

CHAPITRE 8

UNE CONCURRENCE ENTRE ASSUREURS POUR L'EFFICACITE PRODUCTIVE

Dans un très grand nombre de pays, l'assurance privée volontaire joue donc un rôle relativement marginal. Cette affirmation semble paradoxale quand on sait que de grandes réformes des systèmes de santé des années quatre-vingt dix ont porté sur l'assurance maladie et ont visé à introduire des mécanismes de marché dans ce secteur.

Comme nous l'avons souligné au terme de la seconde partie, ces réformes avaient pour principal objectif d'obtenir que les organismes concernés concentrent leurs efforts sur la gestion des risques et la recherche de l'efficacité productive en limitant la sélection différentielle. Autour de cette problématique commune, nous avons établi une distinction entre les systèmes de tradition libérale ou sociale.

Quels sont les principes généraux qui caractérisent les évolutions récentes dans ces deux types de système ?

1. Lorsque le marché de l'assurance est traditionnellement concurrentiel, les entités qui couvrent les risques prélèvent directement des primes d'assurance, plus ou moins ajustées au risque, auprès des assurés. L'organisme régulateur estime le niveau de risque des personnes qui sont effectivement couvertes par chaque assureur. Il met alors en place un mécanisme de refinancement croisé entre les assureurs qui couvrent des personnes présentant des risques relativement élevés et les autres. Que la structure de risque de l'assureur résulte du hasard ou de stratégies plus ou moins efficaces de sélection, il peut bénéficier de transferts compensatoires ou être au contraire obligé de financer les malchanceux. Les primes qu'il propose, dès lors, prennent en compte cet ajustement et ne reflète plus sa structure de risque mais son efficacité productive relative.
2. Lorsque le régulateur introduit une forme de concurrence ou de responsabilité financière des organismes gérant le risque maladie, on part d'une situation initiale dans laquelle des entités publiques sont financées au prorata de leurs dépenses. Les

ressources destinées à financer le système sont prélevées par l'Etat auprès des ménages bénéficiaires et les contributions sont généralement liées au revenu. Afin d'inciter les organismes à être plus efficaces, l'Etat met progressivement en œuvre un financement prospectif normatif, tout en maintenant la solidarité du financement. Il collecte les fonds auprès des ménages et alloue aux entités assurant la gestion de la couverture des capitations ajustées au risque.

Si les problématiques dans ces deux types de systèmes convergent, on constate aussi que les voies empruntées pour obtenir le résultat souhaité, mis à part le mode de prélèvement auprès des ménages, sont comparables : dans les deux cas, l'objectif est d'obtenir que les entités qui assurent la gestion des risques, reçoivent un paiement normatif ajusté au risque. Dans le premier cas, le régulateur met en place un ajustement des risques ex post ; dans le second, il verse aux assureurs une capitation et procède à un ajustement de cette capitation en fonction des risques ex ante¹. Parallèlement, on maintient (ou on introduit) la possibilité pour les assurés de choisir un opérateur parmi ceux qui lui sont proposés. Ceci renforce théoriquement les incitations des opérateurs à rendre un service coût-efficace et à améliorer la qualité des services rendus². L'objet de ce chapitre est d'approfondir l'étude technique de ce mode de gestion de la concurrence.

Au préalable notons que les entités qui prennent part à cette concurrence fortement encadrée peuvent être des assureurs privés, des caisses publiques, voire des groupes de prestataires de soins. Lorsqu'il n'est pas utile de les distinguer, nous les qualifions de façon générique d'Organismes Gestionnaires du Risque (OGR), terminologie qui témoigne du fait que l'efficacité productive est au cœur du questionnement.

Nous avons précédemment signalé que la sélection différentielle en assurance n'a pas été étudiée de façon formalisée dans la littérature économique. De même, les analyses théoriques des outils nécessaires pour limiter la sélection en stimulant la recherche de l'efficacité productive dans le cadre de l'assurance maladie sont rares. Pour autant, une littérature assez riche s'est développée depuis plusieurs années sur ce sujet. Elle s'inspire généralement de l'observation d'un système précis.

¹ Afin de lever une ambiguïté possible, rappelons que la notion de "paiement prospectif" fait référence à une situation dans laquelle un paiement normatif est déterminé en début de période. On oppose à cette notion celle de paiement rétrospectif qui qualifie une situation dans laquelle on rembourse les coûts de production lorsque le service a été produit. Dans les deux cas présentés ici, le paiement est prospectif normatif, mais les ajustements peuvent être réalisés ex ante ou ex post.

² Le concept d'efficacité productive rend compte simultanément des dimensions coûts et qualité.

Dans la première partie de ce travail, nous dégagons les éléments de convergence de ces travaux sur les conditions nécessaires à l'établissement d'une concurrence qui porte véritablement sur l'efficacité productive. Cette approche vise dans le même temps à rendre compte des contraintes techniques que le régulateur (qui n'est pas nécessairement l'Etat) doit surmonter, des arbitrages qu'il doit réaliser et des questions qui restent ouvertes (Section 1). Puis, dans un second temps, nous étudions plus précisément les expériences des différents pays en déterminant si les conditions nécessaires que nous avons identifiées sont remplies. Ce travail de recensement, qui se veut exhaustif, permet dans le même temps de tirer un premier bilan des réformes en cours et de montrer que l'organisation du financement et de l'activité des OGR selon les principes édictés n'est pas une condition suffisante pour atteindre l'objectif souhaité (Section 2).

1. REGLEMENTATION DE LA CONCURRENCE : ENJEUX ET PRINCIPES

La responsabilisation financière incite théoriquement les OGR en concurrence à améliorer leur efficacité productive. Nous avons toutefois vu que la sélection est la conséquence directe du versement d'un paiement prospectif mal ajusté au risque. Il n'est dès lors pas étonnant que l'organisation du mode de financement des OGR soit au cœur des solutions proposées. L'essentiel de la régulation passe par l'élaboration d'un mécanisme tel que les entités en concurrence ne soient pas tentées de sélectionner les assurés (1.1). Toutefois, pour parvenir à une solution satisfaisante, la concurrence doit se dérouler dans un cadre particulier et nous dégagons les conditions supplémentaires auxquelles les OGR et les assurés doivent se plier (1.2).

1.1 L'ajustement au risque

La solution naturelle pour réduire la sélection, du point de vue économique, consiste à mettre en place un mécanisme de financement tel que l'OGR ne puisse réaliser, à efficacité productive constante, un profit plus élevé s'il couvre une catégorie de risque plutôt qu'une autre. Cette condition est remplie si le financement obtenu pour chaque personne est actuariellement juste. Dans ce cas, des OGR en concurrence, qui sont incités à augmenter leur efficacité productive, pourraient même avoir intérêt à se

spécialiser dans l'offre de service pour les hauts risques, notamment pour bénéficier d'économies d'échelle³(Bowen, 1995, Kronick *et al.*, 1995).

Nous nous sommes attachés à détailler, dans la première partie de ce travail, les raisons pour lesquelles les niveaux de risque que présentent les individus diffèrent. Nous avons souligné que, pour pouvoir produire le service de couverture des risques, un assureur doit proposer à chaque assuré une prime actuarielle. Plus précisément, il différencie théoriquement les primes selon les facteurs de risque qui contribuent à expliquer la variance explicable. Par la suite, nous avons montré, d'une part, que l'assureur n'est pas toujours en mesure d'avoir accès à l'information dont l'individu dispose sur son propre niveau de risque, et d'autre part, qu'il pouvait choisir de pratiquer une tarification grossière des risques. Nous nous interrogeons désormais explicitement sur la faisabilité technique, pour le régulateur, d'un ajustement au risque, qu'il soit réalisé ex ante ou ex post. Au préalable, il convient de choisir un critère de jugement qui permette de déterminer la qualité d'un tel ajustement par rapport à une situation théorique idéale (1.1.1). Puis, nous nous demandons si, et par quels moyens, il est possible de mettre en œuvre un tel ajustement (1.1.2). Nous montrons enfin que l'on peut mettre en place des mécanismes qui complètent le dispositif et pallient ses défauts (1.1.3).

1.1.1 Critère d'évaluation de la qualité d'un ajustement au risque

Dans un premier temps, nous présentons le critère à l'aune duquel la qualité d'un ajustement au risque quelconque se mesure. Ceci nécessite d'une part, de caractériser d'une situation d'ajustement parfait et d'autre part, de trouver un critère de comparaison d'un mode d'ajustement particulier avec cette situation idéale.

Dans un premier temps, il faut déterminer quelle proportion de la variance est explicable, autrement dit quel est le R^2 maximal qui peut être prédit⁴. De nombreuses études proposent des estimations de la part de la variance explicable, mais les résultats diffèrent selon les échantillons, même lorsque les auteurs utilisent des spécifications économétriques similaires. Pour van Vliet et van de Ven (1992), elle est environ de 14%. Newhouse estime pour sa part qu'elle représente entre 20 et 25% de la variance totale (Newhouse, 1996, Newhouse *et al.*, 1989). Wolfe et Goddeeris (1991), dont l'objectif

³ Rentabilisation d'investissements spécialisés, d'un investissement dans la formation de personnels soignants...

⁴ cf Chapitre 3, § 3.1.1.

n'est pas de déterminer le R^2 maximum, obtiennent un R^2 de 22% (tableau 4 pp.448). On peut noter de plus que le R^2 maximal dépend du type de dépense que l'on observe : il est beaucoup plus élevé pour les dépenses ambulatoires (environ 50%, Newhouse *et al.* 1989) que pour les dépenses hospitalières (8%, *ibid.*).

Il n'existe donc pas d'étalon unique et les études qui visent à estimer la qualité d'un ajustement au risque choisissent en général, de façon assez arbitraire, un seuil d'environ 20%.

Dans un second temps, on estime la variance expliquée par les facteurs de risques utilisés par l'assureur ou le régulateur (le R^2 empirique) et on la compare au R^2 maximal choisi. Plus la part de la variance expliquée dans la variance explicable augmente, meilleur est l'ajustement au risque.

La pertinence d'un tel critère de jugement reste partielle, même s'il est couramment utilisé. Newhouse (1996) souligne ainsi qu'en ajoutant un facteur de risque, on peut augmenter la part totale de la variance expliquée mais diminuer la part de la variance expliquée dans certains sous-groupes. A titre d'illustration, on peut citer l'étude de Howland *et al.* (1987) : les mêmes facteurs de risques (mesures physiologiques et utilisation passée de ressources) expliquent 10% de la variance des hospitalisations observées dans les deux années suivantes pour les hommes et 3% pour les femmes.

L'existence d'un tel phénomène soulève plusieurs réflexions. D'une part, l'introduction d'un facteur de risque qui explique mieux la variance des dépenses d'un sous-groupe que d'un autre peut générer une incitation à une sélection différentielle, en faveur du groupe dont la variance est mieux rendue au détriment de l'autre. En prolongeant le raisonnement, on peut se demander s'il ne serait pas intéressant d'utiliser des facteurs de risque différents pour des populations disjointes. Luft (1995) estime que le programme de recherche ne devrait pas porter sur l'étude de la variance des dépenses individuelles dans un échantillon représentatif de la population mais sur les dépenses de groupes vulnérables caractérisés par des facteurs de risques particuliers⁵. De telles approches compliqueraient les mécanismes d'estimation et d'allocation, mais il serait intéressant de mener une analyse comparative en utilisant les différentes méthodes sur un même

⁵ "Concern about the poor explanatory power of models focusing on individual expenditures is misplaced if the real question is how well one can predict for groups, particularly groups reflecting non random risk selection" (Luft, 1995, pp.28).

échantillon, afin de déterminer celle qui rend le mieux compte de la variation des dépenses entre individus.

A défaut, pour juger de la véritable qualité d'un ajustement au risque, il ne faudrait pas se fonder sur la seule observation du R^2 . Il est préférable de tester la robustesse de l'ajustement en comparant les dépenses prédites aux valeurs observées sur des sous-populations spécifiques ou, comme le suggèrent van de Ven et van Vliet (1993), sur plusieurs intervalles de dépenses.

1.1.2 La tarification du risque en pratique et ses limites

Muni d'un critère de jugement, le régulateur doit choisir les critères qu'il utilise effectivement pour procéder à l'ajustement des risques. Cet exercice nécessite de prendre en compte un certain nombre de contraintes techniques (a). Une recherche très prolifique s'attache à déterminer les facteurs de risque qui devraient être pris en compte (b). Compte tenu de ces éléments, nous proposons un bilan provisoire de l'état de la recherche et des perspectives (c).

a) Les contraintes techniques de l'ajustement au risque

Que la tarification idéale du risque soit mise en œuvre par l'assureur directement, par un organisme chargé de la compensation des risques ex post, ou par une institution chargée de déterminer et de verser aux OGR une capitation normative, le choix des critères d'ajustement doit respecter un certain nombre de contraintes techniques (van de Ven et van Vliet, 1992). Ces contraintes sont plus ou moins fortes selon le contexte.

1. La valeur du critère sur laquelle est fondé le calcul doit pouvoir être estimée sans erreur de mesure.
2. Chaque critère doit rendre compte d'informations suffisamment structurelles pour que le paiement reflète une tendance valable sur l'ensemble de la période et ne caractérise pas seulement l'individu au moment où la mesure est effectuée.
3. Le coût d'extraction de l'information ne doit pas être trop élevé et, en règle générale, l'information doit être relativement facile à obtenir.
4. Les critères choisis ou les modes de collecte de l'information doivent respecter la vie privée des individus. Cette contrainte est plus ou moins forte selon le contexte et la qualité des personnes qui collectent l'information. Si la personne remplit un

questionnaire que l'employeur transmettra à l'assureur, il sera plus réticent à révéler certaines informations que s'il s'adresse à un médecin tenu de respecter le secret médical qui transmettra une information statistique agrégée décrivant les risques couverts par chaque OGR.

5. L'information ne doit pas être manipulable. Dans le cas où l'assuré paie une prime qui est fonction de son risque, il n'a pas intérêt à révéler à l'assureur des informations qui prouvent qu'il est un haut risque. Mais la manipulation peut aussi être le fait de l'OGR : dans la mesure où sa dotation est fonction des risques qu'il couvre, il a intérêt à surévaluer ces risques afin de recevoir un montant plus élevé. Les circuits d'information doivent être élaborés pour minimiser les possibilités de manipulation et l'institution qui détermine les paiements ou les ajustements, doit se ménager la possibilité de vérifier l'information.
6. Les critères ne doivent pas créer d'incitations perverses, notamment d'incitations à une moindre efficacité pour les OGR.
7. Enfin, il faut que chaque "cellule" définie par le croisement des critères comprenne un nombre suffisant d'individus pour pouvoir évaluer le paiement, sans commettre d'erreur de mesure. Nous avons souligné que c'était là une des raisons pour lesquelles l'assureur, qui n'a accès qu'à l'information contenue dans ses propres fichiers, pouvait chercher à éviter de couvrir certaines personnes. Un organisme qui centralise les informations émanant d'un grand nombre d'OGR peut avoir un avantage à cet égard. Mais un tel système centralisé comporte aussi des limites : il faut que chaque OGR soit en mesure de restituer périodiquement une grande masse d'information détaillée et standardisée sur la base de laquelle l'organisme central estimera la capitation ou l'ajustement. En outre, dans l'idéal, l'information restituée doit aussi permettre de distinguer les OGR qui produisent de façon peu efficace de ceux qui sont victimes d'une sélection défavorable (Swartz, 1995). Or, les systèmes d'information existant ne permettent pas toujours de respecter ces conditions. Aux Etats-Unis, ainsi il n'est pas rare que les plans passent eux même des contrats avec des groupes médicaux sur une base de capitation (ce qui est, après tout, un moyen de les inciter à réaliser un effort de gestion optimal) ou qu'ils salarient les prestataires de soins. Dans ce cas, les plans

n'ont pas nécessairement accès à une information détaillée de nature médicale (sur l'état de santé, les pathologies...) et ne sont pas en mesure de la restituer⁶.

b) Quelles variables utiliser dans l'ajustement au risque ?

Compte tenu de ces contraintes, nous présentons les principaux résultats de la recherche empirique, qui vise à déterminer les facteurs de risque les plus pertinents pour l'ajustement au risque.

Variables socio-démographiques

Le recours à la consommation médicale varie tout d'abord en fonction de l'âge et du sexe. Ces facteurs sont quasiment toujours pris en compte pour estimer les risques : les dépenses de santé sont plus élevées au début et à la fin de la vie et les dépenses des femmes sont plus élevées que celles des hommes pendant leur période de fécondité. Ces deux données sont donc des facteurs de base dans la détermination des dépenses de santé.

On sait par ailleurs que le niveau et le mode de recours aux soins varient en fonction de l'environnement économique et social des individus. On peut donc théoriquement envisager de prendre en compte un certain nombre de variables comme le niveau de revenu, le niveau d'étude, la profession (qui peut aussi refléter des facteurs de risque particuliers), le fait d'être employé, de vivre seul... En général, ces variables sont prises en compte sous forme d'indicateurs, comme le fait d'être au chômage ou de bénéficier de certaines prestations sociales réservées aux personnes défavorisées. Les Britanniques, dans le cadre de la réflexion sur l'allocation macro-économique des ressources aux régions⁷ ont tenté d'élaborer un indice synthétique de caractéristiques socio-économiques, mais cette réflexion n'a, semble-t-il, jamais été étendue au calcul de l'ajustement des risques dans le cadre plus micro-économique de l'allocation à des OGR. L'inconvénient majeur de certaines de ces variables est leur caractère transitoire (revenu, chômage) et endogène (l'individu peut les influencer).

Des variables caractérisant l'environnement peuvent aussi permettre de rendre compte de différences dans le coût de la prise en charge des patients. Ainsi, le degré d'urbanisation

⁶ Ce manque d'information des plans sur les patients et les soins effectivement administrés soulève des difficultés dans le cadre de l'estimation d'indicateurs de qualité.

⁷ Procédure RAWP : *Risk Allocation Working Party*

peut refléter des variations de prix des infrastructures ou de mode de recours liées à la densité et à la nature de l'offre. Toutefois, on hésite parfois à prendre en compte ce type de variable par crainte de figer des situations peu efficaces. Par exemple, si la densité plus élevée de praticiens en ville encourage des comportements d'induction de la demande, les différences de dépenses observées entre milieu rural et urbain ne sont pas justifiées du point de vue de l'efficacité. La question plus générale posée ici est de savoir si l'ajustement au risque ne doit prendre en compte que des variations d'état de santé, ou s'il doit prendre en compte des facteurs déterminants des coûts comme les prix et l'offre de soins (van de Ven *et al.*, 1995). Ceci dit, les variables géographiques (code postal) peuvent aussi permettre de rendre compte indirectement de différences de type socio-économique. Par exemple, Kronick *et al.* (1995,51) montrent que le nombre d'admissions et de journées d'hôpital est beaucoup plus élevé quand la proportion de personnes vivant sous le seuil de pauvreté dans une zone géographique augmente. Le choix d'inclure ou non des variables de type géographique est donc assez délicat.

L'ensemble de ces caractéristiques socio-démographiques (âge, sexe, région, degré d'urbanisation, revenu, chômage, niveau d'éducation, taille de la famille...) dont on peut imaginer qu'elles sont disponibles de manière routinière ou qu'elles peuvent être aisément collectées en respectant la vie privée des individus, ne permettent pourtant d'expliquer qu'une faible proportion de la variance totale (autour de 3%), soit entre 10 et 20% de la variance explicable (Newhouse, 1996, van de Ven et van Vliet, 1992).

Indicateurs de l'état de santé

L'ajustement cherche avant tout à rendre compte du niveau de risque des individus. Or, ce risque est largement lié à l'état de santé général. Une recherche foisonnante s'est développée ces dernières années, surtout aux Pays-Bas et aux Etats-Unis, pour élaborer des critères d'ajustement qui reflètent l'état de santé. Nous ne détaillons par ici les résultats⁸ mais nous rendrons compte des principales pistes de recherche et des enjeux qui président au choix des différentes variables.

- L'état de santé perçu (physique ou psychologique), ainsi que la mobilité, sont des informations en général obtenues dans des enquêtes, sur une base déclarative. Elles ont une bonne capacité prédictive mais elles restent en partie subjectives et peuvent

⁸ Lamers (1997) van de Ven et al. (1992) van de Ven et al. (1995) proposent des revues de littérature empirique détaillées.

aisément être manipulées. Au delà des problèmes techniques de collecte de l'information que leur utilisation poserait, il semble difficilement envisageable de les inclure dans une procédure d'ajustement au risque.

