

© OECD, 2003.

© Software: 1987-1996, Acrobat is a trademark of ADOBE.

All rights reserved. OECD grants you the right to use one copy of this Program for your personal use only. Unauthorised reproduction, lending, hiring, transmission or distribution of any data or software is prohibited. You must treat the Program and associated materials and any elements thereof like any other copyrighted material.

All requests should be made to:

Head of Publications Service,
OECD Publications Service,
2, rue André-Pascal,
75775 Paris Cedex 16, France.

© OCDE, 2003.

© Logiciel, 1987-1996, Acrobat, marque déposée d'ADOBE.

Tous droits du producteur et du propriétaire de ce produit sont réservés. L'OCDE autorise la reproduction d'un seul exemplaire de ce programme pour usage personnel et non commercial uniquement. Sauf autorisation, la duplication, la location, le prêt, l'utilisation de ce produit pour exécution publique sont interdits. Ce programme, les données y afférentes et d'autres éléments doivent donc être traités comme toute autre documentation sur laquelle s'exerce la protection par le droit d'auteur.

Les demandes sont à adresser au :

Chef du Service des Publications,
Service des Publications de l'OCDE,
2, rue André-Pascal,
75775 Paris Cedex 16, France.

LA MESURE DE LA QUALITÉ DES SOINS A L'HÔPITAL : L'ÉTAT DE L'ART QUELLE INFORMATION DONNER AU PUBLIC ?

par

Gérard de Pouvourville et Etienne Minvielle*

Résumé

La question de la qualité des soins délivrés par les établissements hospitaliers est devenue un sujet central de la régulation des systèmes de soins pour plusieurs raisons : la crise de financement des services, l'émergence d'une plus grande exigence des individus à l'égard des professionnels de santé, liée entre autre à la diffusion plus massive d'informations sur les soins de santé, enfin, le développement des technologies nouvelles de traitement de l'information. L'ensemble de ces éléments a conduit à remettre en cause le monopole de la profession médicale et à revendiquer le développement de dispositifs externes de monitoring, voire de contrôle, des services rendus.

Ces exigences se sont traduites par l'apparition d'un champ de recherche nouveau dans le secteur de la santé publique, la recherche portant sur la mesure de la qualité des soins, et notamment des résultats de ces soins (outcomes), et par des initiatives des payeurs ou des agences publiques pour intégrer cette mesure des résultats dans l'allocation des ressources aux hôpitaux. Parallèlement, ces mêmes acteurs et la presse ont mis à la disposition du grand public les résultats de travaux comparant des établissements entre eux sur une sélection d'indicateurs. Ces initiatives sont très inégales en nature et en quantité dans les différents pays de l'OCDE. Du fait d'une régulation concurrentielle du système de soins de santé, les États-Unis ont été très en pointe tant du point de vue de travaux méthodologiques que de la mise en place de dispositifs de contrôle externe. La diffusion de ces travaux dans les pays européens et d'autres pays de l'OCDE a été plus lente, en partie pour des raisons politiques, mais aussi pour des raisons de retard en matière de systèmes d'information informatisés.

Le but de ce chapitre est de tirer les leçons des expériences conduites dans une sélection de pays. Deux points seront particulièrement traités. D'une part, on fera une revue de l'état de l'art en matière de méthodes de mesure de la qualité des soins et des résultats, en plaçant l'accent prioritairement sur le domaine de la mesure des résultats cliniques, ou en termes d'état de santé, des soins délivrés dans les établissements hospitaliers. On se limitera dans l'exposé au secteur des soins aigus. D'autre part, on analysera l'utilisation qui a été faite de ces méthodes par des instances externes aux établissements, payeurs, agences publiques, presse grand public, en identifiant quelles ont été les finalités poursuivies, les modalités et difficultés de mise en œuvre, enfin, quelles en sont les évaluations à l'heure actuelle et les recommandations que l'on peut en déduire.

Introduction

La mise sous contrainte de ressources des systèmes de soins de santé des pays développés a conduit à une attention accrue de la part des financeurs publics et privés aux coûts unitaires dans la

* Centre de Recherche en économie et gestion appliquées à la santé, INSERM/CNRS, 80 rue du Général Leclerc, 94276 Le Kremlin Bicêtre Cedex France : gdepouvo@kb.inserm.fr et minvielle@kb.inserm.fr

délivrance des services. Assez vite, cependant, est apparue la préoccupation de la qualité des soins délivrés : il ne fallait pas que la réduction ou la modération de l'évolution des coûts se traduise par des soins de moins bonne qualité. Sur ce plan, plusieurs lignes d'action ont été conduites par les différents acteurs. On distingue quatre logiques principales : une logique *professionnelle*, de rationalisation des pratiques, une logique *manageriale*, s'inspirant des pratiques de gestion de l'amélioration continue de la qualité dans l'industrie manufacturière, une logique *concurrentielle*, visant à exiger de la part des producteurs de soins la publication d'indicateurs sur la qualité des soins, enfin, une logique *de politique publique*, les pouvoirs publics jouant un rôle moteur dans la publication de tels indicateurs soit au nom de l'« accountability » des services publics, soit au nom d'une logique de réglementation correctrice de l'asymétrie d'information entre usagers potentiels des services et producteurs de soins. Les deux dernières logiques aboutissent au même résultat – la publication d'indicateurs de qualité des soins permettant d'évaluer la performance des services et de les comparer entre eux – mais les modalités d'action ne sont pas identiques : dans un cas, ce sont les acheteurs de soins (le plus souvent les assureurs) qui se servent du bras de levier de la mise en concurrence pour exiger des producteurs la production d'information, dans le deuxième cas, le mode d'intervention est la réglementation.

Ces logiques se sont mises en place à des rythmes et des modalités différents selon les pays. Les États-Unis d'Amérique ont commencé très en avance des autres pays de l'OCDE. Par exemple, l'Agence fédérale de financement des soins de santé (HCFA), gestionnaire du fonds Medicare, a publié dès 1986 des taux de mortalité ajustés sur les Diagnosis Related Groups. La Joint Commission for the Accreditation of Healthcare Organizations (la JCAHO) publie un guide présentant les différents systèmes de mesure de la performance des hôpitaux (JCAHO, 2001). En 1991, plusieurs organismes privés commerciaux ou sans but lucratif ont créé le National Committee for Quality Assurance (NCQA), dont le rôle est de fournir des informations sur les différents plans d'assurance maladie offerts aux Américains. Dans ce pays, cette avance peut s'expliquer par quatre caractéristiques : l'obligation pour le gouvernement fédéral et ceux des États de rendre des comptes régulièrement aux instances parlementaires sur les actions mises en place, la croyance forte dans la concurrence comme moteur principal de régulation des activités économiques et en conséquence l'existence d'un système pluraliste d'assurance maladie, enfin, dans le système de soins de santé, la forte pression exercée par le système judiciaire en matière de réparations des erreurs et des fautes professionnelles. Le Royaume-Uni s'est aussi lancé en 1989 dans une politique de mise en concurrence des structures hospitalières par des groupements de médecins généralistes, avec

Encadré 1. **Le Hospital Report 2001 en Ontario : les indicateurs de soins**

La comparaison des hôpitaux en matière de soins est fondée sur la sélection de dix affections, pour lesquelles on va mesurer quatre types d'indicateurs. Les affections sont les suivantes :

- Infarctus du myocarde
- Accident vasculaire cérébral
- Insuffisance cardiaque
- Pneumonie communautaire
- Asthme
- Hémorragie digestive
- Cholecystectomie
- Hystérectomie
- Prostatectomie
- Libération du canal carpien

Les indicateurs sont les suivants : le nombre de complications rapporté au nombre de cas, les réhospitalisations, la durée de séjour, l'accès à la technologie ou à des soins en hôpital de jour.

la publication régulière de données comparatives sur la performance des hôpitaux. L'Angleterre s'est limitée à la publication d'indicateurs reflétant plus la qualité du service rendu (par exemple les files d'attente) (NHS, 1995) que les résultats cliniques, et ce n'est que depuis un an que le National Health Service rend public des données sur la mortalité hospitalière. Mais en 2001, une initiative privée intitulée Dr Foster¹ a mis à la disposition du grand public sur un site internet des données comparatives des établissements hospitaliers publics et privés, qui n'ont pas manqué de susciter de vifs débats, compte tenu de la méthode d'ajustement des taux de mortalité. L'Écosse publie régulièrement des indicateurs de résultat et de processus de soins (Shaw, 1997). Le Canada publie une enquête annuelle dont l'objectif est de comparer la qualité du service rendu par les services de santé, analysés par zone géographique (Marshall, 2001). Le travail est réalisé par le Canadian Institute for Health Information (CIHI, 2001). Cette enquête annuelle complète une initiative de la province d'Ontario, qui a publié des « Hospital Report Cards » sur 95 établissements provinciaux. En France, ce sont les media qui ont pris l'initiative en 1999 de la publication d'un palmarès des établissements hospitaliers (*Science et Avenir*, 1999). Ce palmarès a depuis été régulièrement repris par d'autres périodiques (*Le Figaro Magazine*, *Le Point*). Depuis, le ministère de la Santé, par l'intermédiaire de sa direction des études, élabore un modèle français de publication des données comparatives sur la performance des hôpitaux (DREES, 1999, 2001), ainsi que l'Agence nationale d'accréditation et d'évaluation des soins (ANAES) (voir encadré 1).

L'objet de ce texte est de rendre compte de ces expériences, en se limitant à la mesure de la performance des structures hospitalières. Ceci exclut les travaux nord-américains fort nombreux qui portent sur l'évaluation, non seulement des producteurs de soins de santé pris individuellement, mais des plans de santé offerts au public par des « Managed Care Organizations ». Dans ce dernier cas, c'est l'ensemble des services offerts par un plan qui fait l'objet d'une évaluation². On se limitera également à la mesure des résultats cliniques des actions de soins, c'est-à-dire aux mesures de la performance des soins en termes de diminution de la morbidité, de mortalité évitée, d'amélioration des paramètres biologiques et cliniques des malades. Ceci exclut l'impact des soins sur la qualité de vie des patients, qui reste encore à un stade très expérimental (Cooper *et al.*, 2001). Sont également exclues les dimensions de qualité de service qui sont plus liées au confort matériel et psychologique du patient, qui peuvent être mesurées soit par des indicateurs spécifiques (longueur des listes d'attente) soit par des enquêtes de satisfaction auprès des patients (Coulter et Cleary, 2001)³. Enfin, on ne prend pas en compte ici les travaux portant sur les méthodes de gestion dites d'amélioration continue de la qualité (Total Quality Improvement) dont l'objectif est de s'assurer en permanence qu'un établissement contrôle les facteurs déterminants de la qualité des soins.

La première partie du texte présentera ce que nous appelons la « technologie » de la mesure de la performance des hôpitaux. Le premier paragraphe rappelle quelles sont les principales caractéristiques des services de soins à l'hôpital, le second paragraphe les conséquences qu'elles induisent pour la mesure de la qualité.

