

Impact du passage à la T2A: une modélisation pour l'hôpital public

September 30, 2008

Abstract

Nous analysons les conséquences du passage d'un financement par dotation globale à un financement prospectif lié à l'activité et tentons de déterminer en quoi la réforme pourrait être préjudiciable ou bénéfique à la santé financière des hôpitaux publics. Nous développons un modèle théorique qui lie la demande des patients aux décisions conjointes des médecins et des managers d'hôpitaux pour analyser la façon dont le nouveau mode de financement affecte les décisions prises par l'ensemble des agents et donc l'équilibre budgétaire des établissements.

Classification JEL: I18, D21

Mots-clés: Tarification à l'activité, Hôpital public, Equilibre budgétaire

1 Introduction

L'histoire du système de financement des hôpitaux publics en France peut être décomposée en trois phases distinctes. Avant 1984, le budget des hôpitaux s'obtenait par la multiplication du nombre de journées d'hospitalisation par un prix de journée préalablement fixé. Ce système incitait évidemment les hôpitaux à allonger au maximum les durées de séjour. Entre 1984 et 2004, le budget des hôpitaux publics et des hôpitaux privés participant au service public (PSPH) était constitué d'une dotation globale. D'après la Mission Tarification à l'activité du Ministère de la Santé, ce système de financement générait des effets pervers en ne liant que très faiblement le budget à l'activité réalisée et en conduisant soit à la constitution de rentes de situation, soit à un manque de financement pour les structures les plus actives. Depuis 2004, le système de tarification à l'activité (T2A) a été introduit en France avec une mise en place progressive du financement prospectif qui devait aboutir en 2012. Le calendrier a été toutefois bouleversé très récemment. En effet, les dernières années de la phase de transition ont été abandonnées puisque le financement T2A, à hauteur de 50% en 2007, est passé directement à 100% le 1^{er} janvier 2008. Dès lors, la question de la santé financière des établissements - qui était en partie assurée par une dotation globale qui leur restait attribuée indépendamment de leur niveau d'activité - se pose dès aujourd'hui avec d'autant plus d'importance.

Aux Etats-Unis, les systèmes de paiements prospectifs dans le secteur hospitalier sont considérés, d'un point de vue théorique, comme supérieurs aux autres systèmes (Ma 1994, Newhouse 1996). Par contre, leur mise en place a modifié le comportement des hôpitaux et des médecins qui ont répondu aux nouveaux incitants économiques de la réforme : diminution de la durée de séjour, surcodage des diagnostics patient, sélection des patients, sélection des procédures et spécialisation des hôpitaux dans les activités lucratives, sélection des médecins consommant peu de ressources, substitution des patients vers des structures en dehors du système de paiement prospectif (centres de chirurgie ambulatoire, externalisation des activités radiologiques...) ou encore transfert plus rapide des patients vers des structures de moyen et de long séjour (Hodgkin et McGuire (1994), Ellis (1998), Louis et al. (1999) ou Serden et al. (2003)). En France, les bénéfices attendus de la réforme T2A sont de plusieurs ordres : une plus grande médicalisation du financement, une responsabilisation des acteurs et une incitation à s'adapter, une équité de traitement entre les secteurs et le développement des outils de pilotage médico-économique (contrôle de gestion) au sein des hôpitaux publics et privés. Comme aux Etats-Unis, certains effets pervers sont attendus et pourraient être analysés. Mais une question reste relativement peu documentée dans la littérature : quelle conséquence peut avoir la réforme T2A sur la pérennité financière des établissements ? Alors que le système de dotation globale déconnectait en grande partie le budget de l'activité produite, les hôpitaux profitaient largement de cette marge de manoeuvre pour établir leur équilibre budgétaire en jouant sur leur niveau d'activité et en étant

assurés d'une reconduction budgétaire avec un taux directeur. Avec le paiement prospectif, toutes les options deviennent possibles: des profits si le coût est inférieur au tarif, une augmentation de l'activité puisqu'elle sera automatiquement financée ou des difficultés financières dans le cas d'un tarif défavorable ou d'une réduction de la demande.

Dans ce travail, nous analysons les conséquences du passage d'un financement par dotation globale à un financement prospectif lié à l'activité et tentons de déterminer en quoi la réforme pourrait être préjudiciable ou bénéfique à la santé financière des hôpitaux publics. Nous développons un modèle théorique qui lie la demande des patients aux décisions conjointes des médecins et des managers d'hôpitaux pour analyser la façon dont le nouveau mode de financement affecte les décisions prises par l'ensemble des agents et donc l'équilibre budgétaire des établissements.

Nous proposons, à la suite des travaux de Custer et al. (1990), Willke et al. (1991), Dor et Watson (1995) et de Boadway et al. (2004), une modélisation adaptée au cas français. Nous reprenons la modélisation d'un jeu non coopératif entre un médecin et un gestionnaire hospitalier comme celle développée par Custer et al. (1990). Cependant, alors que ces auteurs se sont intéressés uniquement aux interactions entre ces acteurs du point de vue de l'efficacité productive, nous intégrons une fonction de demande pour prendre en compte l'impact de leurs décisions sur la satisfaction des patients. Nous intégrons également l'idée d'une fonction de production de santé introduite par Willke et al (1991) dont les arguments sont définis par les variables de décision du médecin et de l'hôpital. Nous adaptions enfin le modèle de Dor and Watson (1995) qui comparent deux types de paiements prospectifs (un paiement unique à partager entre l'hôpital et le médecin ou des paiements distincts) en nous focalisant sur des hôpitaux publics dont l'objectif économique n'est pas nécessairement la maximisation du profit.

Notre analyse est centrée sur un couple composé d'un hôpital et d'un médecin représentatif coordonnant leurs actions afin de produire conjointement des séjours hospitaliers homogènes. Ces séjours hospitaliers s'adressent à des individus sensibles à l'effort déployé par le médecin, à la durée d'hospitalisation et aux services fournis par l'hôpital pour améliorer la prise en charge et le confort des patients. Dans ce contexte, nous analysons l'impact de l'introduction d'un financement forfaitaire sur les décisions de l'hôpital et du médecin et la façon dont ces décisions affectent la santé financière des établissements. Le modèle que nous développons se situe à un niveau microéconomique dans lequel nous considérons deux agents représentatifs au sein d'un seul établissement. Bien qu'il soit évident que des contraintes d'environnement comme la concurrence ou des contraintes macroéconomiques comme le système d'enveloppe fermée qui accompagnent la réforme de la T2A existent, nous ne les intégrons pas directement dans la modélisation. Par contre, nous intégrons la réaction des hôpitaux et des médecins face à certains de leurs effets comme une baisse ou une augmentation des tarifs, une variation exogène de la demande à laquelle ils font face ou une modification de la structure des coûts hospitaliers.