- Les mesures physiologiques (pression sanguine, cholestérol...) permettent d'expliquer une partie non négligeable de la variance explicable mais elles ont un coût de collecte très élevé et leur prise en compte nécessite d'avoir accès à des informations complémentaires qui permettraient de les interpréter : un niveau de cholestérol faible peut en effet résulter d'un traitement approprié (Newhouse, 1996). Or, une personne qui présente un taux de cholestérol faible parce qu'elle suit un traitement régulier consomme plus de ressources qu'une personne qui a "naturellement" un taux identique.
- Les variables de style de vie (obésité, consommation de tabac et d'alcool) semblent par contre améliorer faiblement le pouvoir prédictif des modèles (à court terme et au niveau individuel).
- La mortalité est corrélée avec les dépenses de santé qui sont souvent très élevées dans les mois qui précèdent un décès (Beck et Zweifel, 1995). Outre le fait que l'ajustement en fonction d'une telle variable est nécessairement réalisé ex post, la logique selon laquelle on devrait augmenter la rémunération d'un OGR quand le nombre de personnes décédées augmente semble contestable. De fait, les taux de mortalité sont parfois utilisés dans les processus d'allocation macro-économique (au Royaume-Uni) mais jamais dans des calculs d'allocations individuelles.
- L'utilisation passée de ressources médicales est un bon estimateur des dépenses futures. Il est même le meilleur lorsqu'il est utilisé seul. Van Vliet et van de Ven (1992) montrent par exemple que les 10% de la population qui ont les dépenses annuelles les plus élevées une année, auront pendant (au moins) les 4 années suivantes des dépenses en moyenne deux fois plus élevées que les personnes qui appartiennent aux mêmes groupes âge-sexe⁹. Toutefois, ce critère d'ajustement est critiquable à plusieurs égards. Tout d'abord, l'utilisation passée reflète l'ensemble des déterminants de la demande (y compris éventuellement l'offre). En lui accordant une importance trop grande, on peut pérenniser ou encourager des comportements de recours et des

⁹ On trouve un résultat inverse pour les personnes qui ont les dépenses les plus faibles. L'observation des dépenses passées est un moyen peu coûteux de sélection différentielle pour les OGR.

pratiques médicales inefficaces ou inflationnistes. En fait, on décourage les efforts de maîtrise des coûts puisque, plus les dépenses augmentent, plus la rémunération prospective des OGR est élevée (Jones, 1990). Ensuite, la capacité prédictive de l'utilisation passée dépend de la nature du besoin de santé. En effet, une consommation élevée une année peut s'expliquer par un problème aigu ou chronique. Dans le premier cas, la consommation une année n'est pas un aussi bon estimateur des dépenses subséquentes que dans le second.

- Les informations de nature diagnostique semblent pouvoir échapper à ces critiques et les travaux récents explorent cette piste de façon systématique. Des variables indicatrices de la présence de certaines pathologies diagnostiquées lors d'hospitalisations une année ou détectées au travers de la consommation de certains médicaments permettent d'améliorer la valeur prédictive des modèles pour les années qui suivent. On peut comptabiliser le nombre de pathologies chroniques et les pondérer par une variable de coût relatif (Kronick *et al.*, 1995, van Vliet et van de Ven, 1992). La recherche s'oriente vers l'élaboration de groupes de diagnostics qui donnent lieu à des prises en charge d'intensité comparable. Ash *et al.* (1989) ont développé le modèle de référence pour l'hôpital¹⁰ et des modèles ont aussi été construits pour des pathologies détectées en ambulatoire¹¹ (Lamers, 1997, pp. 29 et s., Newhouse et Chairman, 1998, pp. 75).

La principale critique adressée à ces modèles est qu'ils reposent sur des informations manipulables par les OGR ou les prestataires de soins. Les chercheurs tentent donc de classer les diagnostics en fonction du pouvoir discrétionnaire dont les médecins disposent pour les diagnostiquer et d'ajuster la pondération en fonction de ce critère.

Il reste cependant une difficulté à surmonter (Newhouse, 1994, pp. 141) : si les pratiques de traitement diffèrent radicalement entre par exemple les HMOs et des assureurs traditionnels, sur quelle référence doit-on se baser pour estimer le coût moyen de la prise en charge d'une pathologie ? Le moins élevé, au risque d'ériger une sous-estimation en norme ? La moyenne ? Ou doit-on estimer le coût moyen de la méthode de traitement qui fait l'objet du consensus scientifique le plus large ? Cette dernière méthode semble objectivement la plus juste, mais elle nécessite de faire appel

¹⁰ *Diagnostic Cost Groups DCGs.*

¹¹ *Ambulatory Cost Groups ACGs.*

à des compétences très spécifiques et complique plus encore le calcul des facteurs d'ajustement.

- L'invalidité, l'incapacité ou le handicap fonctionnel expliquent des variations interindividuelles significatives dans les dépenses de santé, y compris quand d'autres variables au pouvoir explicatif élevé, comme l'utilisation passée de ressources sont prises en compte. Ces informations sont, dans certains cas, recueillies pour des raisons administratives (versement de pensions) et leur utilisation semble être consensuelle (van de Ven et van Vliet, 1992).

En conclusion, signalons qu'à l'heure actuelle, l'utilisation des informations individuelles sur l'état de santé permet, au mieux, d'expliquer environ les deux tiers de la variance explicable sur des échantillons représentatifs de la population totale (van de Ven et van Vliet, 1992). L'utilisation de facteurs de risque multiples communs à des sous-populations d'assurés ne permet en effet jamais de tarifer la composante strictement individuelle du risque et donc d'intégrer la totalité de la variance explicable dans l'évaluation prospective du risque.

c) Bilan et perspectives

Le moyen théorique idéal d'éradication de la sélection, qui consisterait à garantir que les OGR soient parfaitement rémunérés en fonction des risques qu'ils assument et que cette rémunération les incite à maîtriser les coûts, semble donc un objectif irréaliste dans l'état actuel de la recherche. L'opinion de Luft qui qualifie la locution "actuarialement juste" d'oxymore (1995, pp.24) nous semble cependant par trop pessimiste. On peut, en premier lieu, raisonnablement estimer que les progrès de la science permettront à moyen terme de mieux prévoir le risque. Ensuite, de façon très pragmatique, si on instaure une capitation imparfaitement ajustée au risque, les OGR qui sont desservies par ce système ont intérêt à collaborer à l'effort de recherche pour apporter la preuve du fait qu'elles sont moins bien placées que d'autres au regard des critères utilisés, voire pour en proposer de nouveaux.

Il faut en réalité se demander si l'amélioration du processus d'ajustement au risque sera, à terme, suffisante pour éliminer les profits espérés d'une stratégie d'écrémage. L'OGR réalise en effet un arbitrage : il faut que le profit de la stratégie de sélection soit supérieur à son coût, coût de mise en œuvre de la sélection elle-même et coût de leur réputation.

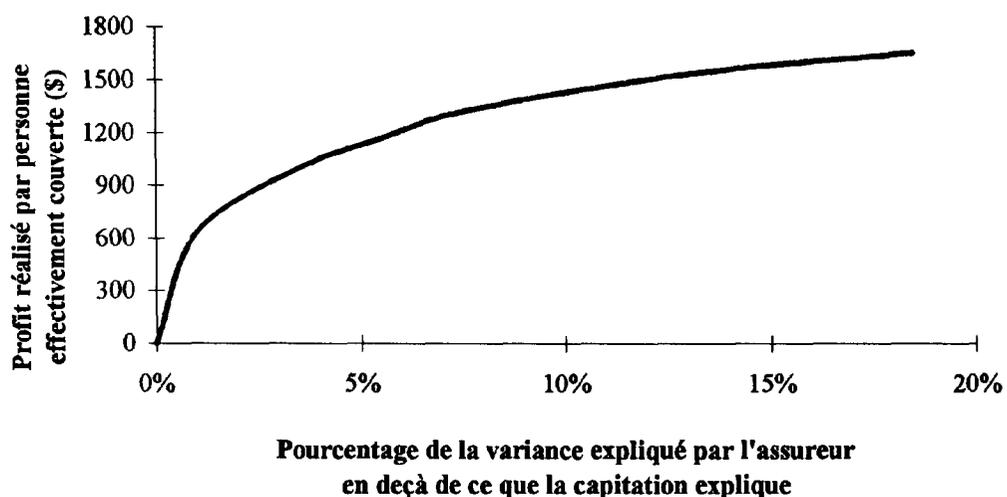
Plusieurs auteurs ont cherché à estimer le profit que les assureurs, ou les OGR pouvaient tirer de la sélection, mais leurs résultats sont divergents.

Newhouse *et al.* (1989) montrent que le profit de l'écrémage n'est pas une fonction linéaire de l'information privée que l'OGR détient sur le risque. Les auteurs considèrent une situation dans laquelle l'Etat verse à l'OGR une capitation ajustée au risque. Ils calculent le profit de la sélection en fonction du différentiel d'information dont bénéficie l'OGR. Ils se fondent sur les hypothèses comportementales suivantes :

1. l'OGR calcule les dépenses anticipées par assuré sur la base de l'information qu'il détient ;
2. si l'information qu'il détient n'est pas prise en compte dans la capitation ajustée au risque qu'il reçoit de l'Etat, il est donc en mesure de déterminer quelles sont les personnes pour lesquelles la capitation est sur ou sous-estimée¹² ;
3. il évince les personnes pour lesquelles il sait que la capitation est sous-estimée (cette stratégie est supposée de coût nul). Le profit qu'il réalise par personne effectivement couverte est alors calculé.

Comme le montre la figure 1, le profit de l'OGR croît avec la part de la variance des dépenses qu'il explique, au delà de la part prise en compte par la capitation ajustée au risque (fonction croissante). De plus, ce taux de croissance est décroissant (fonction concave).

Figure 1 : Profit de la sélection
D'après Newhouse et al. 1989, tableau 4, pp. 50



¹² On se place ici dans le cas où l'OGR reçoit une capitation ajustée au risque déterminée par une entité publique. La situation serait la même si l'assureur déterminait lui-même une prime fondée sur des critères simples (âge sexe) et utilisait l'information supplémentaire dont il dispose pour la sélection. En particulier, il pourrait atteindre le même niveau de profit.

Si l'OGR parvient à expliquer 1% de plus de la variance totale que la capitation ajustée au risque, il réalise un profit substantiel (alors que si la courbe de profit était convexe, il lui faudrait réaliser une estimation bien meilleure pour atteindre ce résultat).

Newhouse extrapole le résultat en affirmant qu'il sera toujours possible à l'OGR de réaliser un profit en améliorant de façon très marginale la capitation.

Cette analyse a été reprise par van de Ven, van Vliet, van Barnevelt et Lamers (1994b, 1995) qui démontrent tout d'abord (sur la base des mêmes chiffres) que Newhouse surestime le profit réalisé sur le premier pour-cent de variance expliquée "en plus" d'environ 60% lorsque la variance expliquée par la capitation est plus élevée¹³. Autrement dit, si la capitation ajustée au risque versée à l'OGR prend un plus grand nombre de facteurs de risque en compte, le profit que l'OGR réalise en l'améliorant de 1% est beaucoup plus faible. En outre, l'analyse de Newhouse ne tient pas compte de plusieurs éléments :

- d'une part, lorsque la variance expliquée par la capitation augmente :
 - l'information dont doit disposer l'OGR pour faire "mieux" est de plus en plus sophistiquée et coûteuse à acquérir,
 - de plus, van de Ven *et al.* (1994b) démontrent que la variance du profit estimé augmente. Autrement dit, lorsque la capitation ajustée au risque est améliorée, le profit réalisé en faisant 1% de mieux est de plus en plus aléatoire et de moins en moins élevé ;
- d'autre part, Newhouse fait implicitement l'hypothèse que seul un OGR sélectionne (parfaitement et sans coût) et que les autres ne réagissent pas, ce qui le conduit à surestimer le profit de la stratégie. En réalité, une stratégie de sélection ne fonctionne que dans la mesure où elle est mise en œuvre par un nombre limité d'OGR, voire par un seul OGR. Si plusieurs OGR mettent en œuvre des stratégies similaires pour sélectionner les risques, ils devront en partager le bénéfice alors que chacun devra assumer le coût de sa stratégie.

¹³ La démonstration est détaillée dans la note technique à laquelle van de Ven fait référence (1994).

En conclusion, il semble que l'ajustement au risque reste à l'heure actuelle nécessairement imparfait et laisse une marge de manoeuvre à l'OGR pour sélectionner les risques. Toutefois, il est permis d'espérer qu'à terme, une meilleure connaissance des déterminants de l'état de santé et de la consommation médicale, permettront une tarification des risques plus fine. A cet égard, il nous semble utile de réfuter l'idée selon laquelle une tarification au risque parfaite gomme toute incitation à l'efficacité. Certains auteurs, notamment Newhouse *et al.* (1989) laissent entendre que si l'intégralité de la variance est expliquée, on retombe dans une situation identique au remboursement des coûts (paiement rétrospectif). Même si on fait abstraction du fait qu'il est impossible de prévoir la variance pure, cet argument reste discutable. En effet, tout paiement déterminé de façon prospective est incitatif dans la mesure où l'OGR peut bénéficier d'une diminution des coûts et en tirer un profit. Si au contraire il ne maîtrise pas les dépenses, il devra assumer un éventuel déficit. Dès lors que le paiement est normatif, il est intrinsèquement incitatif.

Quoi qu'il en soit, à l'heure actuelle, si les techniques d'ajustement au risque ne sont pas parfaites, celles de la sélection ne le sont probablement pas non plus et le profit de la sélection n'est pas nécessairement élevé et certain. Enfin, et c'est l'aspect que nous étudions maintenant, il est possible d'élaborer des modes de financement mixtes qui contribuent à mitiger les effets de la sélection.

1.1.3 Les systèmes de paiement mixtes

Dans le début de cette analyse, nous avons implicitement étudié le cas où l'ajustement au risque, qu'il soit réalisé ex post ou ex ante, porte sur l'intégralité de la dépense. Le calcul de la somme que doit finalement recevoir l'OGR se fonde sur une estimation qui ne prend pas explicitement et directement en compte les dépenses auxquelles il fait face.

Or, l'analyse théorique oppose deux cas extrêmes :

- le paiement rétrospectif : en remboursant en fin de période les dépenses de l'OGR, on ne génère aucune incitation à l'efficacité ;
- le paiement prospectif ou normatif : en versant à l'OGR un paiement prospectif normatif, on crée une incitation maximale à l'efficacité productive, mais aussi à la sélection dans la mesure où la dotation n'est pas parfaitement ajustée aux risques.

Si on sait qu'il est impossible d'ajuster parfaitement le paiement au risque, on peut envisager, pour limiter la sélection, de mettre en place un mode de paiement en partie prospectif et en partie rétrospectif. Newhouse (1996) souligne qu'une solution intermédiaire, à savoir un système de paiement mixte, est probablement optimale du point de vue théorique.

Un certain nombre de raisons pragmatiques plaident aussi (selon les contextes) pour un mode de paiement mixte :

- Newhouse (1994) estime qu'un système mixte est toujours préférable à un système pur. En effet, lorsqu'on rembourse les dépenses observées, on fonde le paiement sur une estimation biaisée des coûts qui ne prend pas en compte l'efficacité ; lorsqu'on calcule un paiement prospectif, on risque de commettre des erreurs dans l'estimation du coût des traitements. Il en tire la conclusion que "*Pricing errors are inevitable...a mixed system is better than a pure system because it averages these errors*"¹⁴ (pp. 143) ;
- ensuite, selon le contexte institutionnel dans lequel évoluent les OGR, ils peuvent être amenés à supporter des coûts plus ou moins fixes et déterminés de façon exogènes. Dans la mesure où ils n'ont pas les moyens d'exercer un réel contrôle sur ces coûts, il ne semble pas logique de les responsabiliser financièrement sur la totalité de la dépense. Aux Pays-Bas, par exemple, toutes les caisses sont obligées de passer des contrats avec tous les hôpitaux et n'ont aucune marge de manoeuvre pour négocier les tarifs. Les dépenses hospitalières des caisses sont donc remboursées intégralement et ne font pas partie de l'ensemble des dépenses sur lesquelles elles sont responsabilisées ;
- enfin, dans les systèmes où le paiement prospectif est destiné à augmenter l'efficacité des organismes qui servent l'assurance maladie (comme les Pays-Bas ou la Belgique), on part d'une situation antérieure où les caisses n'avaient aucune incitation financière à gérer efficacement et recevaient des paiements entièrement rétrospectifs. Pour éviter de les déstabiliser et quand bien même on chercherait à tendre vers un paiement largement prospectif, il faut introduire la responsabilité financière progressivement.

¹⁴ Les erreurs de tarification sont inévitables, un système de paiement mixte est meilleur qu'un système pur parce qu'il permet de répartir, en moyenne, ces erreurs.

S'il on admet qu'un mode de rémunération mixte est optimal, il faut s'interroger sur la façon dont il doit être calibré. Newhouse (1996) propose une revue de littérature complète des modèles dont les conclusions pourraient éclairer ce choix¹⁵, mais il conclut que les solutions théoriques proposées ne sont pas adaptées au cas de l'assurance et que la détermination du partage optimal nécessitera de nombreuses expérimentations. Parmi les pistes de recherches envisagées :

- Newhouse (1994, pp. 142) propose d'augmenter la part du financement rétrospectif (remboursement) quand les dépenses sont très élevées afin de protéger les personnes plus vulnérables à la sélection ;
- Bowen (1995) suggère qu'on pourrait exclure certaines pathologies du paiement prospectif ;
- Luft (1995, pp.30) propose de rembourser les OGR pour la part des dépenses qui dépasse un certain plafond ou de constituer un pool de hauts risques vers lequel les OGR transféreraient ceux qu'ils estiment être des hauts risques et qui serait financé collectivement pas les OGR et / ou l'Etat.

Ce type de solution (notamment les deux dernières) a pour inconvénient majeur de désresponsabiliser assez largement les OGR sur la maîtrise des dépenses lorsque celles-ci sont très élevées. Ce faisant, ils ne sont pas incités à collecter l'information médico-économique sur les personnes concernées. Or, cette information permet d'étudier la façon dont les patients sont pris en charge, de contrôler les pratiques afin de les rendre plus efficaces et, à terme, elle est essentielle à la recherche de l'efficacité.

Ceci dit, aucun de ces auteurs n'explore en détail le fonctionnement de tels systèmes. Van Barneveld *et al.* (1996) constituent une exception à cet égard. Ils proposent un mécanisme original et en développent une analyse systématique sur des données empiriques. Le principe du mécanisme est le suivant : les auteurs partent d'un système où la capitation ajustée au risque (imparfaite) concerne l'ensemble de la population et proposent, pour en pallier les défauts, d'instaurer une mutualisation obligatoire des hauts risques qui fonctionne selon les principes exposés ci-après.

¹⁵ Entre autres, Laffont et Tirole (1993), Ma (1994) qui étudie le cas de l'hôpital, Ellis et Mc Guire (1990) pour les médecins.

1. L'OGR désigne périodiquement un groupe d'individus dont les dépenses vont être prises en charge collectivement par l'ensemble des OGR.
2. La proportion d'assurés dont les coûts seront mutualisés est fixée par les autorités pour chaque OGR.
3. Chacun d'entre eux est libre de déterminer ses critères de sélection et peut, en particulier, exploiter les informations dont il dispose sur ses propres assurés. A titre individuel, l'OGR a donc intérêt à déterminer qui va engendrer une dépense au dessus de la capitation avec autant de précision que possible. Ainsi, si les dépenses passées ne sont pas prises en compte dans la capitation ajustée au risque, l'OGR peut les utiliser pour déterminer qui doit être dans le pool. Les personnes dont l'OGR se doute au départ qu'elles vont entraîner des dépenses supérieures à la capitation ajustée au risque, sont donc intégrées dans ce pool.
4. La stratégie de sélection de l'OGR est détournée et n'aura pas de conséquences néfastes pour les assurés : puisque l'OGR ne supporte le coût que dans la mesure où il participe au financement du pool, il a un intérêt moindre à évincer les hauts risques. Un autre avantage de ce système est qu'il n'oblige pas le régulateur à déterminer des critères d'admission au pool qui pourraient être contestés ou devraient être négociés (comme c'est le cas si on choisit d'exclure certaines pathologies du système de paiement prospectif).
5. En fin d'année, les OGR partagent les dépenses du groupe poolé. La contribution d'un OGR ne dépend que des dépenses de pool et du nombre total de personnes qu'il couvre. Autrement dit, elle ne dépend pas directement des dépenses que ses propres assurés ont générées.

Ce mécanisme ne concerne pas spécifiquement les personnes qui ont des dépenses élevées. En effet, même si une personne génère ex post un déficit que l'OGR ne pouvait pas rationnellement anticiper, le risque d'écrémage disparaît puisque l'OGR n'était pas capable de détecter ce "haut risque". La fonction d'assurance est maintenue, au sens où l'OGR est responsable des variations imprévisibles des coûts autour de la moyenne, mais la tentation de recours à l'éviction des hauts risques a été limitée.

Concrètement, pour les personnes poolées, la mutualisation peut s'appliquer au coût total, au surcoût par rapport à la capitation reçue ou aux dépenses supportées par l'OGR au delà d'un plafond arbitraire. En outre, elle peut être partielle : l'OGR ne mutualise alors pas l'ensemble de la dépense mais une fraction plus ou moins élevée. Cette solution semble préférable pour ne pas générer de déresponsabilisation collective sur la maîtrise des coûts engendrés par les personnes concernées.