Dans la deuxième partie, on présentera les indicateurs les plus utilisés dans les différentes expériences nationales. On discutera ensuite les limites de certains de ces indicateurs.

La troisième partie examine les utilisations qui sont faites de cette technologie et tente d'en cerner les impacts, à partir des textes publiés dans la littérature. On rappelle d'abord quels sont les objectifs visés par les principaux promoteurs de la production et de la publication des mesures de la performance hospitalière. On examine ensuite l'impact de ces actions en adoptant trois points de vue principaux : celui des patients, celui des acheteurs de soins (assureurs publics ou privés), celui des pouvoirs publics.

La dernière partie tire les leçons de l'analyse pour proposer des recommandations d'action.

1. La technologie de la mesure de la performance hospitalière

1.1. Les caractéristiques des soins hospitaliers

L'hôpital est d'abord une structure de production multi-service. Si l'on utilise les outils les plus répandus de mesure de l'activité hospitalière que sont les systèmes de classification des séjours du type Diagnosis

Related Groups, un hôpital peut « produire » entre 600 et 1 500 types de séjours différents, selon le degré de finesse de la classification étudiée. Cette variété pose donc un problème particulier lorsqu'il s'agit de mesurer non seulement l'activité, mais aussi la productivité et la qualité du service. En effet, il faudrait en toute rigueur disposer de mesures adéquates pour chaque « produit ». Cela signifie aussi qu'un hôpital peut être performant pour certaines activités, et moins pour d'autres. Or, aucune expérience actuelle ne permet de couvrir l'ensemble de l'activité d'un hôpital.

La première solution consiste à choisir un petit nombre d'activités, en général en raison de leur forte médiatisation (la chirurgie cardiaque ou l'obstétrique), de l'importance de leur fréquence assortie d'une gravité faible (une intervention courante ne devrait pas faire courir de risques aux patients), de leur caractère électif (risque fort d'interventions injustifiées), de la population cible (les fractures du col du fémur chez les personnes âgées) ou de la disponibilité de bases de données éprouvées (les soins de réanimation). La deuxième solution consiste à rechercher des indicateurs reflétant de façon agrégée les résultats des soins, comme les taux de mortalité ajustés sur l'éventail des cas (Casemix). Lorsque la première méthode est choisie, on ne pourra pas disposer d'une comparaison complète de la performance des hôpitaux. Pour des patients qui souffrent d'une condition morbide particulière non couverte par les indicateurs retenus, l'information publiée n'a donc aucun intérêt. Pour remédier à ce problème, Eddy (1998) propose de pratiquer un changement systématique de pathologies cibles d'une année sur l'autre. Les données de mortalité pose des problèmes spécifiques qui seront abordés plus loin.

La deuxième caractéristique des soins hospitaliers est qu'ils offrent un service complexe à la personne. Le produit hospitalier n'est pas un séjour, mais une prise en charge d'une personne, de sa maladie, de la dépendance induite par cette maladie et par le traitement. Cette prise en charge associe des inputs multiples ; la combinaison d'inputs retenus doit à la fois s'appuyer sur les meilleures connaissances disponibles, être délivrée de façon productive et s'adapter à chaque cas traité : c'est ce que Minvielle (1996) appelle « gérer la singularité (des patients) à grande échelle ». Dans une relation de service, le patient lui-même est un input du processus : du fait de son état de santé à l'admission, de son patrimoine génétique d'une part, du fait de sa plus ou moins grande collaboration aux soins, d'autre part. Enfin, la qualité de la relation interpersonnelle entre le malade, sa famille et l'équipe soignante contribue également au résultat du processus. Pour cette raison, il y a souvent confusion autour de la notion de qualité des soins, un mélange entre la qualité des soins considérée comme un attribut du résultat du processus (la guérison, la compensation d'un handicap, etc.) et la qualité des soins considérée comme un attribut du processus. Il est de coutume dans la littérature spécialisée dans l'évaluation de la qualité des soins de distinguer les deux notions de qualité des résultats (*outcome measures*) et de qualité du processus (*process measures*). Parmi les mesures de résultats, on va également distinguer les résultats intermédiaires (*intermediary outcomes*), comme le contrôle de la tension artérielle ou du niveau de cholestérol, des résultats dits finaux (*final outcomes*) comme la guérison, la rémission ou le décès. Enfin, le fait que le patient soit à la fois le bénéficiaire du processus et une ressource de celui-ci impose, lorsque l'on compare plusieurs hôpitaux entre eux, de prendre en compte les caractéristiques de la clientèle qui peuvent avoir un impact sur les résultats.

D'un point de vue théorique, il est préférable d'utiliser des indicateurs de résultats : ils mesurent l'impact des soins sur l'état de santé des individus. Ces indicateurs de résultats doivent être spécifiques à chaque activité de l'hôpital. Outre cette complexité liée à la variété de l'activité hospitalière, il existe aussi des limites méthodologiques fortes à la mesure régulière et pertinente du résultat des soins. En premier lieu, l'amélioration finale de l'état de santé peut n'apparaître que bien après l'épisode de soins hospitaliers : cela implique de mettre en place un suivi longitudinal des patients, parfois sur une longue période. Or, les données permettant de faire ce suivi ne sont en général pas sous le contrôle des établissements, et leur collecte requiert la collaboration des patients et d'autres producteurs de soins. En deuxième lieu, lorsque le résultat est mesuré par la non-apparition de complications et si celles-ci sont rares, il est nécessaire de disposer d'observations nombreuses pour être sûr de les mesurer avec une précision suffisante. Enfin, il peut exister de nombreux facteurs de confusion dont il faut s'assurer avant d'imputer à un établissement donné des résultats. Pour ces raisons, et quand il existe suffisamment d'éléments convaincants de preuve sur le lien entre une caractéristique du processus et un résultat, on substitue à une mesure du résultat un

indicateur de qualité du processus de soins. Cependant, il n'existe pas une couverture totale des soins hospitaliers par des études bien contrôlées, et la relation entre process et résultat n'est pas toujours bien établie. De ce fait, le respect d'une norme de process n'offre pas une garantie parfaite de qualité du résultat.

La troisième caractéristique de l'activité hospitalière est le caractère stochastique des soins : il n'y a pas une relation déterministe entre l'administration d'un traitement et son effet sur un malade. Aucun traitement n'a une efficacité de 100 pour cent, aucun traitement ne comporte aucun risque. Autrement dit, l'échec d'un traitement peut être dû à des facteurs qui échappent au contrôle des équipes soignantes, ou au hasard.

La quatrième caractéristique de l'activité soignante à l'hôpital est celle d'une pratique médicale en conditions réelles : contrairement à ce qui se passe dans un essai clinique randomisé bien contrôlé, où les patients sont soigneusement sélectionnés, et à l'issue desquels on sait mesurer une efficacité expérimentale, les établissements doivent adapter les traitements de références à des patients qui ne sont pas ceux des essais cliniques. Or, pour des raisons éthiques (priver les patients d'une thérapie qui a fait la preuve de son efficacité) et pratiques (conduire en permanence des essais randomisés) évidentes, ils ne peuvent pas respecter à 100 pour cent des règles de décision fondées sur des données ayant un haut niveau de preuve dans la littérature publiée. Dans ce cas encore, le respect de recommandations de bonne pratique dérivées de la littérature scientifique ne garantit pas complètement l'efficacité des traitements.

1.2. Conséquences pour la mesure de la qualité des soins

La première conséquence des caractéristiques de l'activité hospitalière sur la mesure de la qualité des soins a déjà été évoquée plus haut : aucun système actuel ne couvre la totalité de l'activité des établissements, et les comparaisons inter-établissements se font sur un nombre limité de pathologies. On peut certes faire l'hypothèse qu'un établissement particulièrement performant sur plusieurs activités l'est dans son ensemble, mais ceci reste à démontrer. Un débat est également ouvert sur la publication de classements agréant la performance des hôpitaux sur plusieurs dimensions, comme la mortalité, la réputation et d'autres facteurs (les « league tables »). La procédure d'agrégation repose nécessairement sur une pondération relative des critères entre eux. En général, le choix de ces pondérations est fait de façon empirique par l'équipe qui publie le classement, mais sans fondement scientifique aucun. Au mieux, les poids relatifs des différentes dimensions devraient refléter les préférences observées de la population générale pour chacune des dimensions. Par ailleurs, l'agrégation masque les différences de performance sur chaque dimension, et peut favoriser relativement les établissements moyens sur l'ensemble des critères. De ce fait, les auteurs recommandent de maintenir la publication de résultats non agrégés (Eddy, 1998), de façon à donner au lecteur une information complète.

La deuxième conséquence est liée aux contraintes qu'impose la méthode statistique pour faire la démonstration rigoureuse d'une part, de l'impact d'un hôpital sur les résultats des soins (identification des facteurs de confusion) et d'autre part, de différences significatives entre des établissements. Dans un cas comme dans l'autre, ces contraintes se traduisent d'abord par des contraintes de taille de populations étudiées, ce qui conduit à ne sélectionner que les activités à fort volume, ou bien à collecter des données sur une période assez longue pour atteindre des effectifs suffisants. On rajoute alors un facteur de variation supplémentaire, le temps, qui permet l'apparition de nouvelles technologies ou de nouvelles pratiques. Par ailleurs, cela conduit à publier des résultats comparatifs avec beaucoup de précautions, par exemple l'indication des intervalles de confiance pour les valeurs publiées des indicateurs. L'incertitude sur les résultats est d'ailleurs la deuxième raison pour laquelle les classements relatifs des établissements entre eux sont critiqués : en fonction de la précision de la mesure, un grand nombre d'établissements peuvent avoir une performance équivalente d'un point de vue statistique et se retrouver pourtant à plusieurs rangs d'écart. Ceci va alors nuire à l'intelligibilité des résultats publiés pour un public qui n'est pas assez averti de la méthode statistique, ou ne va pas prendre le temps d'analyser les résultats avec soin. Cela peut également nuire à la capacité d'action des assureurs, si les résultats ne permettent pas vraiment de discriminer les établissements entre eux.

