is perfectly feasible with a private health insurance scheme. In addition, it is perfectly possible simply to forbid creaming off caused by adverse selection for both private and state-run insurers. In the same way, private insurers could be obliged to accept all customers, including those with hereditary pathologies or afflictions discovered at a very early age. As far as paternalism is concerned, the author suggests that mandatory subscription to a health insurance scheme, coupled with suitable publicity concerning risks, would overcome the problem of peoples' short-sightedness.

In conclusion the author states that none of the arguments put forward can justify the claim that private health insurance would be ineffective. Health insurance suffers also from the system of complementary insurance, which the author would like to see replaced by a system of supplementary insurance. For each health service product, a single insurer would be involved, thereby limiting the problems associated with an absence of coordination between multiple insurers.

## 2. Critical analysis of the T2A

In the second part, the author provides a detailed analysis of the implications of the T2A system for financing health care provision. The T2A system applies to all public hospitals and to certain contracted private clinics. It consists of a grid of all the various pathologies, grouped by similarity of tests and acts carried out into GHM (groupe homogène de malades or diagnosis related group), with the tariff for reimbursement for each GHM. The author recognises the improvement that the T2A has achieved compared with the previous system whereby hospitals were reimbursed simply on the basis of the actual costs incurred. Nevertheless, Gilles Saint-Paul believes that certain points should be debated:

- are the tariffs correctly established?
- are the categories really homogeneous?
- does the signal given by the tariffs give rise to the desired reallocation of resources?

### 2.1. Tariff definition and the balance within the healthcare market

While recognising the peculiarities of the healthcare sector, Gilles Saint-Paul nevertheless recommends a microeconomic analysis of the market sector, and to study the implications of the T2A system as it affects the balance of supply and demand in the provision of healthcare.

The T2A represents a definite first step in having the importance of a price signal recognised in the market for providing healthcare. However, it is an asymmetric indicator in the sense that it has difficulty in reflecting the value that the person receiving the care perceives, and can really only act as a cost signal for the providers of the care.

On the demand side, it is difficult for the health authorities to measure the marginal value that patients perceive in the care received. They need to have recourse to ancillary studies and estimations so as to arrive at a value for a “health product”. These methods, however, tend to produce an average value, and not the marginal value, of the care provided. It is thus possible that the health authorities overestimate the value, and recommend tariffs that are too high. And in their turn, the hospitals would then be encouraged to increase the offer more than is required. In addition, such evaluations are also incapable of measuring the external benefits associated with care provision (the most significant being vaccination) that add a social value to the demand for healthcare. One solution for helping to understand the value that patients feel would be to create a form of rationing using waiting lists. The more the scarcity of a particular GHM is felt, the greater would be the increase in its tariff.

On the supply side, the purpose of the T2A would seem much clearer: the system imposes a standard reimbursement based on the average cost

observed from a survey using a representative sample of healthcare establishments. The T2A encourages hospitals to improve their productivity, and to reallocate resources accordingly. But depending on the nature of the cost structure in each establishment, the consequences of an administrative tariff, equal to the average cost, can vary considerably.

If one considers that unit costs are constant and independent of the throughput, the T2A tariff will force the least productive establishments to restructure their activities, or even to close down. And if the unit cost increases with the throughput (difficulty in duplicating the same activity), the T2A tariff could have dramatic consequences because, in that situation, the hospitals would never be able to recover their costs. If increased throughput caused a reduction in unit costs (amortising fixed costs), then a classic case of a natural monopoly would result. Finally, in the situation where, for each establishment, there is an optimal size for minimising the cost of production, the T2A tariffs would have the effect of selecting the more profitable hospitals. That would raise the question of the amount of the throughput, since nothing would guarantee that those establishments with the most profitable activity using the imposed rates would be offering the optimal level of activity.

The T2A is also a powerful tool in displacing the offer of healthcare. For example, certain activities that can be handled in out-patient mode have seen their tariffs increased compared with the same care offered to in-patients. Such adjustments reflect the desire of public health authorities to favour out-patient activities to respond to a change in the demand for certain care. The author, however, wonders how in practice the public health authorities can calculate the level of increase necessary to have the desired effect on patients' behaviour, and for how long this difference in tariff needs to be applied.