Après une présentation du modèle dans la première section, nous analysons l'impact du passage à la T2A sur les décisions du médecin et de l'hôpital. Nous étudions en particulier les conditions sous lesquelles l'hôpital est capable d'accroître son activité tout en maintenant son équilibre budgétaire. Pour analyser les effets liés au changement de financement, nous neutralisons dans cette section les effets dus au volume budgétaire (en considérant que le tarif T2A est égal au coût moyen du séjour sous le financement précédent). Nous relâchons ensuite cette hypothèse dans la section 4 dans laquelle nous étudions le passage à la T2A sans neutralité budgétaire avec un tarif favorable ou défavorable à l'hôpital public. Dans une dernière section, nous étudions l'impact de chocs exogènes sur la demande ou sur les coûts sur l'équilibre budgétaire de l'hôpital après la réforme T2A.

2 Présentation du modèle

A l'image de Dor et Watson (1995), la modélisation est centrée sur un couple composé d'un hôpital et d'un médecin représentatif coordonnant leurs actions afin de produire conjointement des séjours hospitaliers homogènes. L'analyse est menée dans le cadre d'un modèle mono-périodique. Les séjours hospitaliers s'adressent à des individus dont l'état de santé est dégradé et qui ont par conséquent besoin de soins. Les patients sont sensibles à l'effort déployé par le médecin pour les soins prodigués (noté e et qui résume la production et la qualité des actes médicaux et/ou chirurgicaux, des consultations pré ou post opératoires...), à la durée d'hospitalisation (h) et au montant et à la qualité des inputs mis à disposition par l'hôpital pour améliorer la prise en charge et le confort des patients lors de leur séjour hospitalier (notés q et qui incluent les salles et leur niveau d'équipement, les lits et leur équipement ainsi que le nombre et la qualification des personnels soignants, administratifs ou médico-techniques). L'activité de l'hôpital est soumise à une contrainte budgétaire.

Nous considérons une fonction de production de services de santé $S(e, h, q)$ qui dépend du niveau d'effort du médecin, de la durée de séjour et du niveau d'input mis à disposition par l'hôpital. Cette fonction est supposée additive. Il n'y a donc pas d'effets croisés entre les différents arguments de la fonction qui est considérée comme croissante, continue et concave sur e et q . D'où:

$$S(e, h, q) = E(e) + H(h) + Q(q) \text{ avec } S_e > 0, S_q > 0, S_{ee} < 0, S_{qq} < 0 \quad (1)$$

Concernant la durée d'hospitalisation, nous considérons qu'il existe une durée de séjour au-delà de laquelle le niveau de production de santé diminue (par exemple par l'augmentation du risque d'infection nosocomiale, d'apparition d'escarres...). D'où:

$$\bar{h} = \{h \in \mathbb{R}^+ \mid S_h = 0, S_{hh} < 0\} \quad (2)$$

Cette durée de séjour correspond à la durée d'hospitalisation optimale du point de vue du patient. Notons que la structure additive de la fonction de production de services de santé implique que ce niveau optimal ne dépend pas des deux autres arguments.

A l'instar de Willke et al. (1991), la demande à laquelle font face le médecin et l'hôpital dépend de la production de services de santé $S(e, h, q)$ qu'ils proposent à chaque patient. Notons que cette fonction de production intègre à la fois la quantité et la qualité des soins. De sorte que :

$$D = D(S(e, h, q)) \quad (3)$$

Puisque D est une fonction croissante et concave en S et que S est une fonction croissante et concave en e et q alors nous pouvons réécrire la demande des patients comme:

$$D(e, h, q) \text{ avec } D_e > 0, D_q > 0, D_{ee} < 0, D_{qq} < 0 \quad (4)$$

Par rapport à la durée d'hospitalisation, la demande est d'abord croissante en h , atteint un maximum en \bar{h} et diminue ensuite.

Le médecin, qui décide de son niveau d'effort et de la durée d'hospitalisation, cherche à maximiser sa propre fonction objectif qui dépend du niveau de demande et de sa désutilité à l'effort. L'utilité du médecin est donc liée positivement à la satisfaction des patients via la demande et négativement au niveau d'effort qu'il fournit pour soigner les patients. Cette désutilité à l'effort $R(e(q))$ correspond au coût d'opportunité du temps passé à l'hôpital et qu'il ne peut consacrer à une activité libérale ou à ses loisirs. La désutilité est supposée être croissante et convexe en e ($R_e > 0$ et $R_{ee} > 0$). La désutilité est également liée au niveau d'input q mis à disposition par l'hôpital. Nous supposons en effet que les médecins valorisent éventuellement davantage leur travail dans un environnement favorable et que leur temps passé à l'hôpital peut être réduit s'ils bénéficient d'un plateau technique plus développé ou du service de davantage de personnel soignant ou administratif ($R_{eq} \leq 0$). Pour son activité à l'hôpital public, le médecin est supposé être complètement salarié de sorte que son utilité n'est pas liée à sa rémunération. Sous ces hypothèses, sa fonction objectif peut être exprimée comme suit:

$$U^M(e, h, q) = D(e, h, q) - R(e(q)) \quad (5)$$

De son côté, l'hôpital public décide du niveau d'input qu'il met à disposition des médecins et des patients. Comme nous introduisons les aspects économiques auxquels fait face l'hôpital au travers d'une contrainte budgétaire, sa fonction d'utilité dépend de la seule demande exprimée par les patients. Son objectif est donc directement lié à la satisfaction des patients:

$$U^H(e, h, q) = D(e, h, q) \quad (6)$$

Les coûts supportés par l'hôpital et liés à la durée d'hospitalisation d'un patient, $CH(h)$, (frais d'hébergement, frais de blanchisserie, frais de restauration...) sont considérés comme