La troisième conséquence est celle de la qualité des systèmes d'information qui sont requis pour mener à bien de telles analyses. Chaque établissement doit s'assurer que les données qu'il va fournir pour l'analyse comparative sont exactes et complètes. Ceci exige la mise en place de contrôles de la qualité des données en interne et en externe. Les définitions des données doivent être strictement les mêmes d'un établissement à l'autre, au risque de fausser la comparaison. Kritchevsky *et al.* (1999) ont montré comment différentes définitions du taux de césariennes conduisaient à des conclusions variables dans la comparaison entre établissements. Les techniques permettant de contrôler l'impact de facteurs de confusion, par exemple l'ajustement sur les risques des patients pour les taux de mortalité, peuvent requérir des données cliniques très détaillées. Celles-ci sont souvent produites pour la conduite du processus de soins, mais pas toujours systématiquement (elles ne sont pas toutes nécessaires pour la décision médicale) : leur production peut donc exiger un investissement spécifique de la part des hôpitaux soumis à évaluation. En l'absence de dossiers médicaux informatisés et standards d'un établissement à l'autre, il existe donc aussi une troisième limite à l'étendue des activités couvertes par une analyse comparative : celle du coût d'investissement dans un système d'information permettant de couvrir avec un degré de détail élevé l'ensemble de l'activité hospitalière. Un débat est ouvert dans la littérature sur les avantages comparés de bases de données spécifiques, orientées vers la mesure de la qualité des soins sur un segment d'activité précis (par exemple, la comparaison des taux de mortalité pour le pontage coronarien) et des bases de données dites « administratives », c'est-à-dire comportant des données qui sont recueillies régulièrement à des fins autres que l'évaluation de la qualité. Par exemple, les bases de données qui produisent des résumés de sortie (*discharge abstracts*) à des fins de mesure du casemix hospitalier ont été utilisées les premières par Medicare dès 1989 pour publier des taux de mortalité ajustés sur l'éventail des cas. Dans le premier cas, on bénéficie de l'avantage de données spécifiques à un champ de pathologie, donc pertinentes cliniquement. Les résultats seront plus probants, plus faciles à interpréter et plus acceptables pour les cliniciens. Mais cela se fait au prix d'un investissement supplémentaire et d'un champ d'analyse plus réduit. Dans le deuxième cas, en principe, le coût d'investissement initial dans une base de données est faible, le champ couvert est potentiellement large puisque toute l'activité hospitalière est couverte, les données sont produites régulièrement, mais on perd en spécificité dans l'analyse.

2. Les principaux indicateurs

Rappelons qu'il s'agit ici d'indicateurs portant uniquement sur les hôpitaux et qui font l'objet d'une mise à disposition du grand public. Il existe aussi de nombreux travaux d'évaluation des pratiques médicales qui sont publiés dans des revues scientifiques, ou qui servent au monitoring de la qualité des soins au sein des établissements hospitaliers, mais à leur discrétion. On fera la distinction classique entre indicateurs de résultats et indicateurs de *process*.

2.1. Le taux de mortalité comme indicateur de la qualité des soins

C'est *a priori* l'indicateur qui fait l'unanimité : la mort est un événement que tous souhaitent éviter, elle apparaît comme la sanction ultime de soins de mauvaise qualité. Elle a un impact fort sur l'imaginaire de tous et un décès prématuré ou évitable qui semble lié à des dysfonctionnements apparaît particulièrement injuste. C'est aussi un indicateur dont la définition clinique ne souffre pas d'ambiguïté, et qui devrait être repéré avec la plus grande précision dans les établissements hospitaliers, ne serait-ce que pour des contraintes légales (état civil). Depuis les premières publications par Medicare, puis par certains États américains, l'analyse comparative des taux de mortalité a donné lieu à de multiples travaux méthodologiques et continue à être utilisée par les médias dans l'élaboration de leurs palmarès (Ansari *et al.* ; Chalé et Naiditch, 1999 ; Iezzoni, 1997 ; Iezzoni *et al.*, 1995 ; Landon *et al.*, 1996 ; Leyland et Boddy, 1998 ; McKee et Hunter, 1995 ; Thomas et Hofer, 1999).

En fait, la mortalité n'est pas une donnée si facile à recueillir. En premier lieu, beaucoup d'études montrent que les statistiques tenues par les hôpitaux ne sont pas totalement fiables, se traduisant par exemple par des variations importantes et inexplicables dans le temps. Or, comme il s'agit d'un

phénomène rare, des données manquantes ont un impact important sur le résultat. En deuxième lieu, la mortalité intra-hospitalière n'est pas une donnée suffisante pour juger de l'impact des soins et du devenir des patients. D'une part, l'effet final d'une intervention peut se situer au-delà du premier épisode du traitement, et le malade peut décéder plus tardivement à son domicile, ou au cours d'une re-hospitalisation. Un hôpital de soins aigus peut aussi avoir une politique de sortie précoce des malades vers des structures de soins de suite, et la mortalité résultant éventuellement de l'intervention première n'apparaît que dans la deuxième structure. Pour cette raison, il est convenu de mesurer non seulement la mortalité intra-hospitalière, mais aussi la mortalité à 30 jours (ou au-delà si pertinent), ce qui impose aux établissements un suivi des patients après leur sortie. Cette exigence peut alors augmenter l'imprécision des données si le suivi n'est pas totalement fiable.

On retrouve à propos des taux de mortalité les remarques faites plus haut sur les caractéristiques de l'activité hospitalière. Pour la plupart des pathologies traitées à l'hôpital, la mortalité est très faible. Si l'on veut mesurer celle-ci avec une précision satisfaisante pour un hôpital donné, ou pour comparer de façon efficace des hôpitaux entre eux, il est nécessaire de disposer d'une population de départ assez importante. Ceci peut se faire de deux façons : par la sélection de pathologies très fréquentes, par le recueil sur une période dépassant l'année. Mais lorsque l'on travaille sur des pathologies fréquentes, il n'est pas rare qu'elles conduisent également à un pronostic très favorable, et donc à des décès peu fréquents.

Par ailleurs, la littérature sur la mesure des taux de mortalité fait une place importante à la prise en compte des facteurs de confusion, et tout particulièrement aux facteurs de risque présentés par les patients au moment de leur admission. Les établissements peuvent en effet être différents en termes de profils des patients qu'ils recrutent. Les premiers facteurs d'ajustement à prendre en compte sont l'âge et le sexe. En effet, la probabilité naturelle de décès croît avec l'âge, et il y a une différence de mortalité au même âge entre les deux sexes. Mais d'autres facteurs peuvent rentrer aussi en ligne de compte. Pour un même traitement, les patients présentent des profils de risque différents liés à leur état de santé. La nature de ces risques est différente selon les interventions et les pathologies, ce qui requiert de développer des modèles spécifiques, voire des recueils de données spécifiques pour chacune.

Les modèles sont bâtis sur le principe suivant. Une première analyse médicale conduit à identifier des facteurs de risques potentiels dont on va tester l'impact sur la mortalité. Une base multi-centrique de données est ensuite constituée avec des établissements, soit sur une base territoriale (tous les établissements d'un État qui pratiquent une intervention, par exemple), soit sur la base du volontariat, plus rarement sur la base d'une stratification définie a priori. Le traitement des données permet d'identifier les facteurs de risques significatifs et d'estimer leur impact relatif sur la mortalité ou sur la morbidité. On applique ensuite ce modèle aux caractéristiques des patients d'un établissement donné, de façon à calculer des taux de mortalité attendus. On compare ce taux attendu au taux observé dans l'établissement, en faisant le ratio entre les deux. Un taux observé supérieur au taux attendu est le signal d'un éventuel problème de qualité, réciproquement un taux observé inférieur au taux attendu est le signal d'une bonne performance. En général, dans les « league tables » publiés par les médias, ce sont ces ratios taux observé/taux attendu qui sont publiés et qui permettent de classer les établissements les uns par rapport aux autres.

Compte tenu des incertitudes statistiques dont sont entachés les résultats, ces classements sont fallacieux. Rappelons en effet que les taux observés le sont sur la base d'échantillons qui déterminent la précision de la mesure. Les coefficients des modèles statistiques qui servent au calcul des taux attendus sont eux-même estimés avec une précision relative. Cette précision dépend de la taille de la base de données et de sa composition : il peut y avoir une erreur de calibrage si, lorsque l'on teste le même modèle sur une base différente, on n'obtient pas les mêmes résultats, c'est-à-dire les mêmes facteurs de risque et les mêmes coefficients.

Pour ces raisons, les spécialistes remplacent le classement par l'identification d'établissements qui sont significativement différents des autres, les « outliers ». Le terme renvoie à l'idée que l'écart entre taux observé et taux attendu pour certains établissements est trop important pour être dû

uniquement au hasard. Une étude exemplaire de ce type a été réalisée dans les hôpitaux de l'État de New York pour mesurer la mortalité liée au pontage coronarien (Hannan *et al.*, 1990, 1994, 1995 ; Naiditch, 1999).

Malgré ces précautions, la confiance que les auteurs accordent au taux de mortalité (non ajusté et ajusté) pour repérer des problèmes de qualité des soins reste limitée (Ansari *et al.*, 1999 ; Chalé et Naiditch, 1999 ; Iezzoni, 1997 ; Iezzoni *et al.*, 1995 ; Landon *et al.*, 1996 ; Leyland et Boddy, 1998 ; McKee et Hunter, 1995 ; Thomas et Hofer, 1999). Cela est dû à plusieurs raisons. En premier lieu, il existe plusieurs systèmes concurrents d'ajustement pour les facteurs de risques. Iezzoni (1997), Iezzoni *et al.* (1995) et Landon *et al.* (1996) ont comparé la performance de douze de ces systèmes pour quatre pathologies ou traitements : le pontage coronarien, l'infarctus du myocarde, la pneumonie et l'accident vasculaire cérébral non opéré. Si les systèmes ont des performances statistiques globales assez comparables, en revanche, leur comparaison deux à deux donne des écarts qui peuvent être importants, notamment dans le classement relatif des établissements dans les 10 pour cent les meilleurs ou les pires. Autrement dit, le repérage des hôpitaux « outliers » est instable d'une classification à l'autre, même si de façon très globale les classements sont relativement cohérents. Thomas et Hofer (1999) sont arrivés à un résultat identique, mais en utilisant une méthode différente. Ils ont montré que même si l'on disposait d'une méthode parfaite d'ajustement sur le « casemix » et les facteurs de risque, les erreurs systématiques liées à l'identification statistique des « outliers » conduisaient à une valeur prédictive positive assez médiocre de ceux-ci : autrement dit, on trouve un pourcentage non négligeable d'hôpitaux de bonne qualité dans les « outliers » les pires, et réciproquement. Cette valeur prédictive augmente néanmoins avec la taille des échantillons de données procurées par chaque établissement, mais même avec des échantillons de 900 patients par établissement, l'erreur reste mesurable. Silber et Rosenbaum (1997) avaient engagé une critique conceptuelle du lien entre taux de mortalité et qualité de soins. Pour ces auteurs, tous les décès ne sont pas imputables à la qualité des soins. L'état général de certains patients est tel que la mort est inévitable et les soins ne font que la retarder. En revanche, l'équipe soignante doit s'appliquer à prévenir les complications qui vont aggraver le pronostic vital du patient. Le taux de complications survenues en cours d'hospitalisation est alors l'indicateur retenu pour mesurer la qualité des soins, avec deux avantages annexes. Il s'agit d'un phénomène plus fréquent que le décès et requiert donc des échantillons plus petits. Les données sont en principe recueillies de façon systématique dans les résumés de sortie, bien que sur ce dernier point, les auteurs appellent à la plus grande vigilance sur l'exploitation de ce type de données.