Enfin, on peut noter que ce mécanisme est neutre vis-à-vis des assurés eux-mêmes qui, typiquement, ne savent pas que leurs dépenses ont été poolées.

Une simulation de ce mécanisme est proposée dans l'article cité ci-dessus sur la base de données d'assurance collectées pour deux années consécutives (1992 et 1993).

Dans un premier temps, une capitation ajustée au risque (grossière) est calculée en fonction de l'âge et du sexe. Les auteurs font l'hypothèse que chaque OGR va intégrer, en 1993, dans le pool, les personnes qui, en 1992, ont eu les dépenses les plus élevées et que l'intégralité des dépenses qu'elles engendrent sera prise en charge par le pool.

On classe en ordre décroissant les différences entre les dépenses effectives de 1993 et les capitations, puis on regroupe les observations en classes qui représentent chacune 1% de la population totale.

Selon que le pool concerne 1 à 4% des assurés, les auteurs calculent :

- le pourcentage de la dépense totale poolée. Plus celui-ci est élevé, moins l'incitation de l'OGR à maîtriser les coûts est importante. Ce pourcentage est donc un des critères qui permet de déterminer (de façon pragmatique) le pourcentage de la population qui devrait être poolé ;
- la prime à payer par personne couverte pour financer le pool.

On étudie dans quelle mesure le profit / la perte prévisible par sous groupe varie, autrement dit, dans quelle mesure les incitations à écrémer diminuent.

Si 4% des personnes sont poolées, la dépense qui reste à la charge des OGR représente 75% de la dépense totale. Sur les 25% des dépenses poolées, la capitation âge sexe représente 6% de la dépense totale. On voit donc que le critère de choix utilisé par les OGR est simple et efficace (de leur point de vue). Cette impression est confirmée par le test suivant : lorsqu'on partitionne la population que l'OGR conserve à sa charge selon

différents critères - ce qu'il ferait s'il cherchait à trouver de nouveaux critères lui permettant de déterminer quelle catégorie non poolée il faudrait, soit chercher à détourner vers d'autres OGR, soit faire entrer dans le pool dans la période suivante - aucun profil de risque particulier ne se dégage.

Autrement dit, lorsque l'OGR s'est déchargé des 4% des personnes qui ont coûté le plus cher l'année précédente, les informations disponibles ne lui permettent plus de distinguer précisément ex ante des catégories de personnes qui génèrent une perte très élevée. Ceci démontre que le critère simple choisi est le bon.

Ce mécanisme incite-t-il réellement l'OGR à ne plus maîtriser des coûts pour les personnes poolées ? Notons d'abord que la proportion de personnes concernées par le mécanisme est faible : il est, dès lors, vraisemblable que les mesures mises en œuvre par l'OGR pour gérer les soins et augmenter son efficacité productive sur les 96% des personnes non poolées pourront aussi s'appliquer aux 4% poolées. L'argument selon lequel les soins des personnes dont les dépenses sont les plus élevées ne sont pas gérés doit donc être relativisé, en particulier si le pool ne prend en charge qu'une partie de leur dépense et non l'intégralité comme c'est le cas dans cette simulation.

Une dernière série de tests permet de vérifier la robustesse du mécanisme proposé. Sur la base des données disponibles, on génère 11 OGR fictifs auxquels on assigne par simulation des structures de risque délibérément différentes. Initialement, la différence entre le résultat moyen par assuré le plus élevé (profit) et le plus faible (perte) est de 500 florins (pour un coût moyen sur la population de 1700 florins). Si chacun des OGR poolé 4% de sa population, la différence est réduite à 340. Dans une dernière étape, on module le pourcentage poolé par plan en fonction de la perte nette par assuré enregistrée l'année précédente¹⁶. Ceci se justifie parce que la capitation ajustée au risque est très grossière et qu'une perte moyenne très élevée ne peut être entièrement imputée à une mauvaise gestion de l'OGR. En utilisant ce critère, la différence des profits tombe à 265 environ. Au total, la perte effectivement réalisée sur les 1% de personnes dont on avait anticipé, sur la base d'un critère simple, qu'elles engendreraient des coûts largement supérieurs à la capitation, diminue de 80%. Autrement dit, ce mécanisme n'exonère pas l'OGR de toute sa responsabilité dans la maîtrise des coûts, ce d'autant moins que des personnes non

¹⁶ Pour la détermination du pourcentage poolé par plan, le régulateur utilise une information assez simple (cf. van Barneveld, et al., 1996, pp. 141.)

poolées peuvent générer des pertes aussi élevées, sans que cela ait été prévisible : toute mesure destinée à mieux gérer les soins s'appliquera aussi à elles.

Cette analyse détaillée confirme que, même si l'ajustement au risque est imparfait, il est possible d'élaborer des mécanismes originaux qui permettent de contrer les stratégies d'écrémage des OGR. Celui qui est proposé ici repose sur une approche pragmatique et empirique. Les hypothèses sur l'information disponible et sur les comportements des différents agents semblent aussi réalistes.

Au terme de cette analyse, nous sommes en accord avec la position de van de Ven et al. : *"Although we are optimistic about the technical possibilities to prevent cream skimming, the realization in practice requires considerable effort"*¹⁷ (1995, pp.18).

Nous décrirons, dans la seconde partie de ce chapitre, les expériences en cours d'ajustement des risques qui témoignent de la diversité des mécanismes possibles, et qui confirment que cette recherche est prometteuse.

Toutefois, si l'ajustement au risque est une condition nécessaire, il fait que les assureurs et les assurés acceptent de se soumettre à d'autres contraintes.

1.2 Les contraintes supplémentaires

La détermination d'un mode de financement, à lui seul, ne permet pas d'enrayer les comportements de sélection. On observe une certaine convergence dans l'énoncé des autres conditions qui doivent être réunies pour obtenir que la concurrence fonctionne selon les critères souhaités. Nous les présentons ici en soulignant, quand cela sera nécessaire, en quoi les enjeux diffèrent entre les systèmes, initialement concurrentiels, dans lesquels on cherche à limiter les effets des stratégies de sélection, et les systèmes dans lesquels on introduit une forme de concurrence ou de responsabilité financière en cherchant à éviter la sélection.

¹⁷ *optimistes quant aux possibilités techniques d'éradiquer l'écrémage, ..(même si)..la réalisation pratique de cet objectif requiert un effort considérable.*

1.2.1 Définition du panier offert

Dans les deux modèles étudiés, la définition d'un panier standard apparaît comme une nécessité. Les enjeux d'une telle définition sont cependant différents selon le contexte.

Notons tout d'abord qu'imposer une standardisation des garanties constitue en soi un moyen réglementaire pour limiter la sélection des risques (Newhouse, 1994). Nous avons en effet vu qu'en profilant son offre, l'OGR ou l'assureur était en mesure d'attirer des personnes qui présentent un niveau de risque plus ou moins élevé. La standardisation des garanties limite donc cette sélection volontaire ou involontaire par les OGR.

Mais la standardisation est aussi considérée, pour des raisons techniques, comme une condition nécessaire au fonctionnement de l'ajustement au risque (Bowen, 1995, Luft, 1995, Swartz, 1995) même si certains auteurs, comme Bowen, estiment qu'il serait envisageable d'offrir un nombre limité de contrats différents. En effet, pour procéder à l'ajustement des risques, il faut pouvoir distinguer, dans les prix (ou les coûts) d'un OGR, ce qui relève de l'efficacité productive des effets de sélection. Si les contrats proposés par les plans diffèrent par les copaiements ou les niveaux de franchise, il faut pouvoir, en outre, estimer ce qui, dans la différence de prix, relève :

1. du fait que les sommes déboursées par l'OGR sont plus ou moins élevées ;
2. des conséquences sur les frais de gestion de la structure du contrat (les frais de gestion sont moins élevés toutes choses égales par ailleurs en cas de franchise puisque le nombre de dommages remboursés diminue) ;
3. de l'impact des différences de taux de couverture sur le risque moral.

En imposant la standardisation des contrats on peut espérer que le calcul de l'ajustement sera plus juste et - relativement - plus simple.

Les nombreux auteurs qui se penchent sur cette question de l'ajustement au risque s'accordent à dire qu'elle est une nécessité pour promouvoir l'efficacité productive et qu'elle doit s'accompagner d'une standardisation des bénéfices. Ils mettent en avant l'argument selon lequel cette standardisation représente un avantage pour le consommateur qui voit la transparence du marché augmenter et qui peut faire des choix rationnels. En revanche, ils passent sous silence le fait que, par rapport au modèle idéal

de concurrence, elle constitue une contrainte forte sur les choix des individus et se traduit, du point de vue théorique, par une restriction de l'efficacité allocative.

Les pays dans lesquels le contrôle de la sélection accompagne l'introduction d'une dose de concurrence partent en général d'une situation antérieure où toutes les personnes couvertes bénéficient des mêmes garanties. La définition d'un panier standard est une opération qui semble pouvoir soulever moins de difficultés que si l'on part d'un contexte initial où la liberté d'établir son panier couvert est totale. Ceci dit, la définition d'un panier standard nécessite de répondre à plusieurs questions.

1. Dans un premier temps, il faut définir le domaine sur lequel porte la concurrence. On peut, en particulier, exclure certains risques ou certaines populations vulnérables du champ soumis à concurrence. C'est l'option qui a été choisie aux Pays-Bas. Tous les résidents bénéficient de l'AWBZ¹⁸. La réforme inspirée des propositions de Dekker prévoyait d'intégrer l'AWBZ dans le champ concurrentiel mais cette option a finalement été rejetée. Deux arguments essentiels justifient une telle décision : tout d'abord, la qualité est un enjeu beaucoup plus grand pour les soins de long terme et les personnes qui en bénéficient sont probablement, à la fois plus vulnérables à cet égard, et moins à même de se défendre si un OGR ne garantit pas un niveau de qualité suffisant. Ensuite, les études économétriques sur données longitudinales (van Barneveld *et al.*, 1997) ont montré qu'il est difficile de déterminer une capitation ajustée au risque qui gomme toute incitation à la sélection différentielle des patients qui relèvent de ce régime.
2. De façon plus générale, déterminer le panier de biens et services sur lesquels la responsabilité de la collectivité est engagée nécessite de faire des choix (politiques) explicites. Le rapport Santé 2010 souligne que le panier couvert à l'heure actuel en France est très large et que la contrepartie est un taux de remboursement moyen moins élevé que dans un grand nombre de pays. "*Le resserrement de l'éventail pourra conduire à mieux couvrir certains biens au détriment d'autres services dont l'utilité médicale sera jugée moins importante (les cures thermales par exemple)*" (1993, pp.403).

¹⁸ Couverture socialisée et obligatoire des risques médicaux graves (rappel)

3. Enfin, il faut définir les processus de révision du panier au fur et à mesure que des techniques ou des produits nouveaux apparaissent.

La mise en œuvre des modes de financement requis pour limiter la sélection nécessite donc de standardiser les contrats d'assurance et de définir explicitement les biens et services couverts. Dans un contexte au départ concurrentiel, cela signifie que l'efficacité allocative doit être limitée afin d'obtenir l'efficacité productive. Dans un contexte où chacun bénéficie initialement d'une couverture identique, cette obligation peut permettre de redéfinir les priorités, ce qui est un avantage mais qui nécessite de trouver un consensus entre différents acteurs du système (prestataires de services, pouvoirs publics et assurés). Une telle démarche implique des arbitrages politiques souvent délicats. Ainsi aux Pays-Bas, le ministère a tenté d'éliminer les prothèses dentaires du panier standard et a été contraint l'année suivante de les réintroduire¹⁹.

1.2.2 Détermination du périmètre et autres conditions annexes

La seconde condition nécessaire au fonctionnement de l'ajustement au risque concerne la définition du périmètre de cet ajustement. Pour que l'ajustement ait un sens, il faut en effet imposer à un groupe de personnes et aux entités qui les couvrent de participer à l'ajustement au risque.

a) Obligation d'assurance ?

En effet, les assurés qui présentent de faibles niveaux de risque ont intérêt à obtenir la tarification la plus juste de leur risque. De leur point de vue, il est préférable que l'OGR qui les couvre sélectionne les risques. Si l'on admet la nécessité de juguler la sélection des risques afin de promouvoir l'efficacité productive, il faut élaborer un mécanisme de financement dans lequel les signaux prix (s'il y en a) rendent compte de différentiels d'efficacité productive et non de la capacité des OGR à sélectionner. Les modalités exactes de financement détermineront toutefois l'impact final sur les bas risques d'une telle mesure et la nécessité ou non de les contraindre à s'assurer. Si on part d'une situation antérieure où les OGR ne différencient pas spontanément les primes en fonction de critères de risque fins, et si on continue à imposer une forme de *community rating*,

¹⁹ En Israël, le processus de définition et de révision du panier a donné lieu à de nombreux débats. Le sous-titre d'un article qui traite du système israélien est instructif : "*explicit rationing of health care services in a universal system is easier said than done*" (Chinitz et Israeli, 1997).

comme c'est souvent le cas dans les mécanismes de péréquation élaborés aux USA, l'impact pour les bas risques n'est pas nécessairement élevé et peut être compensé à terme par une amélioration de la concurrence qui porte sur la gestion des processus de soins, la prévention... La plupart des systèmes qui émergent aux Etats-Unis s'organisent sur une base largement volontaire, à l'initiative des employeurs ou des centrales d'achat que nous avons évoquées dans le troisième chapitre. Les nombreux avantages que présentent ces formes d'assurance semblent donc, à l'heure actuelle, l'emporter sur les inconvénients pour les bas risques. Toutefois, si le coût net devient trop élevé pour certaines catégories de personnes, par rapport à ce qu'elles pourraient espérer payer dans un autre système que celui qui leur impose une forme de péréquation, et si elles ont la possibilité d'échapper au système, ou de ne pas s'assurer, le mécanisme d'ajustement ne peut fonctionner. On peut alors imposer une obligation d'assurance. Ceci est particulièrement vrai si on veut mettre en œuvre ou maintenir une participation des individus qui n'est pas fonction des risques mais, par exemple, du revenu.

Dans tous les cas, la couverture au sein du groupe concerné, qu'il soit constitué sur une base volontaire ou non, doit être virtuellement universelle (Bowen, 1995).

b) Obligation de participation et contraintes imposées aux OGR

Dans la mesure où les services de plusieurs OGR sont offerts aux membres du groupe d'assurés constitué²⁰, il faut évidemment que tous les OGR participent au mécanisme d'ajustement au risque et que tous soient soumis aux mêmes exigences vis-à-vis de cet ajustement des risques.

Au total, comme le note Bowen (1995, pp.34), l'ajustement du risque doit s'inscrire dans un marché défini. Ce marché peut couvrir tous les individus et tous les OGR, ou il peut être défini de façon beaucoup plus limitative (tous les employés d'une entreprise et tous les OGR qui leur proposent des contrats). En tout état de cause, l'entrée et la sortie du marché doivent être réglementés. Cette contrainte se fonde aussi sur des raisons d'ordre technique : pour que le mécanisme d'ajustement fonctionne, il faut que la durée de tous les contrats soit standardisée et que tous soient souscrits à la même période. Si ces conditions ne sont pas réunies, l'ajustement au risque est compliqué par le fait que la structure de risque de chaque OGR peut varier à tout moment et qu'il doit porter sur une

²⁰ Si les assurés n'ont pas le choix entre plusieurs prestataires, la question de la sélection ne se pose pas.

période différente pour chaque OGR. La durée des contrats est donc en général fixée à un an (mais prévoit des exceptions en cas d'entrée ou sortie pour des raisons justifiées, déménagement notamment). Pour ce qui est du renouvellement, on permet en général aux assurés de manifester leur choix pendant une période d'un mois (dite *d'open enrollment*), au terme de laquelle les anciens contrats arrivent à échéance et les nouveaux prennent effet²¹.

Enfin, pour que l'ajustement au risque prenne tout son sens, les OGR doivent respecter quelques obligations supplémentaires :

- couvrir toute personne du groupe constitué qui en fait la demande (*guaranteed issue*) ;
- renouveler par la suite les contrats des personnes qui le souhaitent (*guaranteed renewal*) ;
- ne pas exclure de la garantie des maladies avérées au moment où la personne entre dans le contrat (*pre-existing conditions*)²².

Ces contraintes réglementaires sont, comme la standardisation des contrats, couramment utilisées pour limiter la sélection par les assureurs, y compris lorsqu'il n'y a pas d'ajustement au risque (Kronick *et al.*, 1995, Newhouse, 1994). Toutefois, elles ne permettent pas à elles seules d'enrayer la sélection et elles restent nécessaires au fonctionnement de l'ajustement au risque.

Malgré les nombreuses contraintes qui leur sont imposées, nous verrons, au travers d'exemples précis de mise en œuvre de mécanismes d'ajustement au risque, que les OGR trouvent un intérêt commercial à offrir des contrats, qui contrebalance les contraintes qui leurs sont imposées et auxquelles ils acceptent en général de se soumettre.

Nous avons, dans cette première partie, présenté les enjeux et les conditions techniques de mise en œuvre d'une gestion de la concurrence efficace. L'élément principal du dispositif est le mode de financement conçu pour éviter que les OGR ne recourent à la sélection. L'objectif ultime est d'obtenir un financement prospectif véritablement actuariel

²¹ Cette tradition américaine, qui simplifie la gestion, avait émergé avant que la question de l'ajustement des risques ne se pose.

²² Il faut toutefois éviter les comportements trop opportunistes des assurés : s'ils ne sont pas obligés de souscrire un contrat et s'ils savent que l'assureur ne peut refuser de les couvrir, ils peuvent attendre pour souscrire l'assurance d'être malades. Luft (1995) suggère dans le cas où l'assurance n'est pas obligatoire, que les maladies avérées puissent être exclues pendant la première année d'assurance (délai de carence), mais qu'elle soient couvertes par la suite, y compris si l'assuré change de plan (il faut alors que l'information soit transmise).

des OGR. Plutôt qu'élaborer une parade réglementaire à chaque modalité de sélection, le régulateur cherche à éliminer les possibilités de réaliser les profits de court terme générés par de telles stratégies.

2. PRATIQUE DE L'AJUSTEMENT AU RISQUE ET DE LA CONCURRENCE TEMPEREE

La seconde partie de ce travail vise à recenser les systèmes dans lesquels un mode gestion de la sélection prévaut, qui s'apparente à celui que nous venons de décrire. Les pays développés de l'hémisphère nord, concernés par ce type de réforme, sont les Etats-Unis, Israël, la Belgique, l'Allemagne, les Pays-Bas, l'Irlande et la Suisse.

En décrivant ces systèmes et leurs évolutions ou réformes récentes, nous verrons en quoi les conditions que nous avons identifiées comme nécessaires au fonctionnement d'une concurrence limitant la sélection sont remplies ou non.

Nous nous attacherons aussi à décrire le fonctionnement concret des différentes méthodes d'ajustement au risque, dont nous n'avons pour l'instant fait qu'esquisser les principes²³. Pour finir, nous illustrerons par cette démarche l'idée selon laquelle l'arbitrage entre les exigences de maîtrise des coûts et la nécessité de lutter contre la sélection, induit une convergence dans la structure des systèmes d'assurance maladie à priori très éloignés. En fait, les deux modèles que nous avons opposés de "limitation de la sélection dans un univers concurrentiel" (modèle américain) et "d'introduction de la concurrence" (faisant plutôt référence à l'expérience des caisses publiques aux Pays-Bas), sont des exemples illustratifs de deux groupes de systèmes que nous étudions successivement :

- les systèmes de type concurrentiel dans lesquels on introduit l'ajustement des risques (2.1) ;
- les systèmes dans lesquels un organisme central va déléguer la gestion des risques à des opérateurs en concurrence (2.2).

²³ Nous renverrons le cas échéant, pour le détail de la présentation, à des encadrés techniques ou à des sous paragraphes clairement identifiés.

2.1 Introduction de l'ajustement en complément de la concurrence

La première série de cas que nous présentons regroupe un ensemble de pays dans lesquels l'assurance est traditionnellement souscrite sur une base volontaire et plutôt concurrentielle. Elle est financée par des primes dont s'acquittent les individus et qui reflètent peu ou prou leur niveau de risque. L'ajustement des risques est progressivement apparu comme une condition nécessaire au bon fonctionnement du marché. L'exemple le plus emblématique est celui des Etats-Unis. L'Irlande, dans un contexte initialement monopolistique, cherche maintenant à promouvoir la concurrence sur le marché de l'assurance volontaire et a mis en œuvre un mécanisme d'ajustement destiné à éviter les effets jugés néfastes de la sélection. Enfin, la Suisse, qui a introduit l'assurance universelle en 1994, tout en maintenant une organisation concurrentielle, a, dans le même temps, mis en place un ajustement des risques.