linéaire en h ($CH(h) = c_h \cdot h$). Les mêmes hypothèses sont faites pour les coûts liés aux soins prodigués par les médecins à un patient: $CM(e)$ est une fonction linéaire en e ($CM(e) = c_e \cdot e$). L'hôpital supporte également des coûts fixes liés au niveau d'input qu'il met à disposition des patients ($CF(q)$, avec $CF(q) = c_q \cdot q$). Ces trois hypothèses de linéarité nous semblent appropriées car il n'y a pas a priori d'économies ou de déséconomies d'échelle à conserver un patient un jour de plus, à réaliser un acte médical supplémentaire ou à engager une nouvelle personne. Ces hypothèses ne contredisent pas le fait - reconnu dans les études de coûts - que les coûts d'une hospitalisation aient une forme en cloche par rapport au temps (croissants à partir de l'entrée à l'hôpital, atteignant leur sommet lors de l'acte chirurgical ou du traitement et décroissants ensuite). Le fait que les traitements ne soient pas repartis de façon uniforme tout au long du séjour du patient (ils sont en fait concentrés vers le milieu du séjour et moins nombreux au début et à la fin) donnent à la courbe de coût cette forme particulière indépendamment des hypothèses faites sur l'évolution des coûts associés à une journée, un traitement ou à un input supplémentaire. La dépense totale de l'hôpital est soumise à une contrainte budgétaire imposée par les pouvoirs publics. Elle est définie comme suit:

$$D(e, h, q) [CH(h) + CM(e)] + CF(q) \leq B \quad (7)$$

Le budget, B , dépend crucialement du mode de financement de l'hôpital. Il peut être considéré comme fixe et exogène au modèle dans le cas d'un financement par dotation globale ou variable et endogène au modèle ($B = D(e, h, q) * \bar{T}$) dans le cas d'une tarification à l'activité avec \bar{T} comme tarif associé à un séjour.

Nous évaluons le passage d'un financement par budget global vers un financement à l'activité sur les décisions du médecin et de l'hôpital. Ces décisions sont considérées dans un cadre non coopératif, dans lequel à la fois le médecin et l'hôpital adoptent un comportement passif, c'est-à-dire ne tirent pas parti de leur capacité à influencer, à travers leur propre choix, sur ceux de leur partenaire. Ceci conduit à un équilibre de Cournot/Nash.

Nous commençons par l'analyse de l'équilibre dans le cadre d'un financement par budget global. Nous étudierons à la section suivante l'impact du passage à la T2A pour le médecin et l'hôpital. Le programme d'optimisation du médecin peut être défini comme suit:

$$\underset{e, h}{Max} U^M(e, h, q) = D(e, h, q) - R(e(q)) \quad (8)$$

Les conditions de premier ordre correspondant à ce programme d'optimisation sont les suivantes:

$$D_e = R_e \quad (9)$$

$$D_h = 0 \quad (10)$$

Les deux conditions de second ordre nous indiquent bien que nous sommes en un maximum:

$$D_{ee} - R_{ee} < 0 \quad (11)$$

$$D_{hh} < 0 \quad (12)$$

Le médecin ne faisant face à aucune contrainte maximise son utilité en arbitrant entre la demande marginale liée à son niveau d'effort et la désutilité marginale produite par cet effort. Il maximise également son utilité s'il choisit la durée d'hospitalisation qui maximise la demande des patients ($\widehat{h}_{bg} = \bar{h}$). Notons qu'il considère le niveau d'input de l'hôpital comme donné.

Le programme d'optimisation de l'hôpital est défini de la façon suivante:

$$\begin{aligned} \text{Max}_q U^H(e, h, q) &= D(e, h, q) \\ \text{s.c. } D(e, h, q) [CH(h) + CM(e)] + CF(q) &\leq B \quad (\lambda) \end{aligned} \quad (13)$$

Les conditions de premier ordre correspondant à ce programme d'optimisation sont :

$$D_q - \lambda [D_q [CH(h) + CM(e)] + CF_q] = 0 \quad (14)$$

$$D(e, h, q) [CH(h) + CM(e)] + CF(q) \leq B \quad (15)$$

De la première CPO, on tire:

$$\lambda = \frac{D_q}{D_q [CH(h) + CM(e)] + CF_q} > 0 \quad (16)$$

La dernière inégalité provient des hypothèses faites sur la fonction de demande et sur les fonctions de coût. La positivité du multiplicateur indique clairement que la contrainte budgétaire sera toujours saturée:

$$D(e, h, q) [CH(h) + CM(e)] + CF(q) = B \quad (17)$$

L'interprétation de cette égalité est claire: l'hôpital a toujours intérêt à augmenter son niveau d'input pour maximiser sa fonction objectif et donc il utilisera la totalité de son budget.

3 Impact du passage à la T2A

Nous abordons dans cette section la question de l'impact du passage d'un financement par budget global à une tarification à l'activité sur le comportement du couple hôpital-médecin. Dans un premier temps, le passage à la T2A est considéré comme totalement neutre du point de vue de la rémunération par séjour hospitalier. En effet, nous nous intéressons aux modifications de comportement des acteurs face au changement de mode de tarification et non face à un éventuel changement du niveau global de dotation (qui est traité à la section suivante). Nous retenons

donc comme tarif à l'activité le budget initial de l'hôpital divisé par la demande optimale dans le modèle sous budget global:

$$\bar{T} = \frac{B}{D(\hat{e}_{bg}, \hat{h}_{bg}, \hat{q}_{bg})} \quad (18)$$

Le passage à la T2A concerne en premier lieu l'hôpital qui est soumis à la contrainte budgétaire¹. Le programme d'optimisation de l'hôpital devient :

$$\begin{aligned} \underset{q}{Max} U^H(e, h, q) &= D(e, h, q) \\ D(e, h, q) [CH(h) + CM(e)] + CF(q) &\leq D(e, h, q)\bar{T} \quad (\lambda^{t2a}) \end{aligned} \quad (19)$$

La première CPO est donnée par:

$$D_q - \lambda^{t2a} [D_q [CH(h) + CM(e) - \bar{T}] + CF_q] = 0 \quad (20)$$

Tout comme pour le cas d'un financement par budget global, l'hôpital a toujours intérêt à saturer sa contrainte budgétaire et sa condition d'optimisation est donc:

$$D(e, h, q) [CH(h) + CM(e)] + CF(q) = D(e, h, q)\bar{T} \quad (21)$$

La proposition suivante résume l'impact d'un passage du budget global à une tarification à l'activité.