Encadré 2. Liste des taux de mortalité les plus souvent étudiés

- Mortalité globale ajustée sur l'éventail des cas traités :
- Mortalité à la suite d'un pontage coronarien
- Mortalité à la suite d'un infarctus du myocarde
- Mortalité suite à une défaillance cardiaque majeure
- Mortalité suite à un accident vasculaire cérébral
- Mortalité des malades atteints de pneumonie
- Mortalité des malades atteints d'une maladie respiratoire obstructive
- Mortalité suite à une hémorragie digestive
- Mortalité des malades opérés pour une prothèse totale de hanche
- Mortalité des malades admis en réanimation
- Mortalité à la naissance des enfants de bas poids

Le taux de mortalité est donc un indicateur de la qualité des soins au premier abord simple et évocateur, mais qui se révèle complexe à calculer de façon rigoureuse à l'usage, et qui est faiblement discriminant entre les établissements une fois tenu compte des biais de mesure. La mesure du taux de mortalité ajusté pour une ou quelques pathologies ne suffit pas pour porter un jugement global sur l'établissement (Rosenthal *et al.*, 1997 ; Iezzoni *et al.*, 1995). Il se prête mal à un classement ordinal du type « league tables ». Enfin, il apparaît incertain qu'il soit un bon facteur prédictif de la qualité des soins, un fort taux de mortalité (resp. un faible taux) n'étant pas toujours synonyme de mauvaise (ou de bonne) qualité des soins. Cependant, le travail qui a été accumulé pour le mesurer de façon rigoureuse a produit une connaissance utile pour les professionnels de deux façons : l'investissement réalisé dans l'étude des dossiers des patients de façon à s'assurer de la qualité des données et à interpréter les résultats, et l'identification des facteurs de risques, ont permis de progresser en faisant baisser les taux de mortalité pour certaines pathologies, comme en témoigne l'expérience de l'État de New York en matière de pontage coronarien (encadré 2).

2.2. Les indicateurs de résultats intermédiaires

Cette notion couvre plusieurs phénomènes. Le résultat d'une partie du processus de prise en charge est un résultat intermédiaire. On parle aussi de résultat intermédiaire lorsque l'on ne sait pas guérir une maladie, mais seulement stabiliser certaines de ses manifestations. En matière de pathologie cancéreuse, la réduction de la taille d'une tumeur est un résultat intermédiaire. On mesure un résultat intermédiaire quand le résultat final ne s'observe qu'au bout d'une période longue par rapport à la durée de l'hospitalisation. Un exemple est celui du retour à la marche dans le cas d'une fracture du col du fémur chez les personnes âgées : le résultat final n'est pas observable à la sortie de l'hôpital qui a réalisé le traitement aigu, mais après la période de rééducation. Enfin, la fréquence d'un événement indésirable évitable est aussi un résultat intermédiaire, comme le taux d'infections nosocomiales, c'est-à-dire acquises pendant le séjour hospitalier, le taux de complications pour des interventions chirurgicales spécifiques ou l'importance de la iatrogénie liée à la prescription de médicaments. La mesure de la fréquence de morbidités évitables couvre une large part des travaux sur les résultats intermédiaires.

Celle-ci a fait l'objet de nombreux travaux dans les années récentes, sous l'appellation d'« outcome research ». Aux États-Unis, la Joint Commission for Accreditation of Health Care Organization publie un document qui recense les outils et méthodes disponibles (le projet ORYX). En revanche, ils sont encore peu utilisés dans les enquêtes comparatives entre les établissements, alors même qu'ils apporteraient une information plus spécifique aux usagers, en identifiant les performances des établissements pour des pathologies différentes. Ceci est lié à plusieurs raisons.

Comme pour l'étude des taux de mortalité, l'analyse comparative des résultats intermédiaires entre établissements requiert des précautions méthodologiques importantes, dont on a déjà présenté les principales : taille d'échantillon, ajustement sur les caractéristiques des patients, qualité des données, nécessité de recourir à des recueils spécifiques de données cliniques ou biologiques. Ceci limite l'exploitation rapide de données comparatives. Deuxièmement, la variété même de l'activité hospitalière est un frein à la généralisation de tels travaux sur un grand nombre d'établissements.

Ceci conduit à privilégier des indicateurs de performance qui sont génériques, et s'appliquent à l'ensemble de la structure. Le taux d'infections nosocomiales présente l'intérêt d'être un indicateur de résultats intermédiaires qui n'est pas spécifique à une pathologie donnée. Il mesure l'incidence de cas d'infections acquises au cours de l'hospitalisation rapportée au nombre de séjours, et sans relation directe avec la maladie qui est à l'origine de celle-ci. La mesure de ce taux est complexe, car elle requiert une documentation très rigoureuse du dossier médical. Les infections nosocomiales peuvent être imputées à des négligences des équipes ou au non-respect de consignes d'hygiène, mais il existe aussi un risque incompressible de transmission d'agents infectieux dans un lieu aussi ouvert que l'hôpital. Enfin, une infection nosocomiale peut être due à une défaillance technique de l'infrastructure qui n'est pas immédiatement décelable, comme dans le cas d'épisodes de légionellose. La documentation précise des cas peut être ressentie comme une forme d'auto-accusation par les équipes

soignantes, qui éviteront d'en rendre compte si l'épisode se termine sans séquelles majeures pour le patient. Ceci implique que les résultats des enquêtes d'incidence ne doivent pas conduire à des sanctions (ou des récompenses) individuelles directes, mais plutôt à la recherche d'amélioration du fonctionnement du service. Dans le cas contraire, les données risquent d'être mal renseignées. Ceci est partiellement contradictoire avec une publication large de cet indicateur auprès du public, qui peut laisser craindre des plaintes des usagers, voire des poursuites en justice.

Les infections nosocomiales sont en fait un bon exemple de la façon dont les établissements doivent rendre compte de la qualité de leur gestion des risques. Une surveillance efficace des infections nosocomiales requiert certes une bonne observance de règles très strictes d'hygiène, mais aussi un dispositif coûteux de dépistage des sources d'infection. Mais cette observance des règles doit s'appuyer sur une confiance forte entre les différents acteurs de l'hôpital, de telle façon que les contrôles que la surveillance exige soient ressentis non pas comme des menaces de sanctions, mais comme une contribution à l'amélioration du travail collectif et des résultats auprès des patients. Dans ce cas seulement, les infections seront bien documentées. Autrement dit, un établissement médiocre pourrait fort bien rapporter des taux faibles (car peu observés) alors qu'un établissement efficace et transparent pourra être confronté à la publication de données peu flatteuses en comparaison. La comparaison de taux d'infections entre établissements requiert donc que l'on s'assure en même temps que chaque hôpital est réellement organisé pour surveiller efficacement ce risque (Humphreys et Emmerson, 1993 ; DREES, 2001).

Le nombre d'accidents thérapeutiques suite à des erreurs de prescription médicamenteuse est un indicateur de résultat intermédiaire qui présente les mêmes caractéristiques que le taux d'infections nosocomiales. La responsabilité des personnes est cependant plus engagée que dans le cas des infections, où l'on ne peut contrôler tous les agents infectieux présents dans un établissement. De ce fait, l'observation de ces accidents est encore plus difficile à mener. Il est considéré comme l'une des principales sources d'erreurs médicales, et son coût estimé est considérable. Mais il n'a pas fait l'objet jusqu'à présent de publications comparatives entre établissements.

2.3. Les indicateurs de process

Les indicateurs de *process* sont descriptifs et évaluatifs du processus de prise en charge. Ils font parfois référence à une norme de pratique – « une prophylaxie des événements thrombo-emboliques est indiquée en chirurgie orthopédique » – et l'indicateur signale alors le taux d'accord sur la norme. Dans le meilleur des cas, la norme a été établie suite à des travaux expérimentaux démontrant la supériorité d'une pratique. Dans de nombreux cas, il n'y a pas nécessairement de preuves expérimentales complètes, mais un faisceau d'éléments. Enfin, la bonne pratique peut être établie selon les dires d'experts. On utilise alors des indicateurs de process comme substitut à la mesure des résultats intermédiaires ou finaux sur lesquels ils ont un impact. Cela permet de contourner le problème de la rareté de certains événements indésirables.

Les processus de prise en charge comportent de nombreuses phases, qui font intervenir un grand nombre d'acteurs. Il est alors nécessaire d'identifier les étapes du processus qui ont l'impact le plus fort sur le résultat des soins. Cette analyse permet de définir des événements sentinelles qui vont signaler l'imminence d'un événement critique dans le processus avec un risque élevé de dégradation de la qualité des soins. L'identification de ces signaux est importante pour le pilotage au quotidien des soins, mais ne peut pas, compte tenu du nombre de ceux-ci, faire l'objet d'une publication systématique mise à la disposition du public. Cette publication serait équivalente, par exemple, à l'affichage public en continu de la température d'un flux d'acier en fusion dans une fonderie, et n'a pas de signification pour les usagers. C'est pourquoi, là encore, la recherche s'est orientée vers des marqueurs de process simples, faciles à recueillir et génériques. Certains sont de nature clinique, d'autres non.

Le taux de césariennes et le taux de naissance par voie basse après un premier accouchement par césarienne sont des indicateurs cliniques de process. On peut d'ailleurs considérer aussi cet indicateur comme un résultat intermédiaire, le résultat final étant l'état de santé de l'enfant et de la mère. On a considéré ici qu'il s'agissait bien d'un indicateur caractéristique du processus de prise en charge,

puisqu'il s'agit d'une modalité d'accouchement. Le taux de césarienne est un indicateur de la qualité des soins pour la raison suivante : la césarienne est une intervention chirurgicale nécessitant une anesthésie, présentant donc plus de risques pour la mère et l'enfant qu'un accouchement naturel. Il existe des indications médicales pour la césarienne, mais la croissance (à des rythmes divers) du nombre de ces interventions dans les pays développés a conduit les professionnels à s'interroger sur leur bien-fondé, sur la justification d'un risque chirurgical et anesthésique. Les études ont montré qu'au-delà des indications justifiées médicalement, la césarienne pouvait être utilisée comme technique de programmation de l'accouchement ; que l'utilisation abusive d'un déclenchement artificiel du travail pouvait également conduire in fine à une césarienne ; enfin, dans certains pays, que la césarienne était utilisée comme moyen très indirect de contraception, fondée sur la croyance qu'une deuxième grossesse après césarienne était dangereuse pour la mère, à cause d'une fragilité de l'utérus liée à la cicatrice laissée par l'intervention. Dans tous les pays développés, la diminution du taux de césarienne a été fixée comme un objectif souhaitable. De plus, la croyance qu'un accouchement par voie basse après une césarienne était dangereuse a été combattue, par des actions de formation et d'information tant auprès des professionnels que des usagers.