2.1.1 Les Etats-Unis

Dans le système américain, ces dernières années, la sélection est devenue un enjeu de plus en plus important sur l'ensemble des marchés d'assurance volontaire (a). Des formes d'organisation émergent sur les marchés privés qui apportent une réponse locale et limitée aux problèmes posés par la sélection dans le cadre de la *managed competition* (b).

a) La généralisation du problème de la sélection

Aux Etats-Unis, les marchés sur lesquels la sélection est la plus flagrante sont les marchés individuels et celui des petites entreprises. Une des raisons pour lesquelles ces marchés sont particulièrement vulnérables est que les assurés ne sont pas groupés par d'autres liens que ceux de l'assurance (Gauthier *et al.*, 1995). Lorsque les risques sont élevés, les personnes concernées ne trouvent pas d'assureur qui accepte de les couvrir ou les contrats ne sont pas abordables. Il n'existe pas de mécanisme structurel qui permette de grouper les risques comme dans les grandes entreprises.

Par contraste, jusqu'à une période assez récente, le marché des grandes entreprises était relativement épargné car les risques y sont plus stables et plus homogènes. Les bénéficiaires de l'assurance de groupe ont un emploi stable à plein temps (les autres catégories d'employés sont rarement couvertes), ce qui constitue un premier élément

d'homogénéisation des risques, même s'il ne s'applique pas à l'ensemble des ayants droit (autres membres du ménage). La péréquation entre niveaux de risque est donc plus aisée, parce que les personnes s'assurent de façon solidaire et non à titre individuel. Ceci dit, la sélection est devenue un enjeu au fur et à mesure que les employeurs ont multiplié les choix d'assurance des individus (Jones, 1989, Jones, 1990, Luft, 1995). Même si la couverture reste quasiment universelle au sein des grandes entreprises, dès lors que plusieurs contrats sont offerts, les conditions du marché individuel sont partiellement recréées et nous avons vu que la segmentation spontanée compromettrait le fonctionnement de ces systèmes (cas de Harvard, chapitre 5).

Aux Etats-Unis, sur l'ensemble des marchés de l'assurance volontaire, on cherche à restreindre la capacité des plans à segmenter les risques afin de les obliger à se concurrencer sur les prix, le service et la qualité.

b) L'émergence de la "managed competition"

L'assurance volontaire est en général souscrite à titre individuel dans une gamme de contrats sélectionnés, négociés et partiellement financés par les gros employeurs (privés ou publics). Les Etats encouragent la création de centrales d'achats qui jouent auprès des PME un rôle similaire.

Selon la terminologie proposée par Enthoven (1988, 1994), les employeurs et les centrales d'achat doivent jouer le rôle d'acheteur avisé ou de "sponsor" dans la "concurrence régulée" (*managed competition*). Ce modèle idéal stipule que les consommateurs sont sensibilisés au prix à payer, qu'ils peuvent exercer un réel choix qui incite les assureurs à offrir des contrats (et éventuellement des services) efficaces. Le "sponsor" est un intermédiaire entre assureurs et assurés qui doit, pour mener à bien sa tâche :

- présélectionner un nombre limité de plans en amont (il est en mesure de se doter d'outils plus performants que les assurés pour juger de leur qualité) ;
- prendre des mesures pour que les salariés puissent faire leur choix sur la base de critères simples, transparents, et en nombre limité. Pour ce faire, il peut stimuler ou exiger la standardisation des contrats et imposer des conditions de tarification - en général de type *community rating* - qui sont uniformes ;
- imposer aux plans d'accepter tout membre du groupe (*guaranteed issue et renewal*).

En outre, le sponsor, par définition, effectue une péréquation des risques au sein d'un groupe relativement moins homogène que s'il était constitué sur un marché d'assurance individuelle. Il centralise les demandes et obtient ainsi un pouvoir de marché qui lui permet de négocier les prix, notamment parce qu'il diminue les frais de gestion et de marketing des plans (Diamond, 1994). Pour garantir que la concurrence se déroule sur les bons critères, il faut enfin que les sponsors mettent en place une compensation des risques entre les plans participants.

L'expérience montre que les employeurs ne jouent pas leur rôle de sponsor. Seuls 43% des salariés ont le choix entre plusieurs plans (NEHIS 1993²⁴). Hibbard, Jewett *et al.* (1997) montrent que la plupart des employeurs ignore les indicateurs de résultats médicaux sur les hôpitaux proposés par les plans. En outre, seuls 31% des employeurs diffusent une information aux salariés (dite *performance information*), et encore, cette information privilégie les performances en termes de services (rapidité à gérer les remboursements) et très rarement sur les performances cliniques.

Dans ce contexte, les sponsors ayant mis en place une compensation des risques restent encore assez rares. Un total de cinq exemples est recensé dans la littérature. Nous avons choisi, lorsque l'information est disponible²⁵, d'en présenter le fonctionnement de façon assez détaillée, pour témoigner de la diversité des solutions possibles.

le Health Insurance Plan of California (HIPC)

Le HIPC est une centrale d'achat californienne, destinée aux PME et gérée par le MRMIB²⁶ dont le mécanisme d'ajustement, mis en place en 1996, est décrit et analysé par Shewry *et al.* (1996) et Bertko (1998). Le mécanisme a été élaboré en concertation par le MRMIB, les plans qui participaient à la centrale d'achat et des consultants.

Afin d'en comprendre le fonctionnement, il faut noter au préalable que, pour participer au HIPC, les plans doivent se soumettre à plusieurs obligations (*open enrollement period, guaranteed issue et renewal*). De plus, ils doivent utiliser un *community rating* ajusté sur la base de l'âge, de la localisation, et de la structure de la famille (selon quatre catégories qui ne prennent pas en compte explicitement le nombre d'enfants). Ces facteurs de risque

²⁴ National Employer's Health Insurance Survey

²⁵ Le système de calPers, entité publique qui joue le rôle de sponsor pour les employés de l'Etat de Californie depuis 1962 et celui que l'Etat de New-York a mis en place sur le marché des PME ne seront ainsi pas présentés.

²⁶ Managed Risk Medical Insurance Board

sont donc inclus dans la prime facturée aux assurés et le mécanisme d'ajustement doit porter sur les facteurs non tarifés. Autrement dit, le mécanisme doit gommer, pour un individu résidant dans un endroit particulier et ayant un âge précis, ce qui, dans la différence de prix d'un contrat à l'autre, relève du fait que l'ensemble des personnes couvertes par chaque assureur sont, par exemple, en plus mauvaise santé.

Après concertation, les parties concernées ont choisi d'ajuster les risques en fonction de trois critères non tarifés : le nombre d'enfants couverts, le sexe, et certaines pathologies coûteuses. On calcule pour chaque plan un indice de risque relatif (RAV). S'il est trop éloigné de la moyenne, égale à 1, le plan est taxé ou subventionné.

Encadré 1 Ajustement au risque dans le MRMIB

Pour chaque plan, le MRMIB calcule un *RAV* (*risk assessment value*), fondé sur ces trois composantes, qui reflète la structure de risque relative du plan par rapport à celle de la population totale couverte. Le choix des diagnostics thérapeutiques pris en compte a été effectué sur la base de données de quatre plans représentant 7,2 millions d'années-personnes, et 120 diagnostics ont été sélectionnés sur la base de deux critères :

- ils entraînent des consommations supérieures à la moyenne ;
- leur dénombrement n'est pas trop dépendant des pratiques de codage des plans ou des praticiens. Autrement dit, les possibilités de manipuler l'information pour exhiber une structure de risque plus défavorable qu'elle ne l'est réellement, afin d'obtenir une compensation plus grande, sont limitées. Les chiffres déclarés par les plans sont, en outre, soumis à un audit externe.

Pour finir, seuls des diagnostics entraînant hospitalisation ont été retenus, car les plans n'ont pas d'information suffisamment précise en ambulatoire. Les principales indications retenues sont : SIDA, cancer du sein, maladies pulmonaires chroniques, maladies du foie, maladies ischémiques, arythmies, troubles du pancréas, syndrome de détresse respiratoire.

Une fois la proportion de personnes par diagnostic connue, celle-ci est multipliée par le ratio (coût moyen des soins des personnes atteintes / coût moyen des soins dans la base). On alloue aux individus qui n'ont aucun des diagnostics recensés un poids tel que la moyenne des poids sur la population vaut 1. Ce système revient à allouer aux personnes en bonne santé une dotation déterminée "par défaut" quand on a mis de côté des sommes pour les autres. On pourrait aussi envisager d'allouer aux personnes en bonne santé un poids qui reflète leur coût relatif par rapport aux autres. Cette solution adoptée par exemple aux Pays-Bas (pour d'autres facteurs de risque), semble plus logique sur le plan économique, mais elle favorise moins explicitement les personnes qui génèrent des coûts élevés (ici les malades).

Enfin, on recalcule pour chaque plan en fonction du nombre de personnes atteintes de chaque diagnostic un poids moyen. De façon similaire, on estime un poids relatif de chaque plan qui rend compte de la répartition des sexes et du nombre d'enfants couverts. En multipliant les poids et en normant à 1, on obtient le *RAV* de chaque plan. Il a été convenu de ne mettre en œuvre l'ajustement que pour les plans dits *outliers* (ceux qui sortent de la norme ou qui sont assez éloignés de la moyenne) et dont le *RAV* est éloigné de plus ou moins 5% du *RAV* total qui est par construction égal à 1.

Reste alors à estimer la somme qu'il faudrait transférer ou prélever sur les *outliers* pour les ramener aux bornes de 0.95 et 1.05. Pour ce faire, on calcule une prime moyenne théorique par plan, qui prenne en compte les facteurs qu'il est en droit de tarifer afin que le transfert ne porte que sur les facteurs non tarifables. On calcule un *rated risk factor (RRF)*, qui reflète, pour chaque plan, sa structure âge / région / type de famille, par rapport à celle de l'ensemble de la population couverte. On postule alors que la prime moyenne théorique qu'un plan devrait facturer si sa structure de risque était identique à celle de la population totale est égale à la prime moyenne sur tous les plans (P_M), multipliée par son *RRF*. Compte tenu de sa structure de risque non tarifable, la prime qu'il doit (théoriquement) faire payer pour parvenir à l'équilibre est égale à :

$$RAV_i \times P_M \times RRF_i$$

Si son *RAV* est supérieur à 1.05, la subvention qu'il va recevoir par personne couverte est :

$$(RAV_i - 1.05) \times P_M \times RRF_i$$

Si au contraire son *RAV* est inférieur à la borne limite, il va payer par personne couverte :

$$(0.95 - RAV_i) \times P_M \times RRF_i$$

Des mécanismes sont prévus pour garantir que le fond de péréquation sera finalement à l'équilibre : si on estime par exemple que les sommes prélevées sont insuffisantes pour couvrir les subventions, on prélève des sommes sur les plans dont le *RAV* est légèrement supérieur à 0.95.

En dernier lieu, il reste à expliquer la temporalité de ce mécanisme : l'ajustement calculé une année est en fait mis en œuvre l'année suivante. Par exemple, en t on calcule que le plan i devrait recevoir (ou payer) 10\$ par mois et par personne sur la base des informations collectées en $t-1$. Lorsque la période de renouvellement des contrats arrive, le plan propose pour $t+1$ une prime qui prend en compte ce transfert. Il peut choisir de répercuter intégralement ou non ce transfert dans la prime proposée mais il paiera ou recevra effectivement via le HIPC par personne couverte et par mois la somme arrêtée l'année précédente, transfert compris. Ce mécanisme d'ajustement au risque fonctionne donc de façon comparable à l'*experience rating* : on récupère ou répercute sur l'année suivante les sur- ou sous-dotations de l'année précédente. Pour que le mécanisme fonctionne, il faut donc que les taux de *turn over* ne soient pas trop élevés, même si des ajustements supplémentaires sont effectués en fin d'année ($t+1$). Il n'en reste pas moins que la gestion globale du système est complexe et que l'ajustement au risque est effectué sur la base d'informations collectées deux années auparavant, ce qui complique en particulier la prise en compte de l'entrée ou de la sortie de plans.

Le bilan global de l'expérience après deux années complètes d'existence est positif. La taille de la population totale couverte dans le HIPC a considérablement augmenté et l'ensemble des acteurs en présence est satisfait (Bertko et Hunt, 1998). Quelques défections initiales de plans ont été constatées mais d'autres plans (de taille importante) ont rejoint le HIPC, et le nombre total de plans offerts est resté stable, entre 20 et 25. Les résultats de l'ajustement montrent que les HMO ont en général des structures de risque meilleures (7 sur 22 ont un *RAV* inférieur à 0.95 la première année), et qu'un des deux PPOs offert a dû être compensé assez fortement la première année (\$46 par contrat et par mois). On peut voir dans cet état de fait la conséquence d'un effet d'auto-sélection spontané, qui traduit encore une fois que les plus malades préfèrent les formes plus souples de prise en charge. Au total, la première année, 1% des primes ont été transférées. La seconde année, les transferts totaux ont représenté 0,15% des primes et la

gamme des reversements s'est considérablement rétrécie. Quatre HMO ont reversé des sommes et un PPO a été subventionné. Un des objectifs recherchés, à savoir maintenir un choix aux assurés entre HMO et PPO, a donc été atteint. Les premiers sont plus souvent mis à contribution et les seconds continuent à exiger des primes relativement élevées car l'ajustement est incomplet, mais tous y gagnent en termes de parts de marché et acceptent de jouer un jeu dont le fonctionnement a été élaboré sur la base d'une concertation des acteurs en présence. Tous participent aussi à la recherche pour améliorer le fonctionnement du mécanisme d'ajustement.

le *Buyers Health Care Action Group* (BHCAG)

Le BHCAG du Minnesota regroupe les employés de 28 grosses entreprises et négocie directement l'achat de soins à des groupes médicaux et médico-hospitaliers (systèmes de soins, SdS ci-dessous). Il ajuste les paiements en fonction des risques. Ce système est toutefois assez particulier puisque les SdS sont payés à l'acte et ne sont donc pas "mis au risque" dans l'acception traditionnelle du terme (pas de capitation). L'analyse suivante résume le fonctionnement du système sur la base d'informations collectées par Knutson (1998), Newhouse (1998), Dunn (1998a et 1998b).

Chaque année les SdS proposent un prix mensuel pour prendre en charge un patient "moyen"²⁷. Chaque employeur détermine sa contribution théorique. Celle-ci est en général inférieure au prix le plus faible proposé par les SdS afin d'encourager les employés à faire du prix un critère de choix central. Chaque personne couverte choisit son SdS pour une année et paie à son employeur la différence. Par la suite, l'employeur paie directement les SdS à l'acte.

Une procédure similaire à celle du HIPC est utilisée pour calculer un poids relatif par individu qui reflète son état de santé, puis un poids qui reflète le risque relatif du SdS par rapport à la population totale couverte, dont le risque est normé à 1. Les poids sont plus précisément calculés par groupe de diagnostics ambulatoires (ACG). On pondère la proportion de personnes dans un ACG par le coût moyen de prise en charge observé dans les différents SdS rapporté aux coûts de prise en charge de l'ensemble de la population. Comme les SdS sont payés à l'acte, ils sont capables de faire remonter

²⁷ Hors soins exceptionnels : le BHCAG prend en charge par un autre mécanisme 90% des dépenses qui dépassent un seuil donné.

l'information médico-économique nécessaire. Ces poids sont calculés tous les trimestres sur la base d'informations collectées sur les douze mois précédents.

Le prix mensuel proposé par le plan est multiplié par son risque relatif et constitue pour lui l'objectif à atteindre. Chaque trimestre, on calcule la moyenne des paiements mensuels effectués par le plan sur les douze mois précédents et on la compare à l'objectif calculé pour la même période. Si les paiements sont supérieurs à l'objectif, les honoraires et les prix de l'ensemble des services²⁸ sont ajustés d'un pourcentage tel que, si cette nouvelle grille avait été appliquée, l'objectif aurait été atteint. Inversement, si les paiements moyens sont inférieurs à l'objectif, les prix unitaires sont augmentés. La nouvelle grille de tarifs s'applique pour le trimestre suivant. Autrement dit, si, compte tenu de son niveau de risque, le SdS a utilisé "trop" de soins, il est sanctionné et verra ses prix diminués ; s'il a fait "mieux", la rémunération des prestataires augmente. La différence entre l'objectif et le résultat est mesurée de façon rétrospective mais la sanction ou la récompense sont prospectives.

A l'occasion de la première réévaluation tarifaire, début 1997, on a constaté que les niveaux de risques étaient très différents d'un SdS à l'autre (avec une différence de 35% entre le plus faible et le plus élevé) et les prix ont été ajustés de -10%²⁹ à +7.7%. Le bilan de 1997 a montré que les paiements effectués par les employeurs avaient diminué de 9% par rapport à l'année précédente, le système a donc permis de maîtriser les dépenses. Il a depuis été affiné et il est en cours d'évaluation. Il avait été initialement jugé très complexe, mais il semble que les SdS sont satisfaits (aucun n'a quitté le système et de nouveaux SdS ont même rejoint le BHCAG).

la *Washington State Health Care Authority* (HCA)

La HCA est le sponsor de différentes populations de l'Etat de Washington, dont les employés publics (Dunn, 1998a, Dunn, 1998b, Newhouse et Chairman, 1998, Wilson, et al., 1998). Le mécanisme d'ajustement au risque est en cours de révision. Il ne se fondait avant le début de 1998 que sur un modèle démographique et il intègre progressivement des informations de nature diagnostique. Lorsque ce modèle sera mis en place, les

²⁸ Le groupe médical peut toutefois exclure certains prestataires ou institutions de ce système de point flottant (système utilisé en Allemagne où les honoraires de certains praticiens sont ajustés à la hausse ou à la baisse si leur somme n'est pas égale à une enveloppe macro-économique déterminée). Entre 10 et 30% des prestataires par SdS ne sont pas concernés, mais, dans ce cas, la variation des prix pour les autres doit être plus élevée.

²⁹ Une limite à l'ajustement à la baisse était imposée.

variables d'ajustement seront : l'âge, le sexe, le statut vis à vis de l'assurance (employé, conjoint, dépendant), la structure du ménage et des variables de statut (retraité par exemple), l'information médicale étant prise en compte sous la forme de DRG.

Le déroulement du mécanisme est le suivant : chaque année, les plans proposent une prime par assuré moyen³⁰. Chaque personne effectue son choix de plan et paie la différence entre une contribution de l'employeur déterminée par la HCA et la prime affichée par le plan. Par la suite, les paiements de la HCA versés aux plans sont basés sur les prix proposés par les plans et ajustés au risque. Le modèle utilise des informations démographiques et de diagnostics regroupés par niveaux de coût associés. Par catégorie de risque, on estime une probabilité d'avoir des dépenses positives puis une dépense moyenne quand la personne consomme effectivement. Sur la base de ces informations, un indice de risque relatif pour l'individu est calculé et agrégé au niveau du plan. La contribution totale de la HCA est redistribuée en fonction du risque relatif de chaque plan : le processus d'ajustement consiste à ré-allouer des fonds mais il est neutre au niveau global. Une simulation effectuée en 1997 montre que les assureurs, lorsque l'ajustement est basé sur le modèle démographique, reçoivent entre 90 et 112% de la prime qu'ils proposent dans le processus d'appel d'offre. En ajoutant les informations de diagnostic, la gamme s'élargit (65% à 130%)³¹. Les simulations montrent aussi que ce dernier modèle prévoit mieux les différences de structure de risque entre les coûts.

En conclusion de ces trois exemples, il nous semble intéressant de souligner ou de rappeler plusieurs points de convergence.

1. Dans tous les cas, la motivation principale pour mettre en oeuvre l'ajustement au risque est la crainte de voir se développer la sélection des risques. La population couverte est relativement bien définie, assez large et hétérogène. L'ajustement se fait dans un contexte où les contrats sont standardisés ; les acteurs offrant leurs services sont, chaque année, à la même époque, obligés d'accepter de prendre en charge toute personne qui se présente (*open enrollement*). Ils n'ont pas le droit de tarifier certains facteurs de risque (puisque le *community rating*, voire un prix unique pour tous les assurés sont imposés), mais l'ajustement doit théoriquement les compenser de façon adéquate.

³⁰ Qui varie dans les mêmes proportions pour tous les plans afin de prendre en compte la structure de la famille.

³¹ Ce modèle est mis en oeuvre progressivement et sera utilisé en l'an 2000

2. Pour ce qui est de l'ajustement lui même, les différents programmes semblent tous converger, malgré les difficultés que cela pose, vers une utilisation de l'information de nature diagnostique associée à des coûts relatifs.
3. A l'exception du BHCAG, les modèles se veulent prospectifs. L'information de nature médicale sur le structure de risque est, certes, collectée deux années avant l'ajustement. Mais, pour la période sur laquelle porte l'ajustement :
 - les versements se fondent sur le nombre de personnes effectivement couvertes ;
 - les paiements aux plans sont déterminés ex ante et non en fin de période lorsque les coûts sont observés.

Les modèles reposent donc fortement (et implicitement) sur l'hypothèse que les structures de risque médical des populations couvertes par les différents plans sont stables, ce qui ne va pas sans poser de problèmes lorsque de nouveaux participants rejoignent le champ d'application. En particulier, on peut redouter ce qui se passerait si un nouvel entrant générerait un *turn over* très élevé pour une catégorie de risque particulière (si par exemple, un entrant attirait tous les ménages jeunes et sans enfant). Le BHCAG évite en partie cet écueil car il a adopté un ajustement plus fréquent et il utilise des informations plus récentes (cet ajustement n'est toutefois pas réellement prospectif, comme nous l'avons souligné). Il fait cependant face à des difficultés spécifiques : les ajustements doivent prendre en compte une estimation des dépenses non présentées au remboursement.