To ensure overall budgetary control of health expenditure, the T2A tariffs are adjusted uniformly downwards, as soon as the ONDAM objective is achieved. ONDAM (for Objectif national des dépenses d'assurance maladie) is the provisional budget allocated to public health insurance in France. But uniformity in the lowering of tariff has absolutely no incentive effect on individual behaviour – it would be more effective to reduce the tariffs for those professionals who over subscribe, and not for the profession as a whole. In its current form, the T2A and ONDAM combination creates the problem of the stowaway on board ship. The author therefore suggests it would be better to focus legislation on individual incentives, even if that means forgoing overall control of expenses, since the latter introduces a form of over-determination.

## **2.2. The question of the homogeneity of the GHMs**

Defining the GHMs raises the thorny question of how precise is the mapping of the various pathologies. Not only do they have their own defining characteristics (respiratory, digestive, etc.) but, with each pathology, there is

also the question of how severely afflicted is the patient. Depending on the severity, a flat-rate reimbursement could lead to adverse selection. However, international comparisons suggest that France has developed a particularly detailed version (more than 2,000 GHMs, compared with Germany's 1,200).

In addition, though, there is the inherent heterogeneous nature of the care and treatment offered from one healthcare establishment to another. It is quite possible that a single rate T2A does not lead to homogeneity in the quality of care provided, but rather to different lengths of time on waiting lists – high quality establishments effectively rationing their care and appointments more easily obtained elsewhere. It is fair to say that the link between the T2A tariffs and an acceptable quality of care is not obvious, and the idea needs to be treated with caution.

### **2.3. The question of reallocating resources**

The T2A is designed to give a price signal to care providers. Ideally, this signal should encourage them to adapt their offer as a function of their costs relative to the tariffs in place. But it is far from certain that establishments react systematically in this way. The effect of some recent incentives involving the possibility of bonuses for following this policy have not yet had the time to be measurable. In passing, one should note that the logic of aligning production costs would indicate that certain establishments should cease certain activities, but the strength of the opposition to this idea is sufficiently strong to delay or even prevent such closures.

## **3. Economic policy and health care provision**

The French system of health insurance tries to manage two objectives which push in opposite directions – ensure equal access to health care for all citizens, and put a brake on ever increasing costs. It would seem that increasing the level of the ticket modérateur, thereby installing a dose of individual responsibility, even progressive, should have a positive budgetary effect, without necessarily involving a significant anti-distributive influence.

The author is in favour of a fixed-sum/ceiling mechanism, annualised perhaps, where the sums involved would be linked to the individuals' incomes. By adopting this idea, the government would be choosing to have health insurance working as a vertical redistribution system.

Gilles Saint-Paul proposes an ambitious reform, consisting of establishing individual health accounts. Such accounts would be credited initially with a certain defined sum. Each time any health care 'product' was used, the account would be debited. No-one would be obliged to pay any negative balance but, on the other hand, someone with a positive balance, above a certain threshold, could ask for that to be reimbursed in cash. With such a health account system, someone who makes the effort in terms of

prevention, detection and early treatment could see him/herself with a positive balance over time. This could be a significant financial incentive to adopt such beneficial practices. On the other hand, a system with an annual ceiling with the excess being paid for by the state would be much less of an incentive for those who go beyond such a ceiling, and such people would be much less likely to moderate their consumption or adopt good practice once this ceiling had been reached.

Gilles Saint-Paul believes that such incentives could also be directed towards healthcare professionals. The T2A is clearly a first step and it provides the price signal, but it fails to provide the connection between the tariffs and the incentives for redirecting care. Overcoming this obstacle requires implementing processes for incenting public hospitals, or developing private institutions, be they non-profit-making or otherwise. It would seem desirable also to extend the idea of the T2A to those medical practitioners who operate outside the hospital system in private practice. It is fair to point out, however, that any reorganisation carried out as a result of applying the T2A tariffs would be less effective and also more painful for small scale surgeries than for large hospitals. But the current trend for doctors to form private urban groupings of some size would suggest that extending the T2A concept to these structures would be a help.

#### **4. Comments**

According to Franck von Lennepe, Gilles Saint-Paul's arguments are mainly based on his view that the price signal offered by the T2A system is not sufficiently used by the principal stakeholders in the French healthcare system. But von Lennepe is not convinced that an increased recourse to the mechanism would always be desirable.