Il existe donc une réelle unanimité autour de cet indicateur de performance. Néanmoins, là encore, il est nécessaire de s'entourer de précautions méthodologiques pour mener des comparaisons valides d'un établissement à l'autre. D'une part, les recrutements des centres d'accouchement peuvent être différents du point de vue du profil de risque des mères et donc de la probabilité de recours à la césarienne pour raisons médicales. Il faut donc ajuster les comparaisons en tenant compte de ces facteurs de risque (Aron *et al.*, 1998). D'autre part, Kritschesky *et al.* (1999) ont montré que le calcul du taux de césariennes était dépendant de conventions liées aux caractéristiques des systèmes d'information de chaque établissement, ne permettant pas une comparabilité absolue d'un hôpital à l'autre.

Les indicateurs non cliniques de *process* qui sont les plus fréquemment utilisés sont les durées de séjour, les taux de re-hospitalisation et les volumes d'activité par type d'intervention. Certains périodiques qui publient des « league tables » font l'hypothèse qu'une courte durée de séjour à l'hôpital est synonyme d'une bonne qualité des soins. D'une part, ceci correspondrait au souhait des patients. D'autre part, une courte durée de séjour diminue le risque d'exposition des patients aux erreurs de prescription et aux infections nosocomiales. Cette pression à la diminution de la durée de séjour a aussi un fondement économique car elle a permis une diminution du nombre de lits d'hôpitaux. Mais il faut prendre garde à son interprétation. Une comparaison de durées de séjour doit se faire en ajustant sur différents facteurs : l'éventail des cas traités (le « casemix »), la sévérité de ces cas, et parfois le statut socio-économique des patients. D'autre part, il faut se garder de l'intervention d'une relation causale. Certes, de nouveaux traitements sont apparus qui ont diminué la nécessité d'une hospitalisation longue en augmentant d'ailleurs la qualité des soins, comme les techniques endoscopiques d'exploration et de chirurgie. Dans ce cas, c'est l'amélioration des techniques elles-mêmes qui permettent de diminuer la durée de séjour et la qualité des soins. D'autre part, la durée de séjour est un paramètre manipulable par un établissement. Si celui-ci dispose en aval de places dans des services de convalescence ou de rééducation, la sortie précoce de l'hospitalisation aiguë peut se traduire *in fine* par une convalescence plus longue et une moins bonne qualité des soins. Pour cette raison, la mesure de la durée de séjour doit s'accompagner de la mesure du nombre de sorties considérées comme précoces, et du nombre de re-hospitalisation dans un court laps de temps après un premier séjour (en général 30 jours). En effet, un grand nombre de re-hospitalisations précoces peut être le signe d'un rétablissement insuffisant du patient. Cette mesure n'est elle-même pas complètement satisfaisante : en effet, certains établissements peuvent envoyer leurs malades présentant des problèmes suite à une première intervention vers un autre hôpital.

L'utilisation du volume d'activité par type d'interventions repose sur une idée simple, empruntée à l'économie industrielle : l'expérience cumulée dans l'exécution d'une tâche améliore la performance de l'opérateur. De nombreuses études ont cherché à mettre en évidence cet effet d'expérience dans le système de santé, tout particulièrement pour les actes chirurgicaux ou d'investigation diagnostique (Hillner *et al.*, 2000 ; Nathens *et al.*, 2001 ; Phillips *et al.*, 1995 ; Ritchie *et al.*, 1999). Comme on cherche à montrer l'effet de l'expérience non seulement sur la durée de l'intervention *stricto sensu* ou son coût

(productivité primaire), mais aussi sur la qualité du résultat obtenu, mesurée par exemple par le nombre de complications péri- et post-opératoires ou le décès imputable à l'intervention, il faut s'assurer dans ces études que l'on prend les mêmes précautions méthodologiques que pour les autres travaux de mesure de la performance. En particulier, il faut s'assurer que les patients pris en charge par différents opérateurs sont identiques du point de vue de leur facteur de risque, et que la technique chirurgicale est la même. Certaines études s'intéressent à l'expérience cumulée dans le temps, ce qui soulève alors le problème de l'apparition de techniques nouvelles. Enfin, le niveau d'analyse pertinent n'est pas uniquement l'intervention et l'opérateur, mais le volume global d'activité d'une équipe. Une équipe qui fait peu d'intervention d'un certain type mais qui a une grosse activité globale présente une expérience supérieure à une équipe qui fait peu d'interventions d'un certain type car elle a une activité globale faible. Enfin, l'analyse doit s'appliquer à l'ensemble de l'équipe chirurgicale et pas uniquement à l'opérateur.

Sous réserve de la prise en compte de ces facteurs, cet effet d'expérience a été mis en œuvre pour certaines interventions chirurgicales, comme la prothèse totale de hanche, les interventions chirurgicales sous laparoscopie, la chirurgie cardiaque, la cardiologie interventionnelle, pour ne citer que les principales. De ce fait, le volume d'activité par intervention est souvent utilisé comme critère de *process* dans l'évaluation comparative des établissements, et notamment dans l'élaboration de classements. Cependant, il convient là encore de mettre en garde contre une interprétation trop à la lettre de ces classements. Premièrement, les travaux montrent un lien statistique significatif entre volume et résultats, mais ce lien ne permet pas de passer directement à un classement. Supposons en effet que l'on aboutisse à expliquer 30 pour cent des écarts de performance entre établissement par un effet volume. Il n'en reste pas moins que 70 pour cent des écarts restent à expliquer par d'autres facteurs. Passer directement au classement est source d'identification de faux positifs (des hôpitaux jugés performants à cause de leur volume élevé et qui en fait ne le sont pas), et de faux négatifs (des hôpitaux classés médiocres à cause de volumes plus faibles et qui ne le sont pas). Prenons par exemple le cas d'une équipe chirurgicale réalisant des hauts volume mais au prix d'une cadence élevée de travail du personnel soignant, et de journées de travail longues du chirurgien. Des études ont montré que des équipes chirurgicales proches de la saturation risquaient, compte tenu de la fatigue des intervenants et de la baisse de leur vigilance, d'avoir des résultats de mauvaise qualité. Une autre raison d'éviter des classements ordinaux est que les effets volume ne sont pas continus : un hôpital qui fait 152 interventions par an n'est pas plus performant qu'un hôpital qui n'en fait que 151. Il est donc nécessaire d'identifier des seuils d'activité.

Si l'on se place du point de vue de l'information à donner à l'utilisateur, l'indication du volume ne suffit pas. Il faut également indiquer si l'équipe opératoire respecte des règles en matière de rythme de travail et de temps de récupération. Il faut aussi savoir si d'autres critères de *process* plus spécifiques sont respectés : par exemple, la pratique systématique d'une prophylaxie anti-infectieuse en chirurgie digestive pour éviter des infections post-opératoires. Ces indicateurs de *process* sont généralement étudiés dans les processus d'accréditation. Ceci ouvre donc la question de la publicité donnée à ces rapports, qui contiennent l'information la plus extensive qui existe sur les méthodes de travail d'un établissement.

2.4. Les enseignements

Plusieurs enseignements peuvent être tirés des expériences décrites ci-dessus en matière d'évaluation comparative des hôpitaux. En premier lieu, tous les auteurs insistent sur la question de la qualité des systèmes d'information et du degré de standardisation requis pour s'assurer de la rigueur des études conduites.

L'investissement et la maintenance des bases de données représentent donc un coût qu'il ne faut pas sous-estimer. Il ne faut pas non plus sous-estimer le potentiel de compétences requises pour gérer les bases et pour analyser les données. On a déjà mis l'accent sur les précautions méthodologiques qu'il faut prendre pour tenir compte de facteurs de confusion dans la comparaison inter-établissements. Par conséquent, toute organisation, publique ou privée, qui souhaite développer ce type d'analyse doit

recruter des équipes de statisticiens de bon niveau et s'assurer de la collaboration d'experts cliniciens qui contribuent à l'interprétation des résultats.

Le dernier point concerne la publication de palmarès des établissements, sous la forme d'un classement (les « league tables »). Les médias sont friands de ce type de présentation des résultats, qui a l'avantage d'être simple et spectaculaire et de procurer des bons tirages. Mais il y a de bonnes raisons de penser qu'ils sont faux, au sens où le classement ne reflète pas réellement des différences de qualité de soins entre les établissements. Nous avons évoqué ces raisons : les résultats sont entachés d'incertitude, les indicateurs ne sont pas des mesures directes, univoque, de la qualité des soins. Une autre raison est que ces publications procèdent *in fine* à un calcul permettant de pondérer les différentes dimensions de la qualité entre elles, par exemple en accordant plus de poids à la mortalité qu'à la durée de séjour. Mais cette fonction de pondération n'est fondée sur aucune justification scientifique, politique (le résultat d'un vote) ou économique (enquête marketing auprès des consommateurs pour connaître leurs préférences). Le classement final est donc sensible à ces choix de pondération.

Or cette information est censée aider des usagers à devenir des consommateurs éclairés, en les aidant à choisir les établissements où ils iront se faire soigner. Si cette information est fautive, alors les médias orientent mal ces décisions et pourraient être, en théorie, à l'origine de sévères désagréments rencontrés par les patients. Leur responsabilité pourrait même être mise en cause devant la justice pour fourniture d'informations erronées ayant conduit à dommage à autrui. Cette remise en cause pourrait également venir d'établissements mal classés. A tout le moins, il serait recommandé de publier, non pas des classements, mais des comparaisons par groupe d'établissements ayant des niveaux de performance statistiquement comparables.

3. L'impact des expériences d'évaluation comparative

On analysera cet impact du point de vue de trois acteurs : les fournisseurs de soins eux-mêmes, les usagers, les payeurs. En fait, il existe peu d'étude d'évaluation de cet impact des comparaisons hospitalières.

3.1. Impact sur les fournisseurs de soins

L'impact attendu de la publicité des données sur les fournisseurs de soins est celui de l'amélioration des résultats de chacun. Cette amélioration dépend elle-même de la structure des incitations contenues dans le programme à l'origine. Dans un modèle concurrentiel classique, la mise à la disposition des informations relatives à la qualité des produits n'entraîne pas nécessairement chez les producteurs de basse qualité une incitation à une amélioration de celle-ci. En effet, ils peuvent faire l'hypothèse que certains consommateurs font un arbitrage prix-qualité et seront prêts à dépenser moins. Dans ce cas, il existe pour ces producteurs une place sur le marché. Tout au plus peut-il exister une réglementation définissant un seuil de qualité au-dessous duquel il n'est pas possible de descendre. Pour quelles raisons un producteur va-t-il améliorer sa qualité ? Il le fera s'il pense pouvoir rentabiliser son investissement par une part de marché et un profit accru.

Dans le secteur de soins, on devrait donc s'attendre à ce que la publication comparative de données portant sur la qualité des soins n'ait d'impact en termes d'amélioration de la qualité que sur les fournisseurs de soins cherchant à accroître leur part de marché dans une perspective de recettes ou de profits accrus.