Proposition 1 *L'impact de la T2A sur le comportement de l'hôpital public dépend de l'élasticité de la demande par rapport au niveau d'input. Si cette élasticité est supérieure à l'unité alors l'hôpital augmente son niveau d'input et tire profit du passage à une tarification à l'activité. Cela a pour conséquence d'augmenter l'effort du médecin et la satisfaction des patients. A l'inverse, si l'élasticité de la demande par rapport au niveau d'input est inférieure ou égale à l'unité, l'hôpital n'est pas incité à modifier le niveau de son input et le statu quo prévaut. La T2A n'a donc pas d'impact sur l'hôpital public dans ce dernier cas.*

Preuve:

La variable de contrôle de l'hôpital étant le niveau d'input q , nous allons étudier la variation de la contrainte budgétaire par rapport à une variation de q . En posant:

$$SF_{t2a}(e, h, q) = D(e, h, q) [\bar{T} - CH(h) - CM(e)] - CF(q)$$

¹A côté de la rémunération à l'activité qui finance les séjours hospitaliers, des ressources budgétaires sont prévues dans le cadre de « Missions d'Intérêt Général et d'Aide à la Contractualisation » (MIGAC) en vue de développer les autres missions des hôpitaux publics : activité d'enseignement et de recherche, rôle d'établissement de recours, contrats d'objectifs et de moyens, etc. Notons que l'insertion de ce budget ne modifierait pas les conclusions de notre modèle.

$SF_{t2a}(e, h, q)$ s'interprète comme le solde financier de l'hôpital (revenu - coût). La contrainte budgétaire impose que ce solde soit toujours positif ou nul. La variation de ce solde par rapport à une augmentation du niveau d'input est donné par :

$$\frac{\delta SF_{t2a}(e, h, q)}{\delta q} = D_q [\bar{T} - CH(h) - CM(e)] - CF_q \quad (22)$$

Or, d'après(21), nous avons à l'équilibre budgétaire:

$$[\bar{T} - CH(h) - CM(e)] = \frac{CF(q)}{D(e, h, q)} \quad (23)$$

et, d'après la définition des coûts fixes, puisque $CF(q) = c_q \cdot q$, nous obtenons:

$$CF_q = \frac{\delta CF(q)}{\delta q} = c_q \quad (24)$$

En insérant ces expressions dans (22), nous obtenons:

$$\frac{\delta SF_{t2a}(e, h, q)}{\delta q} = D_q \left(\frac{c_q \cdot q}{D(e, h, q)} \right) - c_q \quad (25)$$

$$= c_q \left(D_q \frac{q}{D(e, h, q)} - 1 \right) \quad (26)$$

$$= c_q (\epsilon_{D/q} - 1) \quad (27)$$

où $\epsilon_{D/q}$ est l'élasticité de la demande par rapport au niveau d'input de l'hôpital. Finalement, comme le coût fixe unitaire est strictement positif, $c_q > 0$, nous obtenons le résultat suivant:

$$\frac{\delta SF_{t2a}(e, h, q)}{\delta q} > 0 \Leftrightarrow \epsilon_{D/q} > 1 \quad (28)$$

$$\frac{\delta SF_{t2a}(e, h, q)}{\delta q} = 0 \Leftrightarrow \epsilon_{D/q} = 1 \quad (29)$$

$$\frac{\delta SF_{t2a}(e, h, q)}{\delta q} < 0 \Leftrightarrow \epsilon_{D/q} < 1 \quad \blacksquare \quad (30)$$

Par conséquent, par rapport à l'équilibre en budget global, l'hôpital peut augmenter son utilité (en augmentant son niveau d'input car la demande est sensible à ce dernier) si l'élasticité de la demande par rapport au niveau d'input est supérieure à l'unité. L'intuition économique de ce résultat est relativement simple: en augmentant son niveau d'input, l'hôpital agit positivement sur la demande et donc obtient un supplément de revenu mais il augmente également son coût variable et son coût fixe. Dans le cas où le revenu supplémentaire généré par l'augmentation de la demande est supérieur au coût supplémentaire, l'hôpital a intérêt à augmenter son niveau d'input puisqu'il satisfera sa contrainte budgétaire tout en augmentant son utilité. La question qui suit naturellement est celle du niveau jusqu'auquel l'hôpital peut augmenter son niveau d'input.

Corollaire 1 . Après le passage à la T2A, le niveau d'équilibre de l'input de l'hôpital public est tel que l'élasticité de la demande par rapport à ce niveau d'input est inférieure à l'unité.

Preuve:

Si l'élasticité de la demande par rapport au niveau d'input est inférieure à l'unité avant le passage à la T2A, nous avons vu que le statu quo prévalait. Si l'élasticité est supérieure à l'unité alors il faut s'intéresser à la variation de l'élasticité par rapport à la variation du niveau d'input:

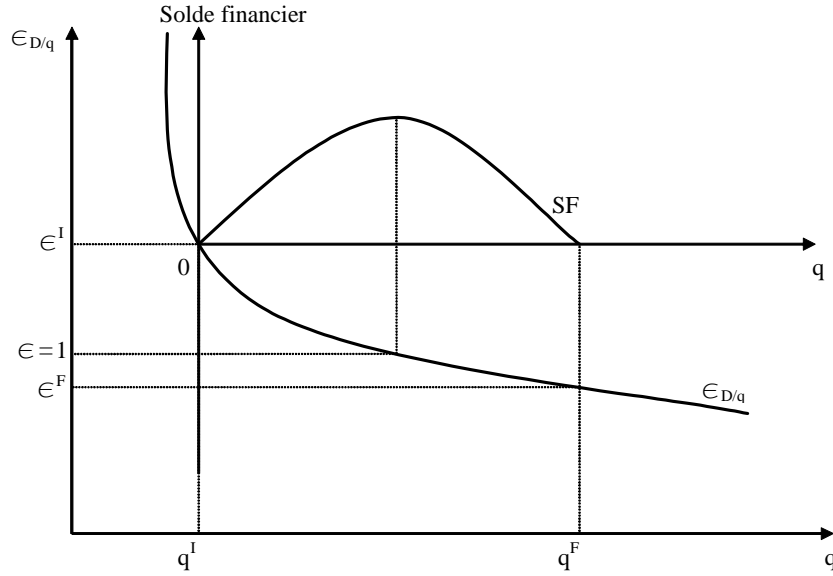
$$\frac{\delta \epsilon_{D/q}}{\delta q} = \frac{\delta \left(D_q \frac{q}{D(e, h, q)} \right)}{\delta q} = D_{qq} \frac{q}{D(e, h, q)} + D_q \frac{D(e, h, q) - D_q q}{D(e, h, q)^2} = D_{qq} \frac{q}{D(e, h, q)} + \frac{D_q}{D(e, h, q)} (1 - \epsilon_{D/q}) \quad (31)$$