4. Quoi qu'il en soit, dans tous les cas, les prestataires de service en présence ont quasiment tous accepté les contraintes qui leur étaient imposées, probablement parce qu'ils étaient largement associés au processus d'élaboration du mécanisme. De plus, de nouveaux prestataires ont proposé leurs services après que le mécanisme a été implanté.
5. Enfin, une étude plus qualitative a permis de montrer que l'existence d'un ajustement diminuait la crainte des plans d'être victimes de stratégies de sélection de leurs concurrents ou d'effets de segmentation spontanée (Knutson, 1998) et pouvait les conduire à limiter leur propre stratégie de sélection (Bertko et Hunt, 1998).
6. Les employeurs et les employés n'ont pas souffert de l'implantation des mécanismes et s'en déclarent globalement plutôt satisfaits (Dunn, 1998b, pp.145).

7. La principale leçon de ces expériences est toutefois que l'application concrète de l'ajustement au risque nécessite : "*a series of trade-offs between risk-adjustment theory, data availability, [and] the willingness of participants to play the game*"³² (Bertko et Hunt, 1998).

2.1.2 Irlande

En Irlande, comme aux Etats-Unis, l'assurance est volontaire. En revanche, elle ne porte pas sur l'intégralité de la couverture comme nous l'avons vu dans le chapitre sept. Néanmoins, elle participe à l'équilibre général du système et le ministère de la santé a jugé bon, pour préserver cet équilibre, de mettre en place un ajustement des risques lorsque le marché s'est ouvert à la concurrence.

a) La fin du monopole de l'assurance volontaire

L'assurance privée volontaire est, depuis 1996, soumise à un système de péréquation. Depuis 1957, une entreprise publique d'assurance maladie, le *Voluntary Health Insurance Board* (VHI) détenait légalement un quasi-monopole sur le marché de l'assurance maladie. Traditionnellement, tous les assurés payaient la même prime et l'assureur ne pouvait refuser de couvrir une personne ou de renouveler son contrat. Ce mode de fonctionnement n'était pas imposé par une réglementation spécifique. En 1992, la troisième directive européenne "non vie" impose aux Etats membres d'ouvrir les marchés d'assurance aux compagnies d'assurance européennes. Le texte prévoit toutefois que certaines dispositions spécifiques mises en place par les Etats sur le marché de la maladie pour protéger le public pourront s'appliquer aux entrants sur le marché. Le *Health Insurance Act* de 1994 fixe explicitement un certain nombre de principes :

- *community rating*. L'assureur ne peut faire varier la prime en fonction du sexe, de l'orientation sexuelle (sic) ou de l'état de santé actuel ou prévisible des assurés. Quelques exceptions sont prévues : les enfants et les étudiants couverts par leurs parents peuvent bénéficier d'un abattement de 50% maximum et les contrats de groupe peuvent obtenir une réduction de 10% maximum.

³² Une série d'arbitrages entre la théorie de l'ajustement au risque, la disponibilité des données et la volonté des participants de jouer le jeu.

- *open enrollment et guaranteed renewal* : toute personne de moins de 65 ans a droit à l'assurance et aucune personne ne peut se voir refuser le renouvellement de son contrat. Les assureurs ont le droit d'imposer les restrictions sur la couverture des pathologies déclarées, mais celles-ci sont strictement réglementées.

Les *Health Insurance Regulations* de 1996 introduisent, à l'initiative du ministère de la santé irlandais, un mécanisme de péréquation des risques entre assureurs afin d'éviter que la concurrence ne passe par la sélection. Avant de présenter ce mécanisme, notons que la réforme du fonctionnement du marché de l'assurance trouve son origine dans la réglementation européenne. Une institution publique en situation de quasi-monopole doit s'ouvrir à la concurrence d'assureurs étrangers. Le gouvernement autorise cette concurrence mais cherche à préserver ce qui est perçu comme un avantage acquis : une tarification de l'assurance indépendante des risques. La dynamique est donc assez différente de celle présente dans d'autres pays : l'introduction de la concurrence n'est ici pas perçue comme une nécessité pour maîtriser les coûts. Ceci s'explique probablement en partie par le fait que l'assurance est volontaire, financée de façon privée, qu'elle ne conditionne pas l'accès au système de soins et qu'elle est plutôt réservée à une partie aisée de la population. Toutefois, nous avons vu que l'assurance privée contribuait à alléger les charges de l'Etat (puisque'elle finance une partie de l'hospitalisation par ailleurs gratuite pour tous). Le maintien du système et de la forme de solidarité qu'il génère entre différents niveaux de risque est jugé primordial.

b) le mécanisme d'ajustement

Un rapport élaboré par un groupe d'experts indépendants et commandité par le ministère de la santé³³ détaille et justifie les choix réalisés dans la méthode d'ajustement puis recommande que certaines améliorations soient apportées le plus rapidement possible. La méthode d'ajustement choisie est résolument rétrospective : l'objectif est de corriger ex post ce qui, dans les différences de coût observées entre assureurs, provient de différences de structure de risque et non de différences d'efficacité. Les éléments de coût qui proviennent du fait que la structure de risque est différente sont corrigés et on remplace la valeur propre observée pour l'assureur par la valeur moyenne dans la population. Les autres éléments ne donnent pas lieu à correction. Pour ce faire, on

³³ *Report of the Advisory Group on the Risk Equalisation Scheme*. Irish Ministry for Health and Children, 8 April 1998.

détermine en première approche les dépenses normatives brutes de chaque assureur. La somme de ces dépenses normatives brutes est normée de façon à ce que les dépenses normatives ajustées totales soient égales aux coûts totaux (le mécanisme est globalement neutre). Il s'articule concrètement autour de la correction des différences de structure par âge et par sexe des populations couvertes (R catégories indicées r) entre I assureurs indicés i .

Les remboursements effectués par l'assureur i avant ajustement peuvent être écrits de la façon suivante :

$$n_i \sum_r \left(P_{ir} \sum_d F_{ird} C_{ird} \right)$$

Avec :

n_i : nombre total de personnes couvertes par l'assureur ;

P_{ir} : proportion de personnes couvertes par l'assureur, appartenant à la catégories âge-sexe r ;

F_{ird} : fréquence d'un type de procédure médicale d dans la catégorie r pour l'assureur i ;

C_{ird} : coût de cette procédure pour l'assureur i .

On estime que P_{ir} reflète la sélection (la distorsion de la structure âge sexe d'un assureur à l'autre) et on le remplace par P_r , la part relative de la catégorie âge-sexe r dans la population totale couverte. A l'opposé, on postule que le coût unitaire C_{ird} relève de la responsabilité individuelle de l'assureur.

Deux scénarios sont ensuite explorés :

- dans le premier, on considère que la fréquence observée des actes médicaux relève de la responsabilité de l'assureur. Les calculs sont alors très simples : le coût moyen par catégorie de risque r de l'assureur n'est pas modifié par l'ajustement et seules les proportions de chaque cellule sont ajustées pour être égales à la proportion moyenne observée dans la population³⁴. Cette méthode revient en fait à assumer que les seuls facteurs qu'il importe de corriger sont l'âge et le sexe, puisqu'on estime que toute hétérogénéité de coût à l'intérieur d'une catégorie r relève exclusivement d'une gestion

³⁴ Les coûts par assureur ainsi calculés sont ajustés d'un facteur α tel que la somme pondérée par ce facteur des coûts théoriques des assureurs est égale à la somme observée des coûts.

plus ou moins efficace, ce qui est jugé par le groupe d'experts comme erroné et potentiellement dangereux ;

- dans le second, on considère que la fréquence reflète des différences de structure de risque et doit donc être remplacée par F_{rd} , fréquence moyenne observée de la procédure dans la population. La validité de cette approche dépend de la nature des "procédures médicales" considérées. Rappelons que l'assurance couvre essentiellement des dépenses hospitalières et des soins de spécialistes en milieu hospitalier. Si on regroupe les dépenses par groupe homogène de malade (*GHM* ou *DRG*), on peut alors postuler que la fréquence observée du DRG dans une catégorie r par un assureur i dépend largement de l'état de santé de la population couverte et que le coût unitaire du DRG est lié à la capacité de gestion de l'assureur. On peut donc remplacer la fréquence du DRG observée par l'assureur par la fréquence moyenne observée dans la population³⁵. Cette approche est recommandée par le groupe d'experts auteurs du rapport d'évaluation du schéma d'ajustement.

L'ajustement mis en œuvre en 1996 constitue une approximation dans laquelle on remplace les DRG par des durées de séjour. Le rapport souligne qu'elle génère des incitations perverses : si un assureur parvient à réduire la durée de séjour pour une pathologie (et donc à être plus efficace), compte tenu du mécanisme d'ajustement, l'opération entraînera une augmentation nette des coûts.

c) Bilan provisoire

Les experts indépendants qui ont mis au point le mécanisme d'ajustement, le jugent imparfait (parce qu'il se heurte à de nombreuses contraintes techniques), mais ils adoptent une position très pragmatique.

Selon eux, dans la mesure où le *VHI* est, pour des raisons historiques, en position dominante sur le marché, les nouveaux entrants, qui se tourneront en priorité vers des catégories relativement peu risquées (entreprises, jeunes...), obtiendront un avantage comparatif qui leur permettra de pénétrer sur le marché. Les défauts du mécanisme d'ajustement serviront donc la concurrence à court terme.

Les experts recommandent néanmoins que l'ajustement sur la base des DRG soit mis en place le plus rapidement possible. On peut enfin remarquer que le rapport préconise une

³⁵ On norme ensuite les dépenses pour que la somme pondérée des coûts théoriques soit égale à la somme observée.

certaine forme de standardisation des contrats. Plus précisément, il recommande que le *community rating* et l'ajustement des risques ne portent que sur un panier de soins jugés essentiels, au motif que l'Etat doit se contenter de garantir l'accès à une couverture de ces soins.

2.1.3 La Suisse

En Suisse, l'assurance maladie était, jusqu'à une période récente, volontaire, servie sur une base non lucrative et très réglementée. Son fonctionnement n'était pas satisfaisant et la loi LAMal, en introduisant l'assurance universelle, a tenté de limiter les effets de sélection, avec des résultats mitigés.

a) Le cadre général : la loi LAMal

Le système Suisse a été profondément réformé en 1994 car l'ancien système générait une inéquité et des problèmes d'accès à l'assurance croissants (OCDE, 1994, pour un diagnostic détaillé). Dans le même temps, les méthodes de tarification limitaient les incitations à l'efficacité : les tarifs étaient calculés en fonction de l'âge à l'entrée en assurance et, passé un certain âge, les assurés étaient de facto captifs de leur assurance.

Rappelons les grandes lignes de la loi LAMal (18 mars 1994) :

- tous les assureurs doivent offrir un contrat qui couvre obligatoirement un panier de services défini précisément par la loi. Une franchise fixée par le Conseil fédéral et une participation de 10% aux frais en deçà, dans une limite annuelle légale, sont prévues au contrat. L'assureur peut proposer une gamme de franchises plus élevées, déterminée légalement, moyennant diminution des primes selon des conditions précises, mais il ne peut offrir plus ou moins de services que ce qui est prévu par la loi ;
- l'assureur ne peut refuser de couvrir ou de renouveler le contrat des personnes qui en font la demande ;
- tous les Suisses doivent s'assurer.

Une des premières conséquences de la réforme a donc été d'introduire une assurance obligatoire et universelle. Le principe du choix de l'assureur et de la concurrence entre caisses qui caractérisaient le système avant la réforme ont été maintenus. Les contraintes qui pèsent sur les assureurs (acceptation obligatoire en assurance) et les assurés

(interdiction de sortir du système si on présente un faible niveau de risque) sont les premiers éléments destinés à garantir une mutualisation large des risques et à éviter la sélection explicite ou l'auto-sélection. En outre, les contrats sont très largement standardisés. Tout assureur qui offre des contrats dits particuliers³⁶, doit aussi offrir le contrat de base et ne peut donc espérer bénéficier d'une sélection favorable massive que s'il est le seul à offrir ce type de contrat. En tout état de cause, il devra tout de même couvrir les personnes qui présenteraient de plus hauts niveaux de risque et qui choisiraient le contrat standard.

Le système de financement renforce ces mécanismes de solidarité entre malades et bien portants et de temporisation de la sélection. Les assureurs sont contraints d'utiliser une forme stricte de *community rating*. Les primes peuvent varier par canton selon le lieu de résidence de l'assuré "s'il est prouvé que les coûts diffèrent selon les cantons et les régions" (LAMal, Titre I Chapitre 5, section 2, Art 61). Sous les mêmes conditions, il peut y avoir jusqu'à trois niveaux de primes par canton. Les enfants (moins de 18 ans) et les étudiants de moins de 25 ans peuvent bénéficier d'une réduction de prime. En dehors de ces conditions, les primes sont uniformes et doivent être approuvées par le Conseil fédéral. Le mode de tarification cherche donc à imposer une solidarité forte entre hauts et bas risques. Dans le même temps, les cantons doivent subventionner les primes des personnes défavorisées.

b) L'ajustement des risques : un essai manqué

Enfin, un mécanisme d'ajustement des risques a été mis en place sans lequel la sélection, du fait des contraintes tarifaires légales strictes, pourrait être assez forte, même si les caisses ou les assureurs qui ont le droit de servir l'assurance légale ne peuvent réaliser un profit sur cette activité. Le mécanisme d'ajustement n'est imposé aux assureurs que pour les dix années qui suivent l'entrée en vigueur de la loi (1996), pour des raisons qui ne sont pas explicitées.

³⁶ Nous avons signalé le contrat à franchise majorée. Les assureurs peuvent aussi offrir des contrats à bonus malus (très réglementés et dans le détail desquels nous n'entrerons pas) et peuvent constituer des HMO (moyennant là encore diminution des primes pour les assurés).

L'ajustement prévu dans l'Ordonnance de compensation des risques dans l'assurance maladie du 12 avril 1995 fonctionne selon les principes suivants :

1. les assurés (de plus de 18 ans) sont répartis en groupes de risque (par tranches d'âge de 5 ans et par sexe) ;
2. on détermine, par canton et par groupe de risque le coût moyen global, et les coûts par groupe de risque ;
3. si, pour un groupe de risque, le coût moyen de prise en charge par l'assureur est inférieur à la moyenne du canton, il paie par assuré une redevance égale à la différence à la moyenne ; inversement, il reçoit une redevance égale à la différence à la moyenne s'il a des coûts plus élevés que celle-ci.
4. la compensation est valable pour l'année courante, elle est calculée sur la base des coûts de cette année là. Une compensation provisoire est fixée en cours d'année (sur la base des informations de l'année précédente) et les règlements définitifs interviennent l'année suivante.

Ce mécanisme d'ajustement, dont la simplicité peut paraître séduisante, semble critiquable à plusieurs égards :

- en premier lieu, il ne prend en compte que l'âge et le sexe, facteurs de risque qui n'expliquent qu'une part infime de la variance explicable (cf. § 1.1.2).
- ensuite, il risque de gommer les incitations à la concurrence puisque l'objectif implicite de l'assureur est d'avoir le même coût moyen par canton que ses concurrents. Si un assureur a la même structure de risque mais est plus efficace qu'un autre, dont le coût est supérieur au coût moyen observé, il doit lui reverser une compensation. A l'inverse, si un assureur n'est pas efficace, il sera compensé. Il n'a donc qu'un intérêt réduit à maîtriser les coûts.
- enfin, comme le soulignent Beck et Zweifel (1995), la limite de 10 ans prévue pour l'ajustement au risque n'a guère de sens, car tant que le mode de tarification légal est maintenu, le risque de sélection est très élevé.

Le fonctionnement de l'ajustement au risque, malgré ces limites, ne semble pas avoir été modifié depuis l'ordonnance de 1995 (en tout cas, aucun texte législatif n'en atteste).

Le bilan actuel de la réforme est mitigé. Quelques créations de HMO ont été abondamment commentées dans la presse. Si les résultats de ces structures en termes de maîtrise des coûts sont plutôt bons, il n'est pas toujours facile de faire la part de la sélection dans les résultats obtenus (Perneger *et al.*, 1995). Un autre effet de la réforme a été d'accélérer la concentration des caisses et on estime, à l'horizon 2005, que trois caisses domineront le marché (Mosca, 1997). Mais la réforme a été mal tolérée par le public. Tout d'abord, les primes ont augmenté très fortement en 1996 (5,4% en moyenne entre 1995 et 1996). Cette augmentation, qui s'explique en partie par l'augmentation du panier de biens et services couverts, était pourtant prévisible et elle correspond à un transfert de charges : les paiements directs des ménages aux prestataires de soins ont diminué. Les observateurs estiment cependant que les assureurs ont largement dépassé les augmentations qui auraient été strictement nécessaires. Le passage au nouveau système a en outre été très douloureux, car le mode de tarification a induit des augmentations individuelles de prime allant jusqu'à 60% (Mosca, 1997). Dans le même temps, de nombreux cantons n'ont pas su mettre sur pied un système de subventionnement qui fasse coïncider l'augmentation générale des primes avec leur réduction pour les personnes à revenu modeste. D'autres, qui ont mobilisé l'intégralité des subventions, n'ont pas réussi à satisfaire l'objectif social de la Loi car les ressources étaient insuffisantes (Dreifuss, 1996). Parallèlement, des réformes portant sur l'offre ont été retardées ou se sont avérées incohérentes, ce qui a renforcé le mécontentement des assurés. Enfin, les assureurs se sont lancés dans une concurrence effrénée selon des méthodes discutables : ainsi, des études (citées par Mosca) ont montré que, dans les cantons où la concurrence était forte, les assureurs sous-estimaient les primes, et prélevaient les ressources complémentaires en majorant les primes sur d'autres régions. Comme la comptabilité des caisses et les informations statistiques disponibles sont très frustrées, il est très difficile de connaître les coûts réels générés par les assurés et donc la pertinence du montant de leurs cotisations. Le système ne semble toujours pas stabilisé. Ainsi, la seconde caisse de la Confédération a décidé de mettre fin en septembre 1998 à son activité d'assurance au titre de la LAMal dans 8 cantons qui comportaient les proportions les plus fortes de personnes âgées. Plus de cent mille personnes se sont donc trouvées dans l'obligation de chercher une autre assurance.

2.1.4 Principales leçons sur les systèmes de tradition concurrentielle

Les exemples issus de ces trois pays permettent de tirer une première série de conclusions.

Tout d'abord, les exemples américains et irlandais, détaillés à dessein, soulignent que la mise en œuvre d'un mécanisme d'ajustement au risque est un processus éminemment complexe. A l'inverse, un système dont la simplicité peut paraître séduisante, ne permet pas de mettre en œuvre les incitations souhaitées (Suisse). Il est vrai que l'expertise est récente et assez peu diffusée, mais il semble nécessaire de rappeler avec Luft que "*We must be aware of the need to adjust for risk differences and not to simply "assume" that an adequate mechanism exists or will be available in time for its effective use*"³⁷.

En second lieu, les exemples américains montrent que la reconnaissance de la nécessité de gérer la concurrence, pour en obtenir des bénéfices en termes d'efficacité productive, peut émerger spontanément. Pour qu'une telle organisation de la concurrence soit mise en œuvre, il faut cependant qu'un sponsor intervienne de façon volontariste et qu'il investisse dans le processus, y compris en termes financiers. Il peut alors parvenir à élaborer un système auquel assureurs et assurés adhèrent sur une base volontaire parce qu'ils y trouvent un intérêt.

Enfin, les systèmes américains et suisse, qui sont plutôt de type libéral, promeuvent implicitement une forme d'équité entre les hauts et les bas risques, objectif qui est, par contre, affiché en Irlande. Les primes payées par les assurés ne tiennent compte que d'un nombre limité de facteurs de risques et l'ajustement a lieu sur d'autres. Dans la mesure où l'information est disponible (mis à part les problèmes de délais de remontée de l'information), on pourrait envisager que le sponsor impose aux assureurs de les utiliser pour offrir des primes mieux ajustées aux risques individuels. Le résultat en terme d'efficacité serait strictement identique, mais l'équité du système y serait moindre. On peut rappeler à cette égard que la péréquation des risques constitue une forme de protection collective contre l'éventualité de devenir un mauvais risque et présente donc aussi un avantage à cet égard.

³⁷ Nous devons avoir conscience de la nécessité d'ajuster les risques et pas seulement faire l'hypothèse qu'un mécanisme adéquat existe ou qu'il sera disponible pour une utilisation effective lorsque le besoin se fera sentir.

Pour finir, il semble nécessaire de rappeler que les systèmes existants restent rares et que leur mise en œuvre est récente. Ils n'ont donc jamais été évalués en termes d'analyse coût-bénéfice. De nombreux employeurs aux Etats-Unis se contentent à l'heure actuelle d'utiliser des outils plus rudimentaires pour limiter la sélection (*guaranteed issue, guaranteed renewal, ..*). La preuve reste encore à faire de l'efficacité plus grande d'un système dans lequel on a introduit un ajustement au risque.