L'activité de soins de santé présente cependant certaines spécificités. La première est que l'élasticité prix/qualité des usagers est forte, c'est-à-dire que les usagers accordent une valeur importante à la qualité et sont prêts à payer beaucoup plus cher pour avoir une garantie forte de son niveau⁴. Autrement dit, la publication de résultats comparatifs pourrait provoquer des mouvements importants de clientèle entre les établissements (ou, dans le cas de systèmes à forte pression médico-légale, comme aux États-Unis, la crainte de poursuites judiciaires accrues pourrait inciter les établissements les plus mal classés à améliorer leurs résultats). Il suffit que les établissements

perçoivent que cette menace est forte pour qu'ils soient incités à investir dans la qualité. Mais la perception de cette menace par un établissement est tempérée, pour les soins hospitaliers, par la question de l'accessibilité géographique (quelle distance est-on prêt à couvrir pour obtenir une qualité supérieure de soins) et le degré de concurrence dans une zone de chalandise (un hôpital en position de monopole local peut n'avoir aucune incitation à changer). Enfin, pour un établissement donné, on peut faire l'hypothèse que la décision de mener une action est fonction du coût perçu de cette action par rapport aux résultats attendus, et qu'elle est aussi fonction de la distance de l'établissement par rapport à la moyenne des hôpitaux auxquels il est comparé. Si cet écart est faible, l'incitation à agir risque d'être faible, l'hôpital pouvant mettre l'écart sur le compte des erreurs de mesure. Pour les établissements au-dessus de la moyenne, on peut également faire l'hypothèse qu'ils ont une incitation faible à renforcer leur effort pour se rapprocher des meilleurs si le coût perçu de cet effort est élevé.

Une deuxième spécificité du secteur des soins de santé est l'intervention d'un tiers payeur, l'assureur, public ou privé. La négociation du rapport qualité/prix se fait entre l'assureur et le producteur de soins. Deux cas de figure se présentent pour les assureurs privés en situation de concurrence. Soit l'assureur juge que le rapport qualité/prix est satisfaisant (par exemple la qualité et les prix sont bas). Dans ce cas, il n'y a d'incitation à changer que si les assurés manifestent leur mécontentement. En revanche, si l'assureur juge le rapport qualité/prix défavorable, il peut menacer l'établissement de lui retirer son agrément ou de baisser le prix. Cette menace est d'autant plus crédible qu'il est un financeur dominant de l'établissement, ou que tous les assureurs adoptent la même stratégie. Dans ce cas, l'établissement sera forcé de mettre en place une action d'amélioration de la qualité des soins.

Dans le cas d'un assureur public, on fera l'hypothèse qu'il est en situation de monopole. Il est alors difficile politiquement de ne pas corriger des différences trop importantes de qualité des soins entre les établissements qu'il dirige. Quatre cas de figures peuvent se présenter :

Coût/Qualité	Basse	Haute
Bas	1	2
Haut	3	4

Le payeur public observe des établissements 1) de basse qualité et de coût unitaire bas ; 2) de haute qualité et de coût unitaire bas ; 3) de basse qualité et de haut coût unitaire ; 4) de haute qualité et de haut coût unitaire. En 1), il doit en principe financer l'établissement en assortissant le financement d'obligations de résultats en terme de qualité. En 2), il observe une population d'établissements de référence. En 3), il doit éventuellement pénaliser financièrement ces établissements tout en leur demandant d'améliorer leur qualité. En 4), il doit demander à ces établissements de baisser leurs coûts unitaires en maintenant leur qualité. Il n'y a donc d'incitations pour un établissement à mener une action que si le payeur a les moyens techniques de mener une analyse qualité/coût unitaire, et les moyens politiques de la mettre en œuvre. En effet, cela passe par un redéploiement des ressources entre les établissements, et ceux-ci peuvent mobiliser leurs ressources politiques locales pour infléchir les décisions prises.

En résumé, pour un établissement dont les résultats sont médiocres, l'incitation à agir provient *a)* de la crainte d'une image négative se traduisant à terme par une perte de recrutement ; *b)* de la distance de son niveau de qualité par rapport à la moyenne des établissements ; *c)* cette crainte peut être tempérée à court terme par la position relative de cet établissement dans son environnement concurrentiel local ; *d)* elle dépend également de la pression exercée par les assureurs ; *e)* celle-ci est fonction de la perception qu'ils ont du rapport qualité/prix.

Ceci veut dire que l'on peut s'attendre à ce que la réaction à la publication de résultats dépende de la situation particulière de chaque établissement, et qu'il soit difficile de prédire une tendance générale. Ce raisonnement vaut cependant à un instant donné. Si la publication des résultats est répétée, démontrant que certains établissements ne font aucun effort, il est probable qu'à terme

ceux-ci seront obligés de mener des actions correctrices. En matière de mesure de l'impact de l'évaluation comparative, cela signifie qu'il faut maintenir l'action sur la longue durée. Dans le cas d'un payeur public, néanmoins, la responsabilité de la correction des écarts appartient à celui-ci, surtout dans le cas d'établissements à basse qualité et à bas coût unitaire.

La littérature conforte-t-elle ces prédictions ? Elle est trop fragmentaire pour permettre des conclusions définitives, mais elle livre quelques indications. Les expériences qui sont citées et qui ont donné lieu à évaluation en termes d'impact sur le comportement des fournisseurs de soins sont principalement des expériences fondées sur la participation volontaire des établissements, qui financent une partie du dispositif. On est donc dans un cas où les établissements estiment *a priori* que la connaissance de leur performance relative est un bénéfice suffisant pour compenser le risque d'une image négative, les consommateurs et les payeurs leur accordant le crédit d'avoir pris le risque d'apparaître comme peu performants. Ils sont dans une logique où ils doivent aller au bout de leurs efforts : si leurs résultats sont médiocres, ils doivent agir pour ne pas perdre la confiance placée en eux par leurs publics de référence.

Une des premières expériences de grande envergure conduites aux États-Unis a été la publication régulière des taux de mortalité ajustés sur l'éventail des cas par le Département de la Santé de l'État de Pennsylvanie. Cette expérience a été assez loin, puisqu'elle a comporté la publication de résultats par médecin. Commencée en 1989, elle a été interrompue en 1993, après de nombreuses controverses sur la qualité des ajustements réalisés et sur la pertinence d'utiliser la mortalité comme indicateur de la performance.

L'exemple le plus fameux à de multiples égards est celui de l'étude de la mortalité après pontage coronarien dans l'État de New York (Hannan *et al.*, 1990, 1994, 1995 ; Chassin *et al.*, 1996 ; Naiditch, 1999). Celle-ci est exemplaire à plusieurs titres. Elle a associé tous les établissements de l'État de New York pratiquant des pontages coronariens. Elle a été financée par le Département de la Santé de l'État, qui s'est engagé sur un financement de longue durée. L'étude a duré de 1990 à 1997. Elle a conduit à un recueil spécifique de données cliniques avec un suivi des patients sur la longue durée, et la qualité des données a fait l'objet d'un monitoring régulier. Ceci a permis de faire un ajustement des résultats sur les risques particulièrement performant. Elle s'est accompagnée de réunions régulières entre les équipes pour analyser les résultats. Les résultats n'ont pas été présentés sous la forme d'un palmarès, mais par l'identification de trois groupes d'établissements : ceux avec des résultats supérieurs à la moyenne, ceux avec des résultats autour de la moyenne, enfin ceux avec des résultats inférieurs à la moyenne. Dans un premier temps, pour favoriser l'adhésion des équipes chirurgicales au projet, il avait été convenu que les résultats ne seraient pas publiés tant que l'étude ne serait pas complètement aboutie. Cette clause a néanmoins été remise en cause en cours de route, à la suite d'une action en justice conduite par un journaliste au nom de l'accès à l'information. Suite à la première publication des résultats, la mortalité a baissé de 41 pour cent, alors qu'elle ne baissait que de 18 pour cent dans le reste du pays et que la gravité des patients augmentait. Une partie de cette amélioration a été due à la sortie du marché des équipes les moins performantes, en général celles dont les volumes d'activité étaient les plus faibles, mais l'étude systématique des méthodes de travail des plus performantes a également permis de faire progresser les autres équipes et de mieux connaître les facteurs de succès du traitement.

Une autre expérience de grande envergure – le Cleveland Health Quality Choice (CHQC) en Ohio – a donné lieu à de nombreux travaux publiés (Rosenthal *et al.*, 1998). Ce projet date de 1988 et résulte d'un effort conjoint de plusieurs groupes d'acteurs : association hospitalière, association de médecins, d'employeurs, chambre de commerce et d'industrie. Il rassemble les données provenant de 30 établissements et couvre les domaines suivants : la mortalité associée à huit pathologies, la durée de séjour rapportée à 13 pathologies, la mortalité et la durée de séjour en réanimation, le taux de césariennes et le taux de naissance par voie basse après césarienne, des indicateurs de iatrogenèse hospitalière et une enquête de satisfaction auprès des patients. Les travaux publiés sur l'impact de ce programme démontrent la diminution du taux de césariennes, l'augmentation du taux d'accouchement par voie basse après une césarienne (Caron et Neuhauser, 1999), la diminution du taux de mortalité

après réanimation accompagnée d'une diminution de la durée de séjour, et ce malgré l'alourdissement de la clientèle (Sirio *et al.*, 1994). Rosenthal *et al.* (1998) citent quatre études de cas d'établissements participant à l'expérience et ayant mis en œuvre des actions d'amélioration des résultats dans trois domaines : les durées de séjour pour un pontage coronarien, la mortalité associée à la pneumonie, le taux de césariennes et le taux d'accouchement par voie basse après césarienne. Une évaluation externe de l'expérience de Cleveland a été publiée par Rosenthal *et al.* (1997) Ces auteurs ont étudié le taux de mortalité pour les huit pathologies retenues à l'origine de l'expérience et ont mené une étude statistique comparant ces taux avant la mise en place du CHQC et après. Ils ont conclu à une diminution significative de la mortalité ajustée sur les risques pour les huit pathologies pour un total cumulé de 13 pour cent sur deux ans, avec un résultat de 20 pour cent pour l'insuffisance cardiaque, les hémorragies digestives, la colectomie et le pontage coronarien. Compte tenu de la nature des données, cette analyse ne permet cependant pas d'en inférer de façon certaine que cette tendance est attribuable au projet. Une expérience très similaire a été mise en place en Californie, intitulée California Hospital Outcomes Project (Romano *et al.*, 1995). Ce projet s'est limité aux trois problèmes suivant : la mortalité après infarctus du myocarde, la chirurgie du disque vertébral et le taux de césariennes. Enfin, Epstein et Kurtzig mentionnent qu'en 1994, 38 États américains avaient constitué des bases de données permettant d'évaluer la performance des hôpitaux (Epstein et Kurtzig, 1994).