Comme au départ l'élasticité est supérieure à l'unité, cette variation est négative car le premier terme de la somme est aussi négatif. Par conséquent, l'élasticité diminue avec l'augmentation du niveau d'input et deviendra inférieure à l'unité. Dans cette zone, le signe de l'expression (31) devient indéterminé et l'élasticité pourrait augmenter avec q . Cependant, elle ne repasse jamais au-dessus de 1. En effet, dans la zone croissante de (31), il est facile de montrer que:

$$\epsilon_{D/q} < 1 + D_{qq} \frac{q}{D_q} \text{ avec } D_{qq} \frac{q}{D_q} < 0 \text{ d'où } \epsilon_{D/q} < 1 \quad (32)$$

Le niveau d'input optimal est celui qui rééquilibre la contrainte budgétaire de l'hôpital. En effet, tant que l'élasticité est supérieure à l'unité, l'augmentation du niveau d'input améliore le solde de l'hôpital (revenu marginal supérieur au coût marginal). Lorsque l'élasticité devient inférieure à l'unité, le solde marginal devient négatif. L'hôpital poursuit donc l'augmentation du niveau d'input jusqu'à équilibrer son solde. Au niveau d'input optimal, l'élasticité de la demande est donc inférieure à l'unité. ■

Ce résultat est illustré par la figure suivante (cas 1). Cette figure nous servira d'outil d'analyse pour étudier les chocs exogènes considérés dans la section suivante. Nous représentons à la fois l'évolution de l'élasticité et du solde financier avec le niveau d'input q .



Effet du passage à la T2A - cas 1

La figure comporte un double système d'axe: le premier est lié aux variations de $\epsilon_{D/q}$ (axe extérieur) et le second est associé au solde financier de l'hôpital (axe intérieur). L'interprétation des courbes doit se faire par rapport à l'élasticité unitaire ($\epsilon = 1$). La situation initiale est ici caractérisée par une élasticité supérieure à 1 ($\epsilon^I > \epsilon = 1$) associée à un équilibre budgétaire (solde financier nul) au niveau initial q^I . Lorsque q augmente, l'élasticité diminue, mais tant qu'elle reste supérieure à 1, le solde financier de l'hôpital augmente et atteint son maximum pour ($\epsilon = 1$). Ensuite, dans la zone où l'élasticité est inférieure à l'unité, le solde financier diminue pour retrouver l'équilibre budgétaire en q^F . Nous vérifions donc qu'en ce point l'élasticité finale est inférieure à l'unité ($\epsilon^F < 1$).

Après avoir traité de l'impact du passage à la T2A pour l'hôpital, il faut s'interroger sur son impact concernant le médecin. A priori, la contrainte budgétaire n'affecte pas le médecin et il n'a donc pas de réaction directe au passage à la T2A. Cependant, l'hôpital a modifié son comportement en augmentant son niveau d'input et donc la demande. Le médecin pourrait donc éventuellement réagir à cette augmentation de la demande en ajustant son effort et la durée d'hospitalisation. Pour répondre à cette question, il convient d'analyser les variations de h et de e par rapport à la variation de q . L'impact d'une variation de q sur h est définie par:

$$\frac{dh}{dq} = - \left(\frac{\frac{\partial CPO}{\partial h}}{\frac{\partial CPO}{\partial q}} \right) = 0 \quad (33)$$

Le médecin choisit à l'équilibre la durée de séjour optimale du point de vue du patient et ce indépendamment du niveau de q qui, s'il influe sur son utilité, ne modifie pas sa décision de durée de séjour. Par conséquent, dans une interaction à la Cournot, la modification du niveau d'input de l'hôpital ne modifie pas son choix initial.

L'impact d'une variation de q sur e est de son coté définie par:

$$\frac{de}{dq} = - \left(\frac{\frac{\partial CPO}{\partial e}}{\frac{\partial CPO}{\partial q}} \right) \geq 0 \quad (34)$$

Une augmentation du niveau d'input de l'hôpital induit toute chose égale par ailleurs une augmentation ou un même niveau d'effort du médecin. Lorsque l'hôpital met plus d'input à disposition du médecin, le coût marginal de l'effort de ce dernier (R_e dans 9) diminue ou demeure inchangé ($R_{eq} \leq 0$). Ce qui conduit le médecin soit à augmenter, soit à ne pas modifier son effort. Deux situations sont donc envisageables. D'une part, la désutilité du médecin peut ne pas être sensible au niveau d'input de l'hôpital et dans ce cas, nous avons $\frac{de}{dq} = 0$ et le comportement du médecin est insensible au passage à la T2A et à la réaction de l'hôpital. La condition sur l'élasticité de la demande détermine donc entièrement l'issue du passage à la T2A. D'autre part, le médecin peut être sensible au niveau d'input de l'hôpital ce qui l'incite à faire plus d'effort ($\frac{de}{dq} > 0$). Dans ce cas, la contrainte budgétaire de l'hôpital s'en trouve modifiée puisque l'augmentation de l'effort du médecin accroît le coût associé pour l'hôpital. Il faut donc analyser si un nouvel équilibre se produit en étudiant les fonctions de réaction des deux agents. On peut montrer qu'un nouvel équilibre s'établit dès lors que la réaction du médecin n'est pas excessivement sensible à la modification du niveau d'input de l'hôpital².

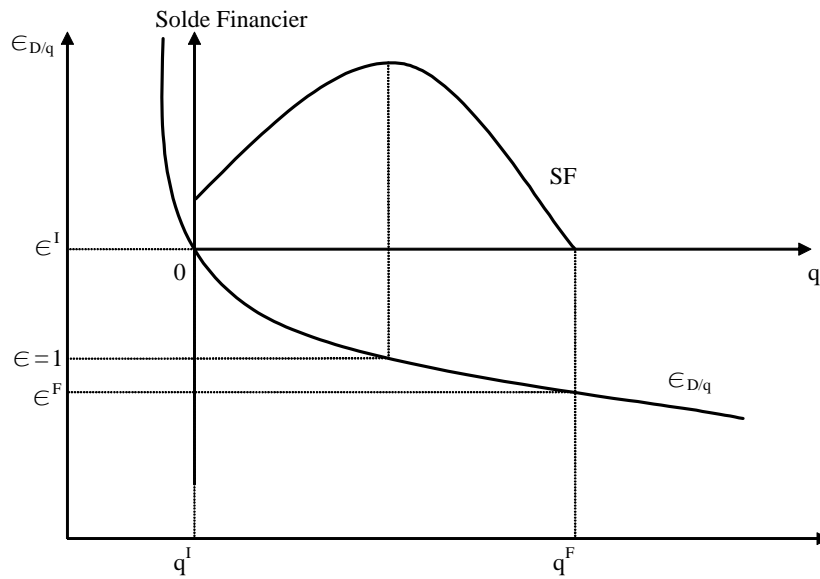
4 Impact du passage à la T2A sans neutralité budgétaire