As far as individual behaviour is concerned, von Lennepe questions the pertinence of the moral hazard *ex ante*. And as for the moral hazard *ex post*, if this exists – and it is clear that reimbursing the totality of health expenditure is an obvious incitement to over consumption – the quantity concerned would seem to be marginal. In addition, by increasing the proportion paid by the patient, all expenditure is reduced, not just that which is considered as ineffective, which suggests treating this idea with some caution. An excessive reduction in expenditure on health – especially that devoted to prevention – could prove to be counter-productive.

Concerning the healthcare providers, von Lennepe salutes the subtlety of Gilles Saint-Paul's analysis of the implications of the T2A system. This idea could also be taken further. Efficiency gains are possible both in the hospital environment and in the local doctors' surgeries (in the area of drug prescriptions for example). The tariff lever is real – as exists in Germany, the T2A tariffs should be on a sliding scale at the level of care provision by establishment, and not at the national level as is the case with ONDAM.



Nevertheless, over and above these price-based incentives, it's the whole healthcare system that needs redesigning. Defining a basket of care items, recommendations on best practice and a far better coordination between the local doctors and the hospital teams are all essential steps, and deserve to be given serious consideration.

Brigitte Dormont agrees with Gilles Saint-Paul both on the pertinence of an economic analysis and on the need for clarity in defining the health service's objectives, as well as measuring its performance. However, she believes that some of the conclusions from this economics-based treatment are pushed to a level that renders them difficult to implement. However, Dormont does recognise that the report highlights certain essential truths – the need to have a clear separation between the role of the standard health insurance and that of the complementary insurance, the inability of the T2A system to respond to three simultaneous objectives, and the thorny problem of incorporating the concept of care quality into the T2A tariff definitions.

But she is disappointed that the report did not include a statement on the current situation with regard to running the health system and show how it had improved, or otherwise, compared with the previous report from the CAE under the guidance of Michel Mougeot.

As for the hospital function, Brigitte Dormont claims that the T2A tariff grids are so precise that they resemble a system of payment by act, something which would itself be likely to create the demand. In addition it is necessary to take account of the two types of moral hazard (that concerning supply as well as that concerning demand) and compare their respective effects on demand for care.

Brigitte Dormont then turns her attention to the question of health insurance, the subject of the majority of her comments.

The key question is not whether the system is public or private, but rather its status (mandatory or optional) and the number of organisations offering the insurance (a single insurer or several in competition with each other). The paternalistic argument put forward by Gilles Saint-Paul is certainly relevant but, over and above the individual's incomplete appreciation of risk, there is a second reason that justifies health insurance being mandatory: a posteriori, critical and emergency care is necessarily the responsibility of the community. It is inconceivable that those who suffer from severe and critical conditions should be left without any care, just because they chose not to be insured.

As Gilles Saint-Paul points out, the tandem of obligatory basic insurance/complementary insurance is a peculiarity of the French system; elsewhere, optional insurance is additional and reimburses expenses incurred other than those that are covered by the mandatory system. The French system may well be a source of inefficiency, but the analysis needs to go deeper. First, for those who do not have access to a company scheme, the subscriptions are on a sliding scale. They can be very high for the more modest households,

effectively creating a barrier to entry. Secondly, the complementary insurers get round the ban on basing subscription rates on the person's state of health by offering a variety of different contracts. They segment the market in such a way that the individual is oriented towards the contract the company prefers he takes. This has the effect of reducing the level of mutualisation along with high subscriptions for certain people. Thirdly, the fact that the majority of complementary insurers compensate for any reduction in the rate of reimbursement by the state system (for certain drugs, for example) can be counter-productive. Fourthly, what remains for the individual to pay can be very high for a small minority of people, those who do not have complementary insurance and those not covered by the system for long-term pathologies.

Brigitte Dormont is convinced that the current health insurance system needs to develop and adapt. This could involve regulating the complementary insurers, prohibiting selection with compensation for the risks involved. One can also imagine a single insurer or a regulated competitive environment for several. Another solution would be to have a capped amount, with everyone paying a fixed sum which would serve to finance those who needed to go beyond the cap as a result of requiring long-term repetitive treatment.

The obligatory part of the insurance should be based on a minimum basket of care, according to Brigitte Dormont. Obviously, defining such a basket, plus understanding how it should evolve in line with technological advances, would necessitate some tough budgetary choices about where to make savings. Finally, the current French system is increasingly using certain medico-economic indicators, in particular the QALY measurement, as a means of evaluating an individual's state of health.