L'expérience de Cleveland est instructive à plusieurs titres. Fondée au départ sur le volontariat et un co-pilotage par les partenaires, avec un principe de décision collective, elle a connu des tensions multiples liées à la diffusion et à l'utilisation des résultats. Un partenaire ayant utilisé les résultats publiés pour renégocier des contrats avec des établissements, mais sans en faire part aux autres membres, les établissements ont menacé de se retirer du dispositif s'il n'y avait pas concertation avant toute utilisation des données à des fins de prise de décision. Deuxièmement, le programme s'est limité à la production d'informations comparatives, mais la mise en place d'actions d'amélioration a été laissée à l'initiative des établissements membres. Le groupe de projet n'est pas intervenu pour éventuellement aider les établissements à mener ces actions. Troisièmement, 75 Pour cent du financement de l'ensemble reste à la charge des établissements, avec une estimation d'environ 3 millions de dollars par an en 1998 pour 30 établissements, soit 100 000 dollars environ par établissement.

Au Royaume-Uni, Bucknall *et al.* (2000) ont montré que la mise en place d'un système d'audit simple sur les pratiques de prise en charge de l'asthme avait eu pour impact une augmentation du suivi des recommandations de bonnes pratiques.

En conclusion, les données publiées suggèrent que la constitution de bases de données comparatives sur les performances hospitalières, lorsqu'elles sont construites sur des principes méthodologiques rigoureux, sur un partenariat et en concertation avec les acteurs hospitaliers, permet la production d'une connaissance très utile pour améliorer ces performances. La publicité des résultats est un point de passage obligé, mais la menace potentielle qu'elle représente pour les établissements qui jouent le jeu doit être tempérée par des règles pré-déterminées d'utilisation de l'information produite, ainsi que du format de publication. Les auteurs soulignent là encore la lourdeur du processus de production de données de qualité. On peut identifier des actions d'amélioration conduites par des établissements à la suite de la publication des résultats, avec des effets positifs sur la performance, mais celles-ci ne sont pas systématiques dans tous les établissements. Il semblerait cependant qu'il y ait un effet global, mais il est difficile de l'attribuer au seul facteur de l'analyse comparative. Il est probable que ces améliorations sont aussi le fruit d'un environnement externe plus exigeant à l'égard des établissements.

3.2. Impact sur les usagers

La publication des données de performance relative des hôpitaux peut avoir un impact si les usagers utilisent ces informations pour choisir le lieu de leur prise en charge. Du point de vue de la théorie économique, cette publicité est justifiée pour réduire l'asymétrie d'information entre prestataires de services et bénéficiaires. Du point de vue de l'action publique, cette publicité est souhaitée au nom de la nécessaire transparence dans le fonctionnement de services publics. Chacun

de ses principes est difficilement attaquant. D'ailleurs, les patients atteints de pathologies chroniques (insuffisants rénaux, patients séropositifs) ont fait la démonstration qu'ils étaient capables d'acquiescer une véritable expertise profane sur leur maladie et les soins dont ils ont besoin : ils expriment réellement une demande autonome de services de santé. Mais ces patients experts ne sont pas la règle, la plupart du temps le recours aux soins se faisant de façon ponctuelle. La question qui se pose alors est : quelle est la meilleure façon de diffuser cette information de façon à ce qu'elle soit réellement utile aux bénéficiaires des services et qu'elle ne les conduise pas à des jugements erronés ? De ce point de vue, les débats autour de la publication de palmarès prennent un relief particulier. Le palmarès est une représentation très simple de la performance relative des établissements. Mais nous avons vu que le classement des établissements, en fonction des taux de mortalité par exemple, n'a aucune valeur scientifique et transmet en fait une information fautive. La « bonne » information est celle qui regroupe les établissements ayant des performances identiques à un intervalle de confiance près, mais même ce type de regroupement n'est pas exempt d'erreurs. Autrement dit, choisir nommément un établissement dans une liste parce qu'il est le premier n'offre aucune garantie qu'il donnera une meilleure qualité de soins que le deuxième, le troisième, etc. Il est même possible qu'il présente un risque de mauvaise qualité. Cette deuxième information « vraie » peut être cependant trop complexe pour être acceptable, elle peut même induire un sentiment de méfiance d'un lecteur profane à l'égard des travaux des chercheurs.

Les décisions pour lesquelles ces informations devraient être utiles sont les suivantes :

- Dans le cas où une personne souffre d'une maladie, quel est le traitement adapté à sa condition ?
- Quels sont les prestataires de soins qui offrent ces traitements, et avec quel niveau de qualité ?

Nous avons trouvé peu de travaux de recherche sur ce thème. D'une part, on ne connaît pas l'impact premier de la publication de données comparatives sur le public : combien de personnes les lisent, avec quel degré d'attention, quelle crédibilité accordent-elles aux informations données. Voelker (1995) signale que les données des enquêtes d'accréditation réalisées par la JCAHO sont disponibles pour la presse et le grand public, mais qu'elles sont fort peu consultées. Si ces informations sont acquises, on ne sait pas le rôle qu'elles jouent éventuellement dans la prise de décision des personnes. Par exemple, on ne connaît pas bien l'impact de la publication d'un palmarès hospitalier sur les modifications de recrutement d'un centre hospitalier. En revanche, Hibbard *et al.* (1996) ont mené des travaux dont l'objectif était 1) d'évaluer la compréhension par des personnes des indicateurs de qualité 2) d'élaborer les modes de communication les plus efficaces des résultats. Leurs travaux visaient à évaluer l'impact des informations comparatives pour faciliter le choix d'un contrat d'assurance maladie. Leurs résultats sont les suivants.

En premier lieu, ces auteurs rappellent que les informations comparatives publiées ne sont pas les seules disponibles pour les bénéficiaires des services. L'expérience passée du recours à un établissement, le bouche à oreille, l'avis du médecin traitant, les enquêtes publiées par des associations de malades ou de consommateurs et les autres sources d'information proposées par les médias sont autant de ressources cognitives concurrentes ou complémentaires. Mesurer l'impact des données publiées nécessite donc de recourir à des dispositifs d'études complexes, au cours desquels on observerait le comportement de décision des personnes en situation « réaliste », soit au cours d'expérimentation, soit au cours d'entretiens. Mais de tels travaux n'ont pas encore été réalisés.

Jewett et Hibbard (1996) soulignent la complexité de l'information relative à la qualité des soins hospitaliers. En utilisant la méthode des « focus groups », au sein desquels il est demandé à un nombre limité de personnes de porter des jugements sur des propositions ou de répondre à des questions, le groupe ayant pour objet la facilitation de la prise de parole et une plus grande explicitation des réponses que lors d'une enquête par questionnaires, Jewett et Hibbard ont trouvé qu'une part importante des personnes avaient une mauvaise compréhension de la signification des indicateurs, soit qu'elles en aient une interprétation erronée, soit qu'elles ne les comprennent pas du tout. Ceci était particulièrement sensible pour les indicateurs calculant un taux d'événements indésirables. Dans ces conditions, il faut se demander comment présenter des informations fiables, c'est-à-dire en indiquant leur degré de précision, pour qu'elles ne soient pas rejetées. Hibbard *et al.* (1997) suggèrent qu'un travail d'explication et de

contextualisation est nécessaire. Il faut expliquer aux personnes destinataires de l'information pourquoi un indicateur a été choisi, en quoi il mesure réellement la qualité des soins, à quelles questions il permet de répondre. Cet exercice peut cependant être rendu difficile, surtout si les indicateurs retenus sont publiés avec les précisions statistiques requises par une bonne méthodologie. Enfin, les auteurs soulignent que la sophistication relative des informations publiées les rend inaccessibles à un public ayant un niveau d'éducation trop bas.

L'agrégation des différentes dimensions de la performance pose aussi problème. Dans les travaux publiés dans la grande presse, cette question est résolue par le choix d'une pondération « maison » de ces différentes dimensions. Mais si cette agrégation n'est pas faite, cela signifie que c'est au lecteur de réaliser ses propres arbitrages entre plusieurs dimensions de la qualité. Or, selon Hibbard *et al.* (1997), les travaux empiriques portant sur la décision montrent que les individus ont beaucoup de difficultés à réaliser des arbitrages entre des dimensions conflictuelles. Dans ce cas, elles tentent de construire des cadres d'analyse où une dimension l'emporte sur l'autre, au risque de sous-estimer ou de négliger des informations importantes.

Pour toutes ces raisons, dans le but d'améliorer l'information donnée aux malades et son utilité dans la décision, Hibbard *et al.* (1997) et Sisk (1998) proposent de donner un rôle important à des acteurs médiateurs qui auraient trois types de compétences. D'abord, la connaissance des questions que se posent les personnes malades devant prendre une décision d'hospitalisation pour une pathologie donnée : il est probable que ces questions auront trait au déroulement de la prise en charge autant qu'aux risques associés ; deuxièmement, la connaissance de la façon dont les structures hospitalières gèrent concrètement ces problèmes ; ceci passe alors par l'accroissement des études fines de *process*, établissement par établissement ; enfin, la connaissance de l'apport et des limites des indicateurs de résultats disponibles conduisant à une capacité d'explication et de contextualisation. Les associations de malades pourraient jouer ce rôle, ainsi que les médecins traitant des personnes, à condition qu'ils aient un accès à toutes ces informations complémentaires. Rendre publiques des données comparatives ne serait pas alors uniquement publier des tableaux de résultats dans la presse, mais faciliter l'accès des personnes à un réseau d'intermédiaires agréés qui seraient au service des personnes.

Enfin, l'accent est mis sur la crédibilité des informations produites. Cette crédibilité recouvre trois dimensions : celle de l'indépendance des producteurs de l'information comparative, celle de la fiabilité des données (liée au degré de standardisation du calcul des indicateurs) et celle de l'exactitude (la qualité des systèmes d'information). Pour labelliser cette crédibilité, devant la profusion des initiatives prises en matière de publication de résultats comparatifs entre les fournisseurs de soins et les plans d'assurance maladie aux États-Unis, une instance d'évaluation et d'accréditation, la Consumer Assessment of Health Plans Study, a été créée en 1996.

3.3. Impact sur les payeurs

Ce point a déjà été discuté partiellement à propos des fournisseurs de soins. En principe, la connaissance de la performance relative des hôpitaux sur la qualité des soins devrait avoir un impact sur le financement de ceux-ci, en fonction du rapport qualité/coût recherché par le payeur. On peut imaginer trois décisions liées à la prise en compte de la performance relative d'un fournisseur de soins : la décision d'adresser ou non des assurés à ce fournisseur ; la décision de contrat sélectif avec un fournisseur, le payeur sélectionnant une gamme de traitement parmi toute la gamme offerte ; enfin, une négociation prix-qualité. Dans cette dernière négociation, un fournisseur de soins de qualité supérieure à la moyenne pourrait également demander un prix supérieur. À l'inverse, un payeur pourrait refuser le prix demandé par un fournisseur ne présentant pas le niveau de soins requis par le premier. Ce dernier arbitrage nécessite de pouvoir réaliser simultanément une comparaison prix-qualité entre fournisseurs de soins.