4.1 Le passage à la T2A est budgétairement favorable

4.1.1 L'élasticité de la demande par rapport à l'input de l'hôpital est supérieure à 1.

Dans cette section, nous analysons le comportement de l'hôpital public au passage à la T2A avec un tarif qui ne respecterait pas la neutralité budgétaire. Dans le cas où le tarif est supérieur au tarif en neutralité budgétaire ($\bar{T} > \bar{T} = \frac{B}{D(e,h,q)}$), la situation est favorable à l'hôpital qui bénéficie d'un solde financier positif lors du passage à la T2A. Ce cas est relativement similaire au cas 1 mise à part la situation initiale. Comme décrit à la figure représentant le cas 2, l'hôpital réagit en augmentant son input jusqu'au niveau q^F . Dans ce cas, la demande augmente et l'hôpital est à son équilibre budgétaire. Par rapport au cas de référence (cas 1 avec neutralité budgétaire), le niveau d'input proposé par l'hôpital est plus important.

²Ce résultat est disponible sur demande aux auteurs. Il s'appuie sur une analyse de la convergence de l'équilibre similaire à celle d'un modèle de type CobWeb. Il dépend donc de la pente des fonctions de réaction des agents.

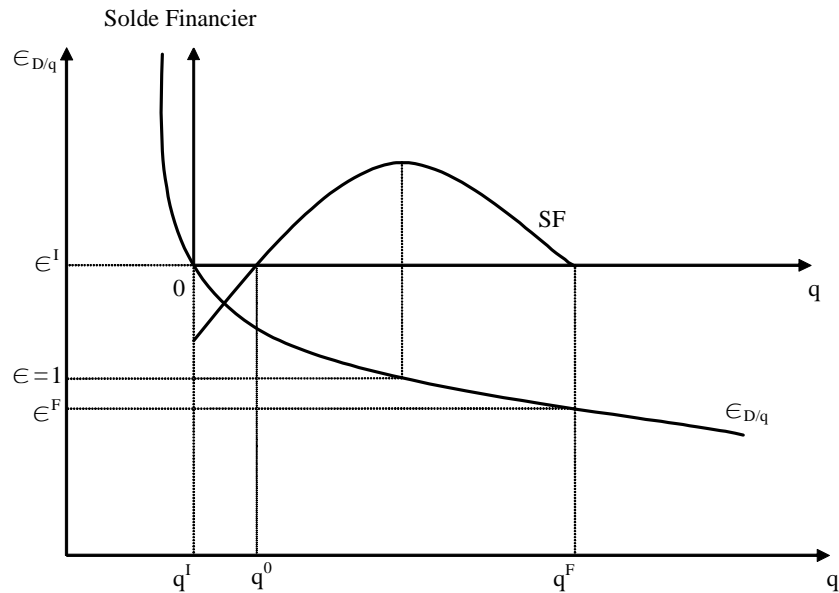


Effet du passage à la T2A - cas 2

4.1.2 L'élasticité de la demande par rapport à l'input de l'hôpital est inférieure à 1.

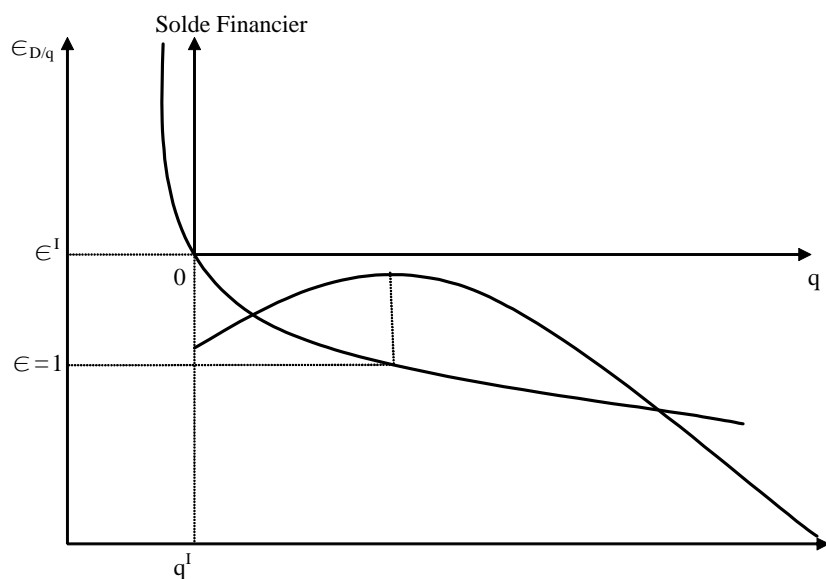
Ici encore, le passage à la T2A est budgétairement favorable à l'hôpital public ($\bar{T} > \bar{T} = \frac{B}{D(e,h,q)}$). Puisqu'il se trouve dans une zone où son élasticité est inférieure à l'unité, l'hôpital se contente d'épuiser son budget. Il en résulte une augmentation du niveau d'input jusqu'au niveau q^F (voir figure représentant le cas 3). La demande augmente et l'hôpital est à son équilibre budgétaire. Par rapport aux deux cas précédents (cas 1 avec neutralité budgétaire et cas 2 avec une élasticité de départ supérieure à l'unité), le niveau d'input proposé par l'hôpital est moins important.

son budget redevient équilibré. Par rapport aux deux premiers cas traités (budget favorable ou neutre), le niveau d'input proposé par l'hôpital est moins important.



Effet du passage à la T2A - cas 4

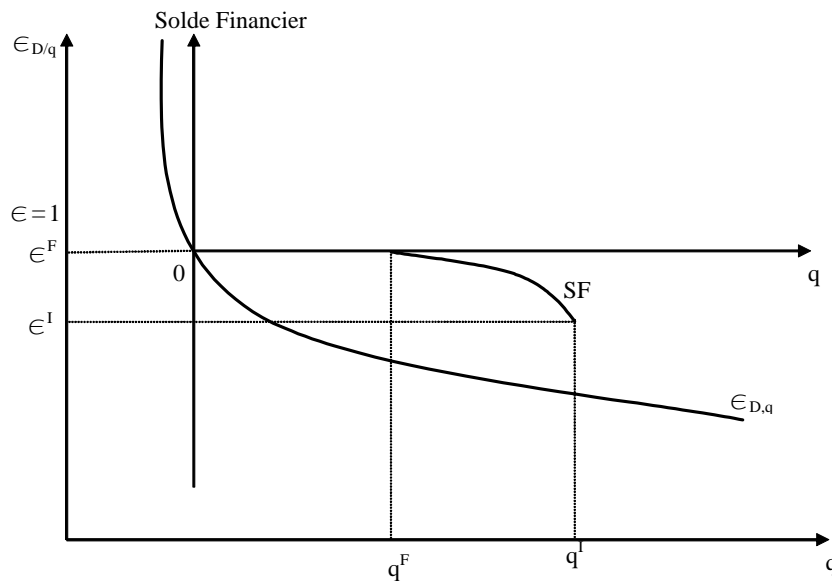
2. L'élasticité de la demande par rapport à l'input est supérieure à 1 et l'hôpital ne récupère pas son équilibre budgétaire. Comme dans la situation précédente, l'hôpital démarre avec une élasticité de la demande par rapport à l'input supérieure à l'unité mais il se trouve en déficit budgétaire au passage à la T2A s'il maintient son input initial (q^I).