2.2 Introduction de la concurrence et délégation de gestion

Dans la seconde série d'exemples que nous étudions, les entités qui mettent en œuvre l'assurance maladie ne sont pas responsables au premier chef de la détermination du montant qui leur permet de faire face à leurs engagements ou de la collecte des fonds. Une entité intermédiaire, publique, remplit cette fonction. Dans ce type de système, la participation au financement de l'assurance est généralement explicitement déconnectée du risque individuel. Des contributions proportionnelles au revenu (impôts ou cotisations) sont prélevées sur les personnes couvertes ou sur l'ensemble de la population. Ceci permet, en théorie, de régler simultanément les problèmes d'accès à l'assurance et d'équité (entre personnes riches et pauvres et entre personnes malades et en bonne santé).

Les fonds sont alloués aux entités en concurrence sous la forme d'une capitation ajustée au risque. Ce refinancement se veut actuariel. Les entités auxquelles on confie la gestion de l'assurance sont alors responsabilisées sur le plan financier et s'approprient tout ou partie du "profit" de cette activité. Le premier objectif dans ce type de système est d'améliorer l'efficacité productive.

Parmi les pays étudiés, certains partent d'une situation antérieure dans laquelle la concurrence était absente (Pays-Bas, Allemagne) ou dans laquelle les organismes n'étaient pas responsabilisés sur d'éventuels déficits (Belgique et dans une certaine mesure, Israël). Le cas de Medicare est un peu particulier. Dans la situation antérieure, les dépenses des personnes couvertes étaient remboursées directement par la HCFA qui fonctionnait comme une assurance traditionnelle. Depuis quelques années, des plans (HMO) peuvent intervenir comme intermédiaires entre les assurés et l'administration. Ils reçoivent une capitation ajustée au risque et sont chargés de gérer le risque des assurés. L'objectif recherché par l'administration reste, là encore, l'amélioration du processus

d'administration des soins, l'augmentation de l'efficacité productive et à terme, la maîtrise des coûts.

2.2.1 Les Pays-Bas : archétype de l'introduction de la concurrence

La mise en concurrence des caisses publiques aux Pays-Bas est l'exemple le plus connu d'introduction d'une forme de responsabilité financière³⁸.

a) Le contexte et les principes de la réforme

L'assurance obligatoire des soins courants est gérée par une trentaine de caisses d'assurance maladie, qui sont des organismes à but non lucratif. Elles couvrent les salariés dont le revenu n'excède pas un certain seuil et offrent toutes les mêmes prestations, définies par l'Etat. Celui-ci prélève sur la population couverte obligatoirement par ce contrat une cotisation proportionnelle au revenu.

Avant la réforme, les caisses bénéficiaient de monopoles régionaux pour la couverture des personnes relevant de l'assurance maladie obligatoire. Elles recevaient du ZiekenfondsRaad (ZFR) un remboursement intégral des dépenses que leurs affiliés engendraient. En cas de déficit global, les taux de cotisation étaient augmentés ou l'Etat complétait le financement. Les caisses publiques étaient avant tout des organismes de liquidation.

La réforme a consisté à autoriser les bénéficiaires du régime d'assurance publique à choisir leur caisse d'affiliation (mise en concurrence sur une base nationale) et à en changer une fois par an (*open enrollement*), tout en maintenant l'obligation faite aux caisses de prendre en charge toute personne en faisant la demande (*guaranteed issue et renewal*). Parallèlement, leur mode de financement a été fondamentalement bouleversé : les caisses sont désormais financées de façon prospective en fonction du profil de risque de leurs assurés (et non plus remboursées systématiquement de manière rétrospective). Elles sont responsabilisées financièrement et donc en principe incitées à minimiser les coûts de gestion et à gérer efficacement les prestations des assurés³⁹. Le coût de l'assurance pour un individu reste, en revanche, complètement dissocié du risque qu'il présente.

³⁸ Une partie de cette analyse a été précédemment développée dans Couffinhal et Rochaix (1998).

³⁹ La stratégie utilisée par l'Etat néerlandais pour mettre en concurrence les caisses publiques d'assurance maladie relève de cette logique qui était sous-jacente aux propositions de DEKKER (Couffinhal et Rochaix, 1998).

Plus précisément, le financement des caisses est assuré par deux sources :

- une capitation ajustée au risque, versée par le ZFR qui centralise les cotisations ;
- une prime versée directement par chaque assuré à la caisse publique auprès de laquelle il choisit de s'inscrire. La contribution des primes collectées directement par les caisses au financement global de leurs dépenses n'est pas très élevée (moins de 6%). Le montant de la prime est obligatoirement identique pour tous les assurés d'une même caisse mais il est librement déterminé par celle-ci. Elle peut la diminuer si elle réalise un bénéfice et elle doit l'augmenter dans le cas contraire. Cette prime reflète donc théoriquement les différences d'efficacité productive d'une caisse à l'autre. Les assurés sont supposés se fonder sur ce critère pour arbitrer entre les caisses publiques. Comme il s'agit d'un financement marginal, c'est avant tout le niveau relatif des primes entre différentes caisses qui doit ici jouer un rôle incitatif. En 1997, l'écart annuel constaté entre la plus haute et la plus basse des primes est de 300 francs. Il semble augmenter au cours du temps : il était de 30 francs en 1995 et de 90 francs en 1996.

b) La détermination du budget des caisses

En l'année n-1, le gouvernement retient un niveau global de dépenses des caisses (hors primes nominales) pour l'année à venir ainsi que sa répartition entre grandes catégories de soins. Le ZFR calcule un budget par caisse sur la base des informations transmises par celles-ci sur les effectifs prévus par catégorie de risque. Ce budget est composé des capitations individuelles modulées en fonction de la classe de risque. En fin d'année, les caisses transmettent les données concernant les risques effectifs et les dépenses engagées. Le budget est recalculé pour tenir compte des effectifs réels et des ajustements sont effectués pour déterminer la part du décalage entre ce budget et les engagements financiers de la caisse qui demeurera sous sa responsabilité.

Calcul de la capitation ajustée au risque

Pour le calcul de la capitation pondérée au risque, le ZFR dispose des coûts moyens par catégorie de risque calculés sur l'ensemble des caisses pour les années précédentes. Connaissant la représentation dans la population globale des facteurs de risque, il alloue des poids relatifs à chaque catégorie de risque. Ces poids prennent en compte la différence de coût observée entre les catégories de risque et leur représentation dans la population. Ils sont ensuite appliqués au budget de la catégorie de soins, en fonction des

caractéristiques de la population effectivement couverte par chaque caisse, afin de déterminer la part qu'elle recevra.

Dans un premier temps (1993 et 1994), la capitation versée aux caisses se base sur l'âge et le sexe (38 classes), critères qui apparaissent rapidement trop grossiers. Depuis 1995, la capitation prend en compte un critère d'incapacité (de travail) et un critère régional, qui reflète probablement à la fois des effets d'offre (le coût des soins) et de demande (le profil de risque des populations couvertes). Ce dernier critère prend plus précisément en compte le degré d'urbanisation des lieux de résidence des personnes couvertes (cinq catégories). Les variables utilisées pour calculer les capitations ajustées au risque sont donc au nombre de quatre (âge, sexe, région et incapacité). Ces facteurs d'ajustement présentent l'avantage d'être facilement observables et contrôlables. Il est peu vraisemblable qu'ils changeront dans les années à venir. En effet, le modèle d'ajustement au risque, si imparfait soit-il, est suffisamment compliqué d'après le ZFR pour que les caisses ne puissent déterminer objectivement les risques pour lesquels elles seraient éventuellement sur-dotées et qu'elles seraient alors tentées d'attirer. En outre, le ZFR estime que l'imperfection de l'ajustement au risque constitue en soi une incitation pour les caisses à contrôler les coûts. Ces affirmations se fondent aussi implicitement sur l'idée que les caisses ne sont pas habituées à évoluer dans un univers concurrentiel, à évaluer le profil assurantiel de leurs assurés et sont peu capables de développer rapidement des techniques de sélection suffisamment sophistiquées pour n'être pas observables et donc sanctionnables.

La capitation ajustée au risque porte sur l'ensemble des soins de ville. Depuis 1996, les soins hospitaliers (qui représentent 55% des dépenses) sont aussi partiellement inclus dans ce système de capitation. Le budget était jusqu'alors calculé sur une base historique. Il est désormais scindé en deux parties représentant respectivement les coûts fixes (63%), qui restent estimés sur une base historique et les coûts variables, calculés selon quatre indicateurs de volume et soumis pour leur part au modèle de capitation ajustée au risque.

Ajustements ex post

Ce système de financement sur la base d'une capitation ajustée au risque est imparfait et s'accompagne d'une mutualisation obligatoire des hauts risques.

Le système de mutualisation mis en œuvre aux Pays-Bas en 1997 s'inspire du modèle théorique que nous avons évoqué dans la section précédente. Il n'opère cependant pas

une mutualisation des coûts engendrés par les personnes que l'assureur choisit ex ante mais des coûts constatés. Il oblige en fait l'ensemble des caisses à supporter collectivement le surcoût engendré par des personnes dont les dépenses individuelles dépasseraient 4500 Florins par an. Ce mécanisme permet une mutualisation de 90% du surcoût et s'opère dans la limite des budgets déterminés par l'Etat pour l'ensemble des caisses. Il ne génère donc pas de déresponsabilisation collective et constitue simplement une nouvelle réallocation du budget global.

Le tableau ci-dessous présente une simulation de ce mécanisme dont l'objet n'est pas tant le réalisme que l'illustration du mécanisme mis en place en 1997 aux Pays-Bas.

Tableau 1 : Mécanisme de mutualisation des hauts risques aux Pays-Bas

	Caisse A	Caisse B	Total
Coûts totaux observés	100 000	50 000	150 000
Nombre de personnes générant un coût supérieur au seuil de 4 500 NLG	5	0	
Coût de chacune de ces personnes	5 000		
Montant mutualisé	$5 \times (5\,000 - 4\,500) \times 90\% = 2\,250$		
Budget initial normatif de la caisse (après ajustement des risques)	95 000	45 000	140 000
Taux de participation au financement du pool	$2\,250 / 140\,000 = 1,61\%$		
Paiement de chaque fond au pool (1,61% de son budget initial)	- 1 527	- 723	2 250
"Subvention" au fond A	2 250		
Nouveau budget des caisses	95 723	44 227	140 000

Responsabilité finale des caisses

Par ailleurs, la responsabilité financière individuelle des caisses publiques est tempérée par une réglementation qui opère un partage des déficits éventuels entre caisses. Ceci permet de maintenir entre elles de façon transitoire une solidarité et se justifie d'autant que l'ajustement au risque de la capitation est imparfait. Ainsi, en 1997 et 1998, 20% des déficits individuels ont été partagés entre les caisses publiques.

Enfin, les déficits effectivement attribués aux caisses font l'objet d'allègements supplémentaires pour tenir compte du fait que leur marge de manoeuvre pour peser sur

les coûts n'est pas encore suffisante, l'Etat réglementant encore largement l'offre de soins. Cette reprise de déficit par le ZFR se justifie aussi par la volonté de ne pas confronter brutalement ces organismes à leur responsabilité financière, alors qu'ils assumaient jusqu'à un passé récent le simple rôle de liquidateur de prestations. On constate toutefois une augmentation rapide de la part du déficit final assumée par les caisses (5% en 1991 65% en 1998). Cette prise en charge du déficit se traduit par l'augmentation des primes nominales supportées par ses assurés pour l'année suivante.

c) Bilan provisoire des effets de la mise en concurrence des caisses publiques

Quand bien même le recul manque pour juger de l'efficacité d'ensemble de la mise en concurrence des caisses publiques, apparaissent déjà quelques difficultés.

Si l'on constate une amélioration de la qualité du service rendu, par la mise en place (par exemple) de comités permettant d'enregistrer les doléances des usagers (*complaints committees*), on ne peut manquer de noter la pérennité des comportements d'écrémage, malgré la présence d'une capitation ajustée au risque. Alors même que la sélection explicite des individus est interdite, la caisse peut mettre en œuvre des stratégies destinées à attirer les bons risques ou à dissuader les mauvais risques de s'assurer auprès d'elle (des comportements de démarchage sélectif ont été observés). Le pan le plus important des bénéfices attendus de la concurrence entre caisses, à savoir leur capacité à négocier avec les prestataires de soins, demeure pour l'instant embryonnaire. Les caisses sont certes obligées de conclure des conventions avec les prestataires, qui portent sur les prix, les quantités et la qualité, mais la marge de manoeuvre demeure réduite du fait de l'existence de budgets globaux et de la structure peu concurrentielle de l'offre de soins. L'évolution du droit néerlandais de la concurrence (qui a pour origine la mise en conformité du droit néerlandais avec le droit européen) devrait permettre dans quelque temps aux caisses de renforcer leur pouvoir de négociation et de limiter la collusion entre prestataires⁴⁰.

Dans l'immédiat, le conventionnement des prestataires leur permet de collecter de l'information qui constituera un bon moyen de procéder à une surveillance de la qualité et du contenu des services (*utilization review*). De même, ces conventions, en rapprochant les caisses et les prestataires, devraient encourager la coopération et l'introduction

⁴⁰ van de Ven, entretien, Avril 1998.

(rendue particulièrement nécessaire dans un environnement de plus en plus compétitif pour les caisses) de techniques de *managed care* ou de *disease management* pour certaines catégories d'assurés. On peut aussi s'attendre à ce que les caisses soient amenées à mettre en place des mesures de réduction du risque moral côté usagers.

L'hypothèse centrale à la mise en œuvre de la concurrence est que les assurés effectuent des arbitrages rationnels entre caisses (en fonction de la prime), renforçant l'émulation entre elles sur le prix et la qualité des services offerts. On constate aujourd'hui que l'arbitrage demeure timide, en raison du différentiel encore peu significatif entre primes (l'incitation financière que représente la prime nominale n'est, semble-t-il, pas suffisante pour entamer des démarches afin de changer de caisse), mais surtout, en raison de l'inertie des comportements des assurés. Même si certaines associations de consommateurs commencent à publier des informations comparatives sur les prix et les services des caisses, la grande majorité des assurés ne met pas encore à profit cette possibilité de "voter avec les pieds". Enfin, il semble que les individus hésitent à exercer leur droit, par crainte de se voir imposer des conditions moins favorables que celles dont ils bénéficient dans leur caisse d'origine en ce qui concerne l'assurance volontaire, dite "du troisième étage", qui couvre les soins non pris en charge par le contrat public (chapitre 7).

2.2.2 La Belgique

La réforme du système belge mise en œuvre au milieu des années quatre-vingt-dix, s'inspire assez largement de celle des Pays-Bas.

a) le contexte général

La quasi-totalité de la population belge est couverte par une assurance maladie obligatoire⁴¹. Des organismes assureurs (OA) ou mutualités à but non lucratif gèrent le régime d'assurance des travailleurs indépendants et celui de la population générale. La Belgique pratique depuis toujours la pluralité des OA, sans restriction légale à la mobilité des assurés d'une caisse à l'autre. La segmentation de la population entre les six grandes catégories de mutualités, regroupées en unions nationales, se fait avant tout selon des

⁴¹ Principales références : "Responsabilité financière des assureurs : le point". *Interface* (1998), Rapport de la commission pour la responsabilité financière des organismes assureurs, Mars 1994, *Calcul de la responsabilisation financière des organismes assureurs pour l'exercice 1995*. Institut National d'Assurance maladie et d'Invalidité, Note CSS 97/8 et Hermesse et Beekmans (1998)

critères linguistiques, religieux et politiques. La concurrence entre mutualités, qui reste marginale, porte, sur des services annexes non directement liés à l'assurance maladie. Conjointement, les unions détiennent un monopole sur le marché de l'assurance.

Le financement de l'assurance maladie obligatoire est principalement assuré par des cotisations patronales et salariales ainsi que des subsides de l'Etat venant combler le manque à gagner pour les personnes exonérées (chômeurs, veuves...). Traditionnellement, les ressources imputées aux différents OA provenaient des cotisations de leurs propres affiliés. Les assiettes de cotisation étant différentes d'une caisse à l'autre, les mutualités ne recevaient pas une proportion des cotisations égale à la proportion de la population qu'elles couvraient⁴² (Nonneman et van Doorslaer, 1994).

b) La réforme du financement et le mécanisme de responsabilisation

La volonté de maîtriser les dépenses de santé a progressivement conduit les autorités à augmenter la responsabilité financière des caisses. Depuis 1990, le ministère de la santé détermine le budget global de financement des dépenses de santé par secteur de soins (hôpitaux, médicaments...). Depuis 1994, un taux maximal de croissance des dépenses est imposé. La loi du 15 février 1993, a mis en œuvre une responsabilisation progressive des organismes assureurs selon des modalités qui se rapprochent de celles utilisées aux Pays-Bas.

Le mécanisme général

Le budget de l'assurance mentionné ci-dessus est réparti entre les mutuelles selon deux clés :

- une première répartition du budget est calculée en fonction d'une capitation ajustée au risque (calcul normatif) ;
- une seconde répartition est calculée en fonction de la part de chaque OA dans les dépenses de soins totales observées pour l'exercice concerné (calcul sur les prestations de l'année).

⁴² Traditionnellement, les mutualités chrétiennes (44.7% de la population en 1990) recevaient une part des cotisations relativement élevée (48.2%) et les mutuelles socialistes (28.9%), une part inférieure (25.3%). Les transferts implicites nécessaires à l'équilibre global du système entre les Flandres plus riche et la Wallonie francophone plus pauvre sont stigmatisés par certains hommes politiques qui veulent que la gestion de l'assurance maladie soit déléguée aux régions, contre la volonté de ceux qui estiment que le système de couverture social est fédérateur et doit relever de la responsabilité fédérale.

Le budget théorique de chaque OA est une somme pondérée des budgets calculés selon les deux méthodes. De 1995 à 1997, 10% du budget total est versé selon le calcul normatif. Ce taux augmente progressivement et devrait atteindre 30% en l'an 2000⁴³.

Lors de la clôture des comptes⁴⁴, cet objectif est comparé aux dépenses comptabilisées. Chaque OA est responsable à raison de 15% de la différence entre les deux montants (bonus ou malus)⁴⁵. Il est en outre prévu que la responsabilité d'un OA soit limitée à 2% de son objectif budgétaire global. Depuis 1995, afin de faire face à leur responsabilité financière, les OA sont tenus de prélever directement auprès des assurés une cotisation uniforme par OA (dont le montant minimum est fixé légalement). Chaque OA constitue un fond de réserve. S'il réalise un bonus, il est versé à la réserve, dans le cas contraire, l'organisme assureur peut tirer sur la réserve, voire utiliser ses fonds propres ou augmenter la cotisation uniforme.

Le calcul du budget normatif

La capitation sur la base de laquelle le budget normatif est calculé représente la dépense que devrait supporter l'assureur en fonction de sa structure de risque et des caractéristiques de sa population assurée. Elle est calculée séparément pour les deux régimes d'assurance (régime général et des travailleurs indépendants) et pour une série d'états sociaux. Le décret royal du 10 juin 1996 donne la liste des états sociaux pour lesquels on calcule une dépense normative. Pour le régime des salariés, on distingue principalement : les titulaires indemnifiables primaires (la majorité de la population), les pensionnés, les invalides ou les handicapés, les veufs, veuves et orphelins, et les personnes non protégées. Les effectifs par OA de chaque état social sont déterminés sur la base des données administratives de l'année pour laquelle on détermine le budget normatif.

⁴³ Pour prendre un exemple, une mutualité dont les dépenses réelles en 1997 s'élèvent à 30% du budget global et dont le budget normatif s'élève à 35% du budget global se verra attribuer : $0,9\% \cdot 30\% + 0,1\% \cdot 35\% = 30,5\%$ de l'objectif budgétaire global. Par construction, le budget global prévu est intégralement alloué.

⁴⁴ En 1997, pour l'année 1995 à titre d'exemple.

⁴⁵ Ce taux était valable de 1995 à 1997. En 1998 et 1999, il est de 20% et de 25% à partir de 2000.