Effet du passage à la T2A - cas 5

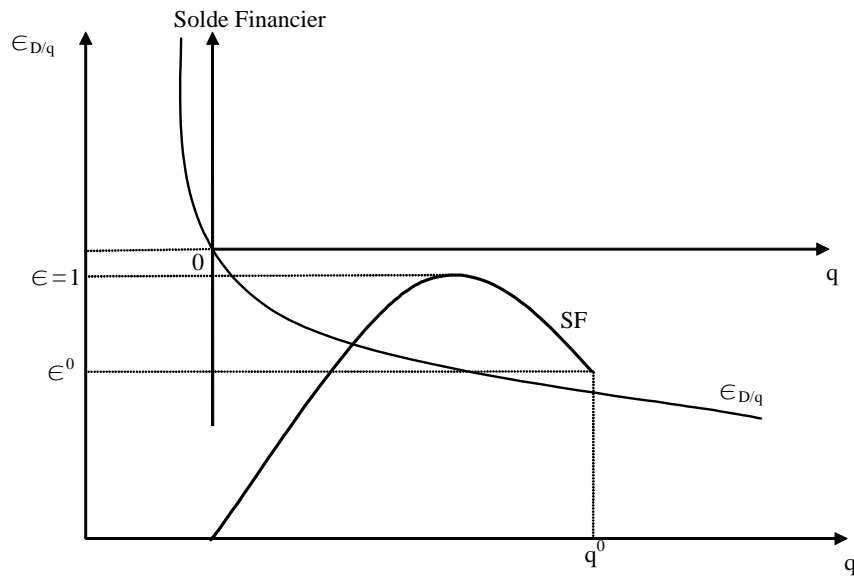
Augmentant son niveau d'input, il se peut néanmoins qu'il atteigne une élasticité égale à l'unité avant d'atteindre l'équilibre budgétaire (voir figure représentant le cas 5). Toute nouvelle augmentation d'input le conduit à creuser son déficit. Par conséquent, l'hôpital ne parvient en aucun cas à retrouver l'équilibre budgétaire et se retrouve en déficit.

3. L'élasticité de la demande par rapport à l'input est inférieure à 1 et l'hôpital récupère son équilibre budgétaire. L'hôpital se retrouve en déficit budgétaire au passage à la T2A s'il maintient son input initial (q^I). Mais puisqu'il démarre avec une élasticité de la demande par rapport à l'input qui est défavorable ($\epsilon^I < \epsilon = 1$), il est contraint de diminuer son niveau d'input pour réduire son déficit ce qui a pour effet d'augmenter l'élasticité de la demande à laquelle il fait face. C'est la situation décrite sur la figure qui représente le cas 6 où il parvient à rétablir l'équilibre budgétaire avant d'arriver à une élasticité unitaire de la demande par rapport à l'input.



Effet du passage à la T2A - cas 6

4. L'élasticité de la demande par rapport à l'input est inférieure à 1 et l'hôpital ne récupère pas son équilibre budgétaire Ici encore, le maintien de l'input d'avant le passage à la T2A (q^I) conduit l'hôpital à un déficit budgétaire qu'il peut réduire s'il diminue son niveau d'input. Cette réduction de q accroît l'élasticité de la demande à laquelle il fait face. Mais lorsqu'il atteint une élasticité unitaire avant d'avoir résorbé son budget (voir figure représentant le cas 7), l'hôpital ne peut pas rétablir la situation initiale de déficit.



Effet du passage à la T2A - cas 7

5 Impact de chocs exogènes sur la demande ou sur les coûts

Nous analysons l'impact d'une diminution de la demande qui peut provenir par exemple d'une perte de parts de marché face au secteur privé et/ou d'une augmentation des coûts résultant par exemple d'une augmentation du coût d'un traitement après une innovation technologique ou thérapeutique sans ajustement du tarif. Nous mettons en évidence le danger que ces chocs exogènes représentent pour l'équilibre budgétaire de l'hôpital. Il est important de garder à l'esprit que, quels que soient l'élasticité de départ et le niveau du tarif proposé (favorable, neutre ou défavorable), le passage à la T2A conduit l'hôpital à un niveau d'input associé à une élasticité inférieure ou égale à l'unité. Nous démarrons donc notre analyse en ne considérant que cette gamme d'élasticités en supposant que les chocs se produisent après le passage à la T2A. Cette section nous montre que quelles que soient les conditions de la réforme ex ante, le passage à la T2A place, ex post, l'hôpital dans une situation vulnérable à une diminution de sa demande ou à une augmentation de ses coûts.

5.1 Une diminution de la demande

Pour l'hôpital, une diminution de la demande peut conduire au déficit. En effet, l'hôpital se trouve dans une zone où l'élasticité de la demande par rapport à l'input est inférieure à l'unité. Si l'hôpital veut récupérer son équilibre budgétaire, il doit nécessairement réduire son input. Il peut y parvenir comme dans le cas 6 déjà analysé mais cette condition n'est pas suffisante, car en diminuant q , il pourrait atteindre une élasticité de la demande égale à l'unité avant d'avoir rééquilibré son budget (cfr. cas 7). Notons par ailleurs que cette diminution de l'input de

l'hôpital renforce cette tendance à la baisse de la demande qui lui est adressée ($D_q > 0$).

5.2 Une augmentation des coûts

Une augmentation des coûts fixes ou une augmentation des coûts variables telle que $\bar{T} > CH(h) + CM(e)$ (le tarif reste supérieur aux coûts variables) peuvent aussi conduire l'hôpital soit à rétablir son équilibre budgétaire en réduisant son input soit au déficit. L'effet sur l'équilibre de l'hôpital est le même qu'une réduction de la demande. Nous renvoyons donc le lecteur au cas précédent.