Les dépenses normatives d'un assureur i , pour un état social r , C_{ir} sont égales à :

$$C_{ir} = n_r \left(\frac{C_r}{n_r} + COR_{ir} \right)$$

avec :

n_{ir} : nombre de bénéficiaires couverts par l'assureur i qui sont dans l'état s ;

n_s : nombre total de bénéficiaires dans l'état s ;

C_s : dépenses réelles globales de l'ensemble des OA pour l'état social s ;

COR_{ir} est un terme de correction pour l'état s et l'assureur i . Il reflète, par état social et pour chaque union de mutualités, l'ajustement en fonction d'une série de facteurs de risque (sexe, âge, situation socio-professionnelle, revenus, composition du ménage, taux d'urbanisation, taux de chômage et de mortalité)⁴⁶. L'estimation des paramètres de correction, sur la base de modèles économétriques, a été confiée à deux équipes universitaires pour l'année 1993. L'arrêté royal de 1996 prévoit que ces paramètres par état social seront ajustés chaque année pour tenir compte de l'augmentation des dépenses et de façon à ce que les termes de correction définitifs de chaque OA multipliés par l'effectif qui lui correspond donnent une somme égale à zéro. L'ajustement prend donc en compte la répartition des facteurs de risque à une date antérieure à celle pour laquelle la clé de répartition normative est calculée. L'arrêté royal prévoit cependant que les facteurs d'ajustement seront ajustés dès que possible si les paramètres recalculés diffèrent de façon significative de ceux qui figurent dans l'arrêté.

c) Quelques éléments de résultat

En 1995, toutes les unions de mutuelles ont obtenu un bonus, alors qu'en 1996, toutes ont dépassé l'objectif budgétaire de 6% environ⁴⁷. Cette uniformité souligne l'importance de choix du montant du budget macro-économique : les facteurs qui agissent uniformément sur les dépenses des unions ont une influence non négligeable sur l'équilibre global du système. En 1998, on observe pour la première fois une différenciation des primes entre les unions de mutualités (l'union de mutualité chrétienne

⁴⁶ Cette liste a été obtenue par recoupement à partir de Hermesse et Beeckmans et l'article d'Interface précédemment cité.

⁴⁷ Le malus est calculé sur la base de 15% de 2% (plafond).

propose une prime légèrement inférieure). Toutefois, la valeur absolue de la prime n'étant pas très élevée, ses concurrents doutent que ceci entraînera des transferts massifs d'assurés.

Comme aux Pays-Bas, la principale limite de la réforme réside dans le fait que les marges de manoeuvre des mutualités pour négocier avec les prestataires sont très limitées. Ainsi l'article d'Interface souligne que : *"si le mécanisme de répartition de l'objectif budgétaire...[est défini] de manière précise et technique, il n'en va pas de même ..de la mise à disposition des outils nécessaires à cette responsabilisation individuelle... [Les conventions entre prestataires de soins et organismes assureurs] valent pour l'ensemble des mutualités et il n'est pas possible pour les prestataires de débattre des tarifs individuellement avec un OA, dans le cadre de l'assurance obligatoire"* (pp.12).

2.2.3 L'Allemagne

Nous avons souligné les similitudes entre le système belge et celui des Pays-Bas. L'objectif de court terme visé par l'introduction de la concurrence entre les caisses publiques allemandes dans les années quatre-vingt-dix était de décloisonner un système dans lequel, pour des raisons historiques, elles se trouvaient dans des situations très différentes au regard des risques qu'elles couvraient et des ressources qu'elles pouvaient prélever. A terme, l'objectif visé par la concurrence est pourtant une meilleure gestion des risques.

a) Le contexte de la réforme

En Allemagne, les caisses publiques, qui couvrent 90% de la population, sont traditionnellement organisées sur une base professionnelle. Les personnes qui ne peuvent être affiliées à ce titre, se regroupent par défaut dans des caisses locales. Le système est très morcelé, puisqu'environ mille caisses œuvrent sur le territoire.

Avant 1995, chaque caisse prélève des cotisations proportionnelles au revenu en déterminant le taux qui lui permet d'atteindre l'équilibre financier. Les caisses sont donc responsabilisées financièrement, mais elles sont en position de quasi-monopole. Par construction, la structure des risques couverts varie fortement d'une caisse à l'autre ainsi que la capacité contributive (le revenu moyen) des adhérents. Les caisses locales cumulent les handicaps : le revenu moyen des assurés est inférieur au revenu moyen des assurés des autres caisses, et elles couvrent des proportions importantes de personnes

âgées, d'invalides... La réunification a renforcé ces inégalités : les taux de cotisations sont systématiquement plus élevés dans les nouveaux Länder que dans les anciens (Boccognano, et al., 1998, pp.132).

La réforme Seehofer, initiée en 1992, cherche en premier lieu à harmoniser le système et à introduire entre les caisses une forme de concurrence qui les incitera à l'efficacité. Les éléments de la réforme sont les suivants :

- toutes les caisses sont tenues de couvrir les mêmes biens et services et d'offrir les mêmes taux de couverture ;
- le symbole de l'instauration de la concurrence est l'introduction d'une plus grande liberté de choix pour les assurés depuis 1997. Ainsi, toute personne a maintenant le droit d'adhérer à la caisse de son choix. Cette liberté est pourtant incomplète, puisque les caisses professionnelles conservent la possibilité de ne pas offrir de contrats aux personnes qu'elles ne couvraient pas traditionnellement, ce qui ne correspond à aucune logique économique, mais relève plutôt d'un lobbying efficace de la part des caisses d'entreprise couvrant une population jeune à revenus élevés ;
- afin que les prix reflètent l'efficacité des caisses et non leur structure démographique et économique plus ou moins favorable, il existe maintenant un mécanisme de compensation des risques qui se traduit par des transferts financiers entre caisses, selon des critères d'âge, de sexe et de revenu des assurés.

b) L'ajustement des risques

Ce mécanisme a été élaboré sous la pression des caisses locales (Allgemeine Ortskrankenkassen, AOK) qui étaient très défavorisées par la situation antérieure, avec le soutien du Parti Social Démocrate (SPD) et des ministres sociaux des Länder.

Le mécanisme est original car il ne cherche pas seulement à compenser les caisses pour leur différence de structures de risque mais aussi pour celle résultant des variations de capacités contributives de leurs membres affiliés.

Dans un premier temps (Files et Murray, 1995), on estime la capacité financière de chaque caisse. Pour ce faire, on détermine par personne *couverte* (assurés et ayants droit) un revenu soumis à cotisation⁴⁸. On multiplie ce montant par une estimation du

⁴⁸ Pour un même niveau de revenu disponible, les caisses dans lesquelles le nombre d'ayant droits est très élevé ont donc une capacité contributive estimée plus faible.

taux de cotisation moyen sur la population (dépenses maladie de l'ensemble des caisses/revenu national soumis à cotisation). On détermine ainsi le montant total que la caisse prélèverait si elle avait un taux de cotisation "moyen".

Ensuite, on calcule un coût moyen par personne selon trois facteurs de risque : l'âge, le sexe et l'incapacité. On calcule alors une structure de coût de la caisse en fonction du nombre de personnes qui répondent à ces critères.

Si la capacité contributive est supérieure aux coûts standardisés, la caisse doit contribuer au financement du pool et vice versa. Les sommes versées et collectées sont déterminées ex ante : la caisse reste responsable de son équilibre financier mais elle peut prendre en compte dans le calcul de son taux de cotisation les compensations qu'elle devra payer ou qu'elle recevra.

c) Bilan provisoire

La compensation a été mise en œuvre pour la première fois en 1994 sur la base d'estimations sur données d'enquête de 1993. Les AOK, relayées par les médecins, soutiennent toutefois que cette compensation est insuffisante pour permettre une concurrence juste, qui serait fondée sur une utilisation efficace des ressources et non sur la sélection des risques. En effet, l'état de santé des personnes défavorisées couvertes par les AOK est bien en dessous de la moyenne nationale. La compensation ne prend en compte le revenu que sous l'angle de la capacité contributive, et celui non de la corrélation entre bas revenus et dépenses hospitalières élevées. Les AOK qui, entre autres, couvrent des proportions beaucoup plus élevées de diabétiques, souhaiteraient que des indicateurs de maladie chronique soient pris en compte dans le calcul des dépenses anticipées. Toutefois, même si elle est imparfaite, la compensation a conduit à un rapprochement significatif des taux de cotisation : l'écart de taux moyen en 1997 dans la partie occidentale du pays entre les caisses d'entreprises et les caisses locales était de 1,1 point contre 2,3 points en 1993. La concentration des caisses a aussi été très rapide. Les effets les plus néfastes de la segmentation ont donc été réduits.

De manière générale, la mise en concurrence des caisses n'a pas eu d'effet sur les coûts médicaux ni sur l'efficacité du système, en grande partie parce que la faible mobilité des assurés amoindrit la menace de sanction contre d'éventuels manquements à l'efficacité. Les assurés ont parfaitement le droit de changer de caisse, mais ils ne le font pas

facilement : ainsi, les caisses locales, dont les taux de cotisation sont parmi les plus élevés, n'ont vu leurs effectifs baisser que de 2%. L'autorisation de changer de caisse a apparemment plutôt été utilisée par les assurés qui déménagent pour rester dans la même caisse. Il semble toutefois que les bons risques (populations jeunes et à revenus élevés) soient davantage mobiles que les autres.

Enfin, l'introduction de la concurrence a incité les caisses les plus exposées (les AOK) à s'impliquer -de façon limitée - dans les expérimentations de type *managed care*. Leurs concurrentes, dont la structure de risques est plus favorable, sont plus réticentes à proposer des formules que leur clientèle traditionnelle, relativement aisée, pourrait boudier : les modèles de type *gatekeeper* ont relativement mauvaise presse en Allemagne parce qu'ils symbolisent une médecine bon marché et de mauvaise qualité (Boccognano, et al., 1998).

2.2.4 Israël

Les système israélien avant 1995 présente quelques similitudes avec le système suisse (non universel et concurrentiel), même si les cotisations sont proportionnelles au revenu (comme en Allemagne). La réforme a introduit l'assurance universelle et a mis à plat le système de financement en centralisant les ressources prélevées. Les caisses en concurrence sont désormais théoriquement rémunérées en fonction des risques qu'elles couvrent.

a) La réforme de 1995

En Israël, l'assurance maladie est traditionnellement servie par des caisses à but non lucratif contrôlées pas des syndicats et les individus peuvent théoriquement choisir leur caisse d'affiliation, comme en Belgique. La caisse dominante, qui dépend du principal syndicat ouvrier (70% du marché) fonctionne en *HMO staff model*⁴⁹. Trois caisses mineures se partagent le reste du marché et passent des contrats avec les groupes de médecins. Les caisses et les syndicats prélèvent des cotisations proportionnelles au revenu sur les employeurs et employés. Les taux sont en principe uniformes pour toutes les caisses mais le processus de financement global manque singulièrement de transparence.

⁴⁹ Les médecins sont salariés.

La réforme de l'assurance maladie israélienne de 1995 vise à répondre à une série de problèmes mis en lumière par une commission parlementaire quelques années auparavant (Commission Netanyahu 1990). L'assurance est volontaire et 4% de la population n'y a pas accès. Le problème est beaucoup plus prononcé pour les citoyens arabes - 12% non assurés - et les personnes résidant dans certaines régions. Certaines catégories d'immigrants sont aussi touchées. Par ailleurs, de nombreuses personnes n'ont pas réellement la possibilité de changer de caisse, soit parce qu'elles savent que leur candidature sera rejetée par une autre caisse - l'écroulement se développe au fur et à mesure que les coûts augmentent -, soit parce que dans leur région, seule la caisse dominante est présente. Celle-ci étant obligée statutairement d'accepter toute personne syndiquée qui en fait la demande, voit au cours des années 80, sa structure de risque se dégrader fortement (12% de personnes de plus de 65 ans versus 6% pour les autres caisses, beaucoup plus de malades chroniques). Comme, en outre, elle couvre des personnes de condition modeste, elle reçoit au total (à âge égal) 30% de moins que les autres caisses. Ses déficits sont financés au coup par coup par l'Etat qui s'évertue dans le même temps à transférer une part croissante du financement vers le secteur privé. Le manque de transparence généralisé du système a des conséquences concrètes pour les assurés. Leurs droits ne sont pas clairement définis et même s'ils sont couverts, ils se voient de plus en plus souvent obligés de payer directement des interventions lourdes.

La réforme de 1995 a donc deux objectifs principaux (Chernichovsky et Chinitz, 1995) : d'une part clarifier les responsabilités de chacun (afin d'inciter les caisses à une meilleure gestion) et d'autre part garantir à tous les citoyens, indépendamment de leurs ressources, un accès à un panier standard de soins servi par la caisse de leur choix. Les principaux éléments du texte de loi sont les suivants :

- l'assurance devient universelle ;
- elle couvre un panier de biens et services clairement définis ;
- chacun peut choisir la caisse d'affiliation ;
- les caisses n'ont pas le droit de refuser la couverture à un individu, ou d'exiger un complément de prime pour financer l'assurance légale. Les caisses doivent donc se faire concurrence sur la qualité des réseaux proposés.

b) Le financement prospectif des caisses

- les cotisations existantes sont regroupées en une cotisation santé unique prélevée de façon centralisée et l'Etat s'engage à financer le complément si nécessaire ;
- les fonds sont reversés aux caisses sous la forme d'une capitation. Celle-ci est cependant très grossière, car elle ne dépend que de l'âge (9 tranches). La pondération de chaque tranche d'âge est calculée par le ministère sur la base d'informations externes aux caisses⁵⁰. Des capitations égales au coût moyen du traitement sont versées aux caisses pour les assurés qui souffrent de quatre pathologies (dont l'IRC et l'hémophilie). Ce versement représente 3.3% du total versé en 1996.

Un bilan récent de la réforme (Rosen, 1998) indique que la concurrence entre caisses a augmenté. Certaines régions ou populations ont désormais plus de choix et la qualité des services s'est globalement améliorée. Parallèlement, il semble que les caisses développent des stratégies d'écrémage plus actives. L'allocation de ressources est incontestablement plus juste qu'auparavant mais toutes les caisses doivent faire face à des déficits importants et croissants, ce qui crée des problèmes politiques entre le ministère de la Santé et les caisses d'une part, qui estiment que le secteur de la santé est sous financé et le Trésor, qui considère pour sa part que les caisses n'ont pas déployé assez d'efforts pour être efficaces. Fin 1997, le Parlement envisage d'autoriser les caisses à introduire des copaiements⁵¹ pour les services du panier de base et à prélever une prime uniforme par assuré, les personnes défavorisées pouvant être exonérées totalement ou partiellement. Les tenants de cette réformes voient là un moyen d'instiller une concurrence limitée par les prix et d'augmenter l'autofinancement des caisses. Les détracteurs soulignaient le côté régressif de la taxe uniforme et le fait que les copaiements seraient plus lourds pour les malades chroniques. Après de nombreux débats, le principe de ces réformes a été approuvé mi-1998, mais les mesures concrètes proposées par les caisses doivent être approuvées par les plans et aucune n'a jusqu'à maintenant été acceptée.

⁵⁰ Données d'enquête (pour l'ambulatoire), données hospitalières, et comptes de la santé (pour la répartition entre types de soins par tranche d'âge). Les consommations pharmaceutiques ne sont pas prises en compte.

⁵¹ Certains copaiements étaient déjà prévus par la Loi mais celle-ci stipulait qu'ils ne devaient pas être augmentés et étendus.

Par ailleurs, le parlement a commandité une étude visant à évaluer dans quelle mesure la capitation devrait être améliorée. Le rapport (Habib et Rosen, 1998), après consultation de nombreux experts et évaluation la faisabilité avec les caisses elles-mêmes, recommande plusieurs améliorations marginales (par exemple, réactualisation plus fréquente de la capitation, inclusion de la pharmacie dans le calcul des pondérations par classe d'âge...) dont nous ne présenterons pas le détail. Quant à l'introduction de nouvelles variables, les conclusions sont les suivantes :

1. en ce qui concerne l'état de santé : les indicateurs proposés sont les maladies chroniques (directement ou par le biais des hospitalisations qu'elles entraînent) et/ou l'incapacité. Le choix des indicateurs devrait dépendre de la fiabilité des informations disponibles et la faisabilité n'est pour l'instant pas garantie. Il est aussi proposé que l'Etat prenne en charge une partie (50%) des dépenses individuelles au delà d'un plafond annuel. Cette approche pose aussi un problème car les caisses ne sont pas capables à l'heure actuelle de calculer un coût par personne et par an. Le rapport estime que la recherche visant à inclure l'état de santé dans la capitation doit être prioritaire ;
2. puisque les dépenses liées à la maternité sont prises en charges en dehors de l'assurance de base, il n'est donc pas jugé nécessaire d'inclure le sexe dans la capitation ;
3. l'introduction d'une variable "zone périphérique" ne fait pas l'objet d'un consensus. Les uns estiment qu'en périphérie les niveaux de consommation sont plus faibles et les coûts d'investissement plus difficiles à rentabiliser et qu'on pourrait, en augmentant la dotation, tenter de corriger ces effets. Les autres répliquent que la loi a déjà amélioré la qualité de l'offre dans les régions plus isolées et qu'augmenter la capitation n'est pas l'incitation la plus efficace à l'investissement. Par ailleurs, ils privilégient une capitation plus élevée là où la consommation l'est (débat entre financement des coûts / financement des besoins). En tout état de cause, les avis divergent sur la façon de mesurer "l'éloignement". Au total, le rapport recommande d'attendre quelques années avant de considérer de nouveau l'inclusion d'une telle variable ;
4. une variable indicatrice de la situation par rapport à l'emploi pourrait être incluse pour rendre compte de différences d'état de santé liées au niveau social (en particulier tant que l'état de santé n'est pas pris en compte en tant que tel). Toutefois, à l'heure

actuelle, le délai nécessaire à la validation de l'information est encore trop long pour envisager l'inclusion de la variable.

La redéfinition de la capitation apparaît en Israël (contrairement à la Suisse) comme un objectif prioritaire qui permettrait de continuer à corriger les inégalités de dotations et qui pourrait limiter les stratégies d'écémage que la mise à plat du système et le renforcement de la responsabilité financière ont avivées. L'approche très pragmatique et progressive adoptée, qui s'inspire du modèle néerlandais, semble pouvoir amener à terme une amélioration du fonctionnement du système.

2.2.5 Medicare

Le dernier système que nous présentons est celui de Medicare. Depuis la fin des années quatre-vingt-dix, chaque assuré peut rester dans le système de paiement à l'acte ou se tourner vers un opérateur privé, de type HMO obligatoirement, qui lui proposera un réseau fermé de praticiens et une filière de soins organisée. En 1997, 14% de la population Medicare était couverte par un HMO. Medicare est administré par la *Health Care Financial Administration* (HCFA) qui, en ouvrant cette possibilité, souhaitait développer un mode de prise en charge et de gestion des soins plus efficace que celui du paiement à l'acte. La HCFA doit en contrepartie jouer le rôle du "sponsor" dévolu à l'employeur dans la concurrence gérée d'Enthoven : choisir les plans et décider du montant de la capitation qu'elle leur verse pour assurer et délivrer les soins aux personnes âgées. Avant 1997, une capitation était calculée par type de bénéficiaire (personne âgée, insuffisant rénal chronique ou invalide) et de soin (partie A et B) sur la base du coût historique moyen de prise en charge dans le système traditionnel de paiement à l'acte dans le comté, ajusté par tranche d'âge, selon le sexe et quelques variables de statut (selon que la personne réside dans une institution - maison de retraite, long séjour -, travaillant, bénéficiaire de Medicaid). 95% de ce "*Adjusted Average Per Capita Cost*" ou AAPCC était versé aux plans. Il est apparu au fil du temps que cette dotation était trop grossière :

- elle encourageait la sélection des risques par les assureurs ou leur donnait une rente trop forte du fait de l'auto-sélection dont ils bénéficiaient (cf. chapitre précédent).
- elle aboutissait à des situations paradoxales car les capitations entre comtés adjacents pouvaient différer jusqu'à 20% et pouvaient varier d'une année sur l'autre (en fonction

des coûts constatés), ce qui ne favorisait pas le développement harmonieux de l'offre par les HMO, particulièrement dans les régions rurales.

En 1997, plusieurs réformes ont été mises en place (Rapport au Congrès, Medicare Payment Policy, 1998). Tout d'abord, la capitation a été modifiée à compter de 1998 :

- une capitation minimum est garantie aux plans, ainsi qu'un taux d'augmentation minimal de 2% par an par comté ;
- la capitation n'est plus calculée désormais sur une base exclusivement locale : elle correspond, pour partie, à la moyenne nationale calculée sur les mêmes critères de risque et, pour partie, à la moyenne locale. En 1998, la proportion du coût local est de 90% et la pondération devrait atteindre 50% en 2003.

La modification du système de paiement doit être neutre pour la HCFA, qui doit déboursier les mêmes sommes que si l'ancien système avait été appliqué. Toutefois, les ajustements permettant de respecter cette contrainte ne sauraient conduire un comté en dessous du minimum garanti.

La loi de 1997 prévoit en outre que la capitation sera ajustée en fonction de l'état de santé à partir de 2000. Par ailleurs, la HCFA cherche à promouvoir une réforme qui contribuerait à renforcer la concurrence entre HMO et à limiter les bénéfices qu'ils tirent de l'écroulement ou de l'auto-sélection. Pour ce faire, la HCFA veut mettre en place une concurrence en deux étapes. En premier lieu, un mécanisme d'enchères permet de fixer le montant de la capitation. La HCFA demande aux opérateurs de fournir leurs prix de manière confidentielle et elle choisit ensuite le montant de la capitation à partir de l'information collectée. En second lieu, les opérateurs démarchent les assurés, en leur proposant obligatoirement le prix qu'ils ont révélé à la HCFA. Cependant, l'instauration d'une telle concurrence pose des problèmes pratiques de mise en place. Tout d'abord, il convient de signaler que les plans sont très réticents et tentent de faire renoncer la HCFA à son projet, ou bien d'en obtenir une version adoucie. Ils craignent en premier lieu qu'en les obligeant à révéler leurs véritables coûts, le mécanisme proposé ne les prive de leur profit (donc d'une forme de rente qu'ils prélèvent sur le système de soins). Ils avancent aussi l'argument selon lequel ce type de procédure à deux étapes augmente notablement les frais d'entrée sur le marché, ainsi que les risques commerciaux. Elle limite aussi la concurrence en introduisant une nouvelle barrière à l'entrée : seuls ceux qui sont passés par le quasi marché ont le droit de proposer leurs services aux assurés. Donc, une fois

l'enchère réalisée, on ne peut autoriser l'entrée de nouveaux plans jusqu'à la prochaine renégociation. Si on agrandit trop l'intervalle entre deux marchés internes, on interdit l'entrée de compétiteurs potentiellement innovants et on accroît la rente de ceux qui sont présents. Si on réduit trop cette période, on augmente notablement les coûts et les risques pour les opérateurs.

Au total, l'expérience de délégation de gestion de Medicare par la HCFA reste connue par les effets de sélection ou de segmentation spontanée qu'elle a suscités pour les personnes concernées. La HCFA cherche à corriger cet effet en limitant la rente dont bénéficient les HMOs qui reçoivent une capitation jugée trop généreuse et en l'ajustant suffisamment bien pour que des personnes qui présenteraient des niveaux de risque élevés puissent, elles aussi, bénéficier des avantages d'une gestion plus globale et plus efficace de leur prise en charge.

2.2.6 Principales leçons des systèmes où la concurrence est introduite

Les cinq systèmes, dont nous venons de détailler les réformes, avaient, à l'origine, des modes de fonctionnement assez différents, comme le rappelle le tableau ci-dessous.

Tableau 2 : Caractéristiques principales des systèmes avant les réformes

	Couverture	Structure du marché	Concurrence	Responsabilité financière
Pays-Bas	universelle*	Monopole	Non	Non
Belgique	universelle*	6 unions nationales	Oui	Non
Allemagne	universelle*	1000 caisses	Non	Oui
Israël	non universelle	4 caisses	Oui	Non
Medicare	universelle*	Monopole	Non	Non

* Pour la population du régime concerné (cf Chapitre 7)

En premier lieu, notons que les réformes ont indéniablement été porteuses d'une harmonisation, puisque, désormais, dans les cinq systèmes, la couverture est universelle, les organismes qui la gèrent sont en concurrence⁵² et sont responsabilisés financièrement de façon explicite et uniforme au sein de chaque système. En outre, à l'intérieur de chaque système, les paniers de biens et services couverts sont identiques.

Au delà de ce premier constat, un certain nombre de points méritent d'être soulignés.

1. S'il n'est pas possible d'estimer l'impact des réformes sur l'équité de tous les systèmes, on peut remarquer qu'aucune ne vise à réduire la solidarité au profit de la responsabilité individuelle ou à remettre en cause ce qui reste considéré comme un acquis social (accès à l'assurance et financement redistributif). Plus précisément :
 - on constate, tout d'abord, que le nombre de personnes couvertes ne diminue dans aucun pays malgré l'introduction de mécanismes concurrentiels. En Israël, il augmente même, mais comme, dans le même temps, le panier de biens et services couvert par l'assurance sociale diminue pour une partie de la population, il est difficile d'estimer l'impact global en terme de couverture. En revanche, désormais, dans ce pays, toute personne est tenue de s'assurer, ce qui signifie en particulier que les bas risques ou les hauts revenus n'ont plus la possibilité d'échapper à un financement solidaire de l'assurance sociale. Dans les autres systèmes, l'assurance demeure obligatoire et les entités en concurrence n'ont pas le droit de refuser de couvrir une personne qui en fait la demande, ce qui permet de garantir une couverture universelle ;
 - avant les réformes, les systèmes étaient déjà essentiellement financés par des cotisations (Pays-Bas, Israël, Allemagne, Belgique) ou des impôts (Medicare) proportionnels au revenu. Ce principe est aussi maintenu. En Allemagne, la concentration des caisses et le mécanisme de compensation - qui vise explicitement à compenser les différences de capacités contributives - ont plutôt amélioré l'équité du financement par la réduction des écarts entre les taux de cotisations. Les réformes ont aussi conduit à une uniformisation des taux de cotisation en Israël et en Belgique ;

⁵² Pour Medicare, cette concurrence se déroule entre les HMOs qui participent à la délégations de gestion. La HCFA n'est pas à proprement parler en concurrence avec les HMOs.

- les Pays-Bas et la Belgique ont ouvert aux organismes en concurrence la possibilité de prélever des primes uniformes sur tous les assurés pour équilibrer leurs budgets. Un tel mode de financement est, par construction, inéquitable. Dans l'esprit initial de la réforme aux Pays-Bas, il était prévu que cette prime représenterait 25% de la dépense, mesure qui a été abandonnée pour préserver l'équité du système. A l'heure actuelle, dans les deux pays, la prime reste très symbolique et constitue, en théorie, un signal de l'efficacité productive relative d'un organisme par rapport à un autre.
2. L'introduction d'une possibilité pour les assurés de choisir l'opérateur gestionnaire de leur couverture a eu un impact très différent selon la "culture" d'origine.

Aux Etats-Unis, où les assurés sont habitués à choisir des contrats et à changer d'assureur, ils semblent utiliser pleinement les possibilités qui leur sont offertes. Plusieurs exemples en témoignent. On constate ainsi que les assurés passent, en fonction de leur état de santé, d'un système à l'autre : quand ils sont en bonne santé, ils choisissent le système HMO et ils reviennent vers le système traditionnel si leur état de santé se dégrade (système dit des "portes de saloon", Morgan et al. 1997⁵³). Dans le même ordre d'idée, les HMOs proposant une période d'adhésion mensuelle, se sont trouvés confrontés au problème du "*hit and run*" : une proportion importante de la clientèle change tous les mois d'assureur, soit pour bénéficier d'une promotion ponctuelle, soit, plus vraisemblablement, pour contourner les plafonds de dépense sur les médicaments (l'assureur les remboursant dans les limites d'un plafond annuel). Quand l'assuré a atteint ce plafond, il se tourne vers un nouveau HMO⁵⁴.

A l'inverse, dans les autres pays, la réforme n'a semble-t-il pas incité les personnes à changer d'organisme gestionnaire. En Israël et en Belgique, une forme de concurrence existait en théorie déjà, mais le choix d'une caisse ou d'une mutualité était, et reste, plus l'expression d'un choix politique, d'une appartenance culturelle que d'un arbitrage économique. Aux Pays-Bas et en Allemagne, les assurés ont plus mis à profit la possibilité qui leur était offerte pour rester dans la même caisse, là où, avant la réforme, ils auraient été contraints de changer (à cause d'un déménagement ou d'un changement d'emploi). Les mécanismes d'incitation par les prix n'ont semble-t-il pas eu un impact très élevé, ce qui n'est pas étonnant en Belgique et aux Pays-Bas où les

⁵³ La question demeure entière de savoir s'ils sont incités par les HMO à passer de l'un à l'autre.

⁵⁴ Entretien avec Arlen Collins, consultant en Managed Care, 1998.

primes sont assez faibles et encore peu différenciées, mais qui est plus surprenant en Allemagne, où les taux de cotisation peuvent varier d'une caisse à l'autre. En tout état de cause, il est probablement trop tôt pour que les habitudes aient changé et il serait intéressant de suivre l'évolution des comportements dans les prochaines années.

3. Pour ce qui est du comportement des OGR en concurrence, il faut, là encore, distinguer les Etats-Unis des autres systèmes.

Les HMOs auxquels on délègue la gestion de Medicare sont familiers avec le fonctionnement d'un marché concurrentiel et, par définition, avec la gestion du risque. Le versement par la HCFA d'une capitation trop peu ajustée au risque leur a permis de réaliser des profits d'autant plus élevés qu'il leur suffisait d'intégrer ces nouveaux patients dans les réseaux déjà existants et qu'il ne leur a pas été nécessaire de réaliser de nouveaux investissements pour améliorer la gestion des risques.

Dans les autres pays, exception faite d'Israël où le fonctionnement en réseau est implanté dans les moeurs, mais où les problèmes auxquels le système est confronté sont de nature plus politique et dont le mode de gestion n'est pas à l'heure actuelle bien défini (Rosen, 1998), l'évolution semble très lente, pour deux grands types de raisons.

- Tout d'abord, aux Pays-Bas et en Belgique, mais aussi en Allemagne, les caisses doivent opérer un véritable changement culturel. Dans les deux premiers pays, elles n'avaient auparavant jamais été responsabilisées et leur rôle se limitait quasiment à la liquidation des prestations. En Allemagne, où la concurrence n'existait pas, la situation était comparable. Elles doivent donc progressivement constituer des systèmes d'information et se doter des outils de gestion du risque, développer une culture de service... Une telle évolution ne peut s'opérer que progressivement et l'on en n'observe pour l'instant que les prémices (accueil d'une plus grande convivialité...).
- Mais surtout, les conditions réglementaires qui pourraient permettre aux organismes de gérer le risque ne sont pas réunies. Dans l'ensemble de ces pays, les marges de manoeuvre des OGR sont restées très limitées et il semble que les Etats répugnent à abandonner des outils de maîtrise des dépenses éprouvés au bénéfice d'une gestion purement marchande. Ils doivent aussi faire face à la résistance des prestataires de soins. On peut cependant noter qu'aucune réforme n'a été stoppée et que certains pays

envisagent d'aller plus loin. Ainsi, en Allemagne⁵⁵, la Ministre Andrea Fisher souhaite confier d'ici 2003, la gestion des investissements hospitaliers aux caisses et non plus aux régions. Les caisses devraient aussi avoir le droit de passer des contrats avec des médecins ou des groupes de médecins.

Ceci dit, il semble encore trop tôt pour tirer un bilan définitif des réformes entreprises dans ces pays qui dépasse le constat simple d'une harmonisation des modalités de concurrence.

⁵⁵ "Réforme de la santé : les caisses Allemandes devraient sortir gagnantes" *Espace Social Européen*, 18-24 juin 1999, pp. 12.

CONCLUSION DU CHAPITRE ET DE LA PARTIE

Au terme de ce chapitre, il semble donc évident que les modes de gestion de la concurrence convergent, comme le tableau 3 le souligne (page suivante). Même si les réformes ou les évolutions constatées répondent à des objectifs prioritaires différents, on peut estimer que toutes prennent en compte le fait que *"the only time risk assessment and risk adjustment are not necessary is when there is a unitary plan enrolling the entire eligible population"*⁵⁶ (Luft, 1995, pp. 23).

Certaines différences fondamentales entre les systèmes traditionnellement concurrentiels ou privés et les systèmes plus socialisés ne sont pourtant pas gommées :

- le mode de prélèvement auprès de l'assuré, sous forme de prime relativement uniforme d'un côté, sous forme de contribution proportionnelle au revenu de l'autre ;
- le degré de contrainte sur les participants. Dans la quasi-totalité des systèmes, assureurs et assurés doivent respectivement participer à l'ajustement au risque et s'assurer. Pour obtenir que la concurrence porte sur l'efficacité productive, on s'éloigne du modèle dans lequel l'assurance est volontaire, au sens où nous l'avons entendu dans le chapitre 7. Deux pays se distinguent à cet égard. Aux Etats-Unis, les assurés ne sont pas tenus de s'assurer, même s'il est toujours plus avantageux de participer au système mis en place par leur employeur que de s'assurer à titre individuel⁵⁷. Les assureurs, pour leur part, choisissent d'offrir leurs services au groupe d'assurés constitué, mais doivent, dans ce cas, accepter l'ajustement au risque. Il est intéressant de rappeler que la réforme Clinton prévoyait, outre l'universalisation de l'assurance, l'introduction d'un mécanisme de péréquation des risques entre les OGR. L'autre exception est l'Irlande, où tout assureur souhaitant entrer sur le marché devra participer à la péréquation mais où les personnes ne sont pas contraintes de s'assurer. Il faut rappeler qu'en Irlande, l'assurance ne concerne pas l'intégralité du risque et que les plus pauvres bénéficient de l'accès gratuit aux soins.

⁵⁶ La seule situation dans laquelle l'ajustement au risque ou son évaluation n'est pas nécessaire, est celle où un opérateur unique couvre l'intégralité d'une population définie.

⁵⁷ Il peuvent cependant préférer ne pas s'assurer ou bénéficier du contrats d'un conjoint en tant qu'ayant droit.

Tableau 3 : Principales caractéristiques des huit systèmes étudiés

	Objectif principal de la réforme	Droits et obligations des assurés	Obligations des assureurs	Mode de prélèvement des ressources	Financement des OGR
Allemagne	Réduction des inégalités et du morcellement A terme : amélioration de l'efficacité productive	Assurance obligatoire Choix de la caisse	Obligation d'accepter en assurance et de renouveler le contrat Pas d'exclusion des maladies déclarées	Cotisation proportionnelle au revenu qui dépend de la caisse	Cotisation perçue directement +/- compensation
Belgique	Amélioration de l'efficacité productive	Assurance obligatoire Choix de la caisse	Obligation d'accepter en assurance et de renouveler le contrat Pas d'exclusion des maladies déclarées	Cotisation proportionnelle au revenu uniforme	Capitation ajustée au risque partielle Financement rétrospectifs
Israël	Universalisation de l'assurance Amélioration de l'efficacité productive	Assurance obligatoire Choix de la caisse	Obligation d'accepter en assurance et de renouveler le contrat Pas d'exclusion des maladies déclarées	Cotisation proportionnelle au revenu uniforme	Capitation ajustée au risque partielle
Medicare	Amélioration de l'efficacité productive	Assurance obligatoire Choix entre le système traditionnel ou un HMO Choix du HMO	Obligation d'accepter en assurance et de renouveler le contrat Pas d'exclusion des maladies déclarées	Impôts et cotisation uniforme pour la partie B, indépendamment du choix du mode de gestion de la couverture	Capitation ajustée au risque
Pays-Bas	Amélioration de l'efficacité productive	Assurance obligatoire Choix de la caisse	Obligation d'accepter en assurance et de renouveler le contrat Pas d'exclusion des maladies déclarées	Cotisation proportionnelle au revenu uniforme Prime uniforme par caisse	Capitation ajustée au risque partielle Financement rétrospectifs
Irlande	Contrainte exogène (Europe) Préserver le mode de tarification équitable entre hauts et bas risques A terme : amélioration de l'efficacité productive	Assurance facultative Choix de l'assureur	Obligation d'accepter en assurance et de renouveler le contrat Réglementation des conditions d'exclusion des maladies déclarées	Prime Community Rating	Primes collectées +/- ajustement
Suisse	Universalisation de l'assurance Amélioration de l'efficacité productive	Assurance obligatoire Choix de la caisse	Obligation d'accepter en assurance et de renouveler le contrat Pas d'exclusion des maladies déclarées	Prime Community Rating	Primes collectées +/- ajustement
USA centrales d'achat	Amélioration de l'efficacité productive	Assurance facultative Choix de l'assureur	Obligation d'accepter en assurance et de renouveler le contrat Pas d'exclusion des maladies déclarées	Prime Community Rating	Primes collectées +/- ajustement

Toutefois, il ne serait pas rigoureux de déduire de cette analyse que les systèmes de santé eux-mêmes convergent. Comme le souligne Polton : "*la convergence des outils de régulation ne signifie pas la convergence des systèmes, qui demeurent très spécifiques en termes d'institutions et de jeux d'acteurs*"⁵⁸.

Au delà de ce constat, illustré dans la seconde section de ce chapitre, il faut rappeler qu'un élément fondamental de différenciation reste le panier de biens et services (uniformisé) sur lequel porte la concurrence "tempérée". Le chapitre 7 nous a conduit à décrire les différences entre les systèmes à cet égard. On peut prolonger l'analyse en montrant en quoi le choix de ce panier a un impact considérable sur la dynamique globale du système par rapport à la problématique qui nous intéresse ici.

1. Tout d'abord, la nature des biens et des services concernés par la concurrence tempérée aura un impact sur les modalités de recherche de l'efficacité productive, voire sur le mode de prise en charge des patients. Si cette concurrence ne portait que sur les dépenses hospitalières, l'OGR aurait intérêt à faire pression sur les hôpitaux, pour diminuer par exemple des durées de séjour, mais aussi à reporter la prise en charge des problèmes de santé des personnes couvertes vers la ville. Ceci n'est pas nécessairement synonyme d'une diminution de la qualité, si tant est que cette prise en charge est organisée. Or, dans une telle configuration, l'OGR n'aurait pas nécessairement les moyens ou la motivation suffisante pour le faire. Plus généralement, l'amélioration de l'efficacité globale passe souvent par la substitution de services et la réorganisation de la prise en charge. Il faut donc éviter d'instaurer des barrières de nature économique à cette substitution et le choix du panier couvert est à cet égard central. Cette préoccupation fait écho à un débat français ravivé à l'occasion des possibilités ouvertes par le plan Juppé de lancer des expérimentations dans le domaine des réseaux de soins : la non fongibilité des enveloppes de dépenses constitue une barrière importante à leur mise en œuvre⁵⁹.

⁵⁸ D. Polton, 1999, *la convergence des systèmes de santé*, Editorial de la lettre du collège des économistes de la santé.

⁵⁹ Les budgets hospitaliers, ceux des cliniques et les honoraires de différentes professions sont soumis à des enveloppes globales, sans possibilité de transfert des unes vers les autres.

2. Ensuite, la conclusion selon laquelle la forme de concurrence étudiée peut permettre de maintenir l'équité du système, voire de l'augmenter, dépend implicitement de l'étendue du panier de biens et services couverts et de l'articulation du système réglementé avec l'assurance volontaire.

Plus le panier sur lequel porte la concurrence tempérée est restreint, plus l'assurance volontaire joue un rôle important dans l'accès aux soins. La sélection dans ces différentes acceptions n'étant pas tempérée sur ce marché privé, on peut retomber dans un système générateur d'exclusion. En Israël, le choix du panier de biens sur lequel porte la concurrence tempérée est controversé pour cette raison.

3. Il reste enfin à souligner que l'assurance volontaire peut constituer un frein au bon fonctionnement de la concurrence tempérée. Nous avons souligné qu'une des raisons pour lesquelles les personnes hésitent à changer de caisse aux Pays-Bas, est qu'elles craignent de ne pas pouvoir souscrire une assurance "du troisième étage" auprès d'une nouvelle caisse. Si les mêmes entités gèrent l'assurance "de base" et l'assurance "volontaire", les assureurs peuvent utiliser cette dernière pour améliorer leur profil de risque et pour sélectionner les assurés. En refusant l'assurance volontaire à une personne, ils sont moins susceptibles de l'attirer pour l'assurance de base. Si l'ajustement au risque sur la couverture de base était parfait, ce qui n'est jamais le cas, les assureurs n'auraient qu'un intérêt limité à utiliser la sélection car ils n'en tireraient de bénéfice que dans le cadre de l'assurance volontaire. En réalité, plus le champ de l'assurance volontaire est large par rapport à celui de l'assurance "de base", plus cette incitation est forte. Elle disparaît, toutefois, si, comme en Allemagne, les caisses chargées de la gestion de l'assurance de base ne sont pas autorisées à couvrir les biens exclus du panier.

Enthoven (1994, pp. 1417) estime que les ventes jointes d'assurance devraient être interdites et que le panier de biens et services, objet de la concurrence tempérée, devrait être suffisamment large pour qu'il n'y ait pas de marché d'assurance volontaire⁶⁰.

⁶⁰ "Tie-in sales should not be allowed. I recommend a fairly comprehensive package of services ... so that there will be no market for supplemental coverage for covered services"

Ceci permet notamment d'instaurer des copaiements qui limitent le risque moral du côté des consommateurs, incitation qui ne fonctionne que si ces copaiements ne sont pas pris en charge par une assurance complémentaire⁶¹.

La dernière interrogation transversale que l'on peut soulever est celle de l'évaluation coût-bénéfice de l'ensemble de ces réformes. Il n'est pas question ici d'y apporter une réponse puisque d'une part, ces réformes ne sont pas abouties et que d'autre part, quand bien même un tel exercice serait possible, il dépasserait largement le cadre de ce travail. Néanmoins, il est apparu clairement que la réglementation de la concurrence nécessitait a minima de mettre en œuvre des mécanismes d'ajustement au risque complexes et coûteux. La gestion du risque par OGR nécessite de procéder à des investissements spécifiques (systèmes d'information, mise en place de nouvelles compétences...). Les gains en productivité sont-ils suffisant pour compenser ces coûts ? Plus généralement, le présupposé théorique selon lequel la gestion des risques est mise en œuvre de façon plus efficace par des organismes en concurrence reste très controversé (Boccognano et al., 1998).

La conclusion de ce travail nous conduira à mettre en lumière l'ensemble des questions qui restent ouvertes et qui sont autant de voies à explorer.

⁶¹ Fonction 3 dans le chapitre 7.