Lorsque l'augmentation des coûts variables est telle que $\bar{T} < CH(h) + CM(e)$ (le tarif proposé est inférieur aux coûts variables), l'hôpital ne peut réagir et est donc conduit au déficit budgétaire. Cette situation renvoie à la section 4.2.1.

6 Conclusion

La comparaison d'un système de financement de type budget global versus un financement à l'activité a été l'objet d'une littérature abondante ces dix dernières années. Cependant, comme la plupart des pays de l'OCDE ont opté pour un mode de financement à l'activité, la question aujourd'hui n'est plus tant de montrer que ce système est supérieur aux autres mais de réfléchir à l'impact de ce mode de financement sur le comportement des acteurs du secteur hospitalier. Il nous semble intéressant d'analyser notamment en quoi le passage à la T2A en France va modifier le comportement économique des médecins et des gestionnaires hospitaliers dans la mesure où ces changements ont un impact sur la demande, le niveau d'effort de chaque agent ou sur le budget des hôpitaux. Le passage à la T2A remet en question l'équilibre financier et la pérennité des établissements, d'autant que la part du financement T2A est passée, contrairement à ce qui était initialement prévu, directement à 100% depuis le 1^{er} janvier dernier.

Notre travail offre une contribution à la compréhension de cette nouvelle problématique en proposant une modélisation de la coordination entre médecins et gestionnaires hospitaliers pour la production de services de santé. Il reste évidemment contingent à tout exercice de modélisation pour lequel des hypothèses simplificatrices sont nécessaires. En considérant une neutralité budgétaire dans laquelle le tarif de la T2A est identique au coût moyen sous budget global, l'hôpital est directement affecté par la réforme au travers de sa contrainte budgétaire. Nous étudions les conditions sous lesquelles l'hôpital peut développer son activité grâce à cette réforme. Nous montrons dans ce travail que la réponse dépend de l'élasticité de la demande des patients par rapport à l'input de l'hôpital. Si cette élasticité est supérieure à l'unité, l'hôpital tire profit du passage à une tarification à l'activité puisqu'en augmentant son niveau d'input, il induit une augmentation de la demande tout en continuant à satisfaire sa contrainte budgétaire. L'impact pour le médecin est positif puisque son niveau d'utilité dépend positivement de la

demande et du niveau d'input de l'hôpital. L'impact pour les patients est également positif puisque la demande traduit la satisfaction des patients. A contrario, si le niveau de départ de cette élasticité est inférieur à l'unité, l'hôpital ne peut pas tirer profit d'une augmentation de son niveau d'input et le statu quo prévaut.

Nous montrons également que la situation est davantage contrastée dans le cas où le tarif de la T2A ne respecte pas la neutralité budgétaire. L'hôpital profite évidemment d'un tarif favorable mais un tarif défavorable peut dans certains cas l'amener à un déficit. De même, nous avons montré que l'hôpital est toujours amené, même si le tarif est favorable ou neutre, à un niveau d'input tel que son élasticité est inférieure à l'unité. Dans ce cas, des chocs exogènes qui viendraient diminuer la demande de l'hôpital ou augmenter ses coûts de production, l'amèneraient au mieux à s'adapter en diminuant ses ressources pour retrouver son équilibre budgétaire ou au pire à subir un déficit qu'il ne pourra résorber seul.

Plusieurs extensions du modèle sont envisageables. La première concerne la prise en compte du caractère multi-produits de l'activité hospitalière. Nous n'avons en effet considéré qu'une seule pathologie qui exclut les effets de substitution ou les possibilités de subventions croisées entre activités rentables et déficitaires. La seconde intègre les différents niveaux de sévérité d'une même pathologie. Comme cela a été montré par Ma (1994) et testé par Ellis et McGuire (1996), la prise en compte de cette hétérogénéité peut modifier le comportement des agents et conduire à de nouveaux équilibres dans lesquels sont présents des effets de sélection de patients. Enfin, la spécification de la fonction de production de services de santé (considérée comme additive) ne permet pas d'intégrer d'éventuels effets de complémentarité entre les différentes variables de décision qui permettraient par exemple de modéliser l'impact du niveau d'input sur la durée de séjour optimale.

Remerciements

Un des auteurs a bénéficié du financement du Ministerio espagnol de Educacion y Ciencia sous le projet SEJ 2006 - 06309/ECON et de la bourse du gouvernement fédéral belge sous le projet PAI P6/09.

Références

- Boadway R., Marchand M. et Motohiro S. (2004) An optimal contract approach to hospital financing. *Journal of Health Economics*, 23, 85-110.
- Custer W.S., Moser J.W., Musacchio R.A. et Willke R.J. (1990) The production of health care services and changing hospital reimbursement – The role of hospital-medical staff relationships. *Journal of Health Economics*, 9, 167-192.

Dor A. et Watson H. (1995) The hospital-physician interaction in U.S. hospitals : evolving payment schemes and their incentives. *European Economic Review*, 39(3/4), 795-802.

Ellis R.P. (1998) Creaming, Skimping, and Dumping: Provider Competition on the Intensive and Extensive Margins. *Journal of Health Economics*, 17(5), 537-555.

Ellis R.P. et McGuire T.G. (1996) Hospital response to prospective payment: Moral hazard, selection, and practice-style effects. *Journal of Health Economics*, 15(3), 257-277.

Hodgkin D. et McGuire T.G. (1994) Payment levels and hospital response to prospective payment. *Journal of Health Economics*, 13, 1-29.

Louis D.Z., Yuen E.J., Braga M., Cicchetti A ., Rabinowitz C., Laine C. et Gonnella J.S. (1999) Impact of a DRG-based hospital financing system on quality and outcomes of care in Italy. *Health Service Research*, 34(1 Pt 2), 405–415.

Ma C.A. (1994) Health care payment systems: cost and quality incentives. *Journal of Economics and Management Strategy*, 8(2), 93-112.

Newhouse J.P. (1996) Reimbursing health plans and health providers : selection versus efficiency in production. *Journal of Economic Literature*, 34, 1236-1263.

Serden L., Lindqvist R. et Rosen M. (2003) Have DRG-based prospective payment systems influenced the number of secondary diagnoses in health care administrative data? *Health Policy*, 65(2), 101-107.

Willke R.J., Custer W.S., Moser J.W. et Musacchio R.A. (1991) Collaborative production and resource allocation: the consequences of prospective payment for hospital care. *The Quarterly Review of Economics and Business*, 31(1), 28-47.