En général, la première décision est prise sur la base d'un processus d'accréditation confié à une agence indépendante. L'accréditation est en principe la pierre angulaire des dispositifs de mesure de la qualité des soins. Les informations détaillées relatives à un établissement donné sont mises à la

disposition des payeurs potentiels et pour une partie d'entre elles auprès du public. Sans doute à cause de leur technicité et de la longueur des rapports, cette information publique reçoit moins d'attention que les palmarès. Le niveau d'accréditation d'un fournisseur de soins est donc un paramètre essentiel de sa relation avec un fournisseur. Mais dans un pays où l'accréditation est en vigueur depuis longtemps (comme aux États-Unis), presque tous les établissements sont accrédités avec un niveau satisfaisant et cela ne permet plus de les discriminer les uns par rapport aux autres.

Il existe de nombreux exemples de contractualisation sélective entre payeurs et établissements : si le niveau de performance pour un segment d'activité n'est pas satisfaisant, le payeur refuse de lui adresser des assurés ayant besoin desdits traitements. Dans les systèmes de soins à fort secteur public, ceci se traduit par des procédures d'agrément sélectif : seuls les établissements rencontrant un certain nombre de critères (en général de *process*) pourront pratiquer tel segment d'activité. Les segments les plus souvent cités sont la chirurgie cardiaque, les centres de cardiologie interventionnelle, les centres de greffes d'organes, les maternités. En dehors des États-Unis, ces procédures d'agrément sont fondées avant tout sur des critères de *process* et non sur des indicateurs de résultats.

Le coût reste un critère dominant de la relation entre payeur et fournisseur, en particulier dans le cas de systèmes concurrentiels d'assurance maladie. Mais aux États-Unis, où cette concurrence est vive, la qualité des plans de santé offerts aux assureurs est devenu un argument commercial important auprès des assurés et de leurs employeurs, qui financent en général ces programmes. Au sein de systèmes de santé où le secteur public joue un rôle important, la qualité et la sécurité des soins sont devenues un enjeu politique important pour les citoyens. Autrement dit, il paraît inévitable que les négociations entre payeurs et fournisseurs fassent une plus large part qu'auparavant à cette dimension de la performance qualitative des soins. Les travaux publiés sur ce thème ont eu pour but principal d'évaluer si la pression sur les prix s'est traduite par une dégradation de la qualité pour les assurés. En revanche, compte tenu du caractère privé de ces négociations, il n'y a pas de travaux publiés permettant de savoir comment les acteurs économiques réalisent des arbitrages entre qualité et coûts.

On peut formuler les hypothèses suivantes. Premièrement, on suppose que le coût unitaire d'une prestation est une fonction croissante de sa qualité : c'est la croyance dominante, la qualité coûte cher. Dans un système où les payeurs sont en concurrence, et mettent également les fournisseurs en concurrence, les premiers ont à leur disposition une information incomplète sur la qualité relative des seconds. Ils ont la garantie de base de l'accréditation. Ils peuvent également utiliser le contrat sélectif. Mais les indicateurs publiés ne couvrent qu'une partie de l'activité des fournisseurs. Ils doivent donc faire l'hypothèse que si un hôpital est bon sur une partie de son activité, il y a une bonne probabilité qu'il le soit aussi sur tout. D'autre part, les indicateurs sont faiblement discriminants entre les fournisseurs : les palmarès sont, selon les normes vigueur, faux. Les fournisseurs peuvent être classés en trois (ou cinq) groupes : les très performants, la moyenne acceptable, les peu performants. Le calcul économique qui établirait pour un payeur l'arbitrage entre une unité marginale de qualité supplémentaire et le coût unitaire de la prestation n'est pas facile à faire sans une étude spécifique de chaque prestataire. De surcroît, il faut attendre deux ou trois années consécutives pour confirmer un premier classement. Enfin, les payeurs agissent sous contrainte géographique : dans une zone donnée, il n'y a qu'un nombre déterminé de fournisseurs. Il n'est pas possible de choisir un prestataire en dehors de la zone de résidence des assurés. Dans ces conditions, il est probable que les payeurs fonctionnent en évitant le plus possible les établissements les moins performants, et en faisant jouer la concurrence sur les coûts (et les prix) entre les établissements dans la moyenne. En revanche, il est probable que ces établissements ne sont pas en position de force face au petit nombre d'établissements très performants. Il restera donc une certaine hétérogénéité de la qualité des soins.

Dans un système d'assurance maladie universel avec un payeur unique, celui-ci affronte les mêmes difficultés pour discriminer finement les fournisseurs en fonction de leur performance et du ratio qualité-coût. Cependant, le payeur doit opérer sous contrainte d'égalité en matière de qualité des soins pour tous les fournisseurs. Dans l'hypothèse où le coût est une fonction croissante de la qualité, alors il est de sa responsabilité d'augmenter les ressources disponibles aux fournisseurs les moins

performants pour les rapprocher des autres établissements. Mais si les résultats sont publiés, le payeur risque d'être soumis à une forte pression des assurés-voteurs pour aligner la qualité vers le haut, et donc augmenter pour tous le niveau de ressources disponibles.

Dans les deux cas, la qualité moyenne augmente ainsi que les coûts. Dans le premier cas, les fournisseurs peu performants sortent du marché s'ils ne se corrigent pas, dans le deuxième cas la puissance publique garantit à tous les fournisseurs les moyens d'aboutir à la performance la meilleure. Cependant, on peut douter de ce scénario. En effet, l'hypothèse « la qualité coûte cher » est peut-être vraie une fois que l'ensemble des fournisseurs a atteint le même stade de performance, mais ceci est loin d'être le cas aujourd'hui. Autrement dit, il est possible que les fournisseurs augmentent leurs performances en baissant leurs coûts unitaires. Par exemple, de nombreuses études montrent qu'il y a une relation positive entre volume d'activité et résultats cliniques. Dans le cas de l'activité hospitalière, à fort coûts fixes, les coûts unitaires baissent aussi en fonction du volume. Adresser des patients à des établissements à fort volume permet donc à la fois d'améliorer les résultats cliniques et de baisser les coûts. Le payeur ne peut donc réaliser un arbitrage qualité-coût que s'il est en mesure de connaître, activité par activité, la relation entre les deux. Or, on a déjà souligné que seule une petite partie de l'activité hospitalière est aujourd'hui couverte par des indicateurs de performance, ce qui signifie que l'on est sans doute loin d'avoir réalisé le potentiel d'amélioration des résultats à moyens identiques. Une autre difficulté s'ajoute : les assurés d'un payeur donné ne présentent pas nécessairement le même profil que la moyenne des patients d'un hôpital donné. Or, les indicateurs de performance portent sur l'ensemble de ces patients. Il peut donc y avoir un biais d'estimation de la performance, dans un sens ou dans un autre selon que les assurés du payeur en question sont plus ou moins à risque.

Il reste donc une forte asymétrie d'information entre assureurs et fournisseurs au détriment des premiers sur la qualité des soins, et ce malgré la production d'indicateurs comparatifs. Il faut que le payeur trouve les incitations qui pousseront le fournisseur de soins à engager des actions visibles pour améliorer les résultats et partager les éventuels gains de productivité avec le payeur. Enfin, pour observer ces efforts, il faut que le fournisseur développe son système d'information au-delà de la production des indicateurs les plus usités aujourd'hui. Par voie de conséquence, cela devrait augmenter les coûts de transaction entre les deux acteurs.

On comprend dès lors pourquoi les études dont l'objectif est de mesurer l'impact de l'intensité de la concurrence sur la qualité des soins donnent des résultats contradictoires, nonobstant les difficultés méthodologiques de mesure de cette concurrence (MacLaughlin et Ginsburg, 1998). D'une part, il n'est pas certain que le payeur puisse réellement observer l'effort fait par le fournisseur pour améliorer sa performance. Dans ce cas, il y a peut-être diminution des coûts avec dégradation de la performance sans que le payeur ne puisse l'observer, ou report des coûts de la qualité sur un autre payeur ou segment d'activité pour lequel la performance n'est pas mesurée, ou encore captation des gains de coût liés à l'amélioration de la performance sans partage avec le payeur. D'autre part, l'intensité de la concurrence risque d'être variable en fonction du segment d'activité de chaque fournisseur.

3.4. Synthèse

Les enseignements que l'on peut tirer de l'impact de la publicité donnée à des indicateurs de performance hospitalière nous viennent avant tout des États-Unis, où cette publicité s'inscrit dans un contexte de concurrence entre payeurs auprès des assurés, et entre fournisseurs auprès des payeurs. Ces expériences sont par ailleurs fondées sur le principe du consommateur informé, à un point tel qu'une initiative, la Consumer Assessment of Health Plans Study, vient garantir la qualité de l'information produite par les différents plans d'assurance maladie offerts. Le poids du consumérisme se traduit également aux États-Unis par le rôle joué par des associations de malades ou de consommateurs dans la publication de travaux comparatifs entre fournisseurs. Ce climat général n'est pas le même en Angleterre, où il existe des « leagues tables » depuis 1993, mais où la publication de données de mortalité n'est intervenue qu'en 2000, ou en France, où pour l'instant la publication de données comparatives de performance a commencé en 1999 et reste le fait de la grande presse. Les payeurs publics de ces deux pays n'ont pas réellement mis en place des mécanismes de mise en

concurrence des fournisseurs de soins fondés sur leur performance clinique. Les impacts observés aux États-Unis sont donc plus importants que ceux que l'on pourrait attendre pour l'instant de pays où le payeur public est en position de monopsonne.

On observe alors qu'il y a eu impact de la publicité des études comparatives sur une partie, mais pas tous les fournisseurs concernés. Certains établissements ont en effet déclenché des actions de correction et sont parvenus à améliorer significativement leur performance sur des traitements particuliers. Ces actions sont d'autant plus systématiques que l'étude a été engagée à l'origine précisément pour améliorer la performance sur un traitement particulier. Dans le cas d'autres expériences comme celle de Cleveland, l'objectif initial est plutôt de mettre à la disposition de l'ensemble des acteurs des données comparatives permettant d'éclairer leur choix, mais la mise en œuvre d'actions correctives reste de l'initiative de chacun. Cette initiative est sans nul doute elle-même dépendante de la position concurrentielle du fournisseur, de sa position de négociation avec les payeurs et de la puissance politique représentée localement par les groupes de consommateurs. Malgré cette limite, ces expériences produisent une connaissance importante et utile à l'ensemble des acteurs sur les facteurs critiques de la performance clinique. Enfin, l'impact sur les patients eux-mêmes est ambigu. Les articles publiés montrent que l'information à caractère clinique est complexe et très souvent mal comprise, les personnes profanes préférant nettement les indicateurs permettant de mesurer la satisfaction des patients que les indicateurs cliniques. La publicité donnée dans la presse aux palmarès aurait un effet dans le court terme mais peu durable. En réalité, on manque d'études permettant de savoir comment ces informations ont été utilisées par les personnes profanes pour prendre des décisions telles que le choix d'un traitement et d'un hôpital. Ceci ne signifie pas qu'il faut s'arrêter, mais qu'il faut trouver des modes de communication plus riches et plus ciblés en fonction des caractéristiques des publics.

4. Recommandations

La publicité donnée à des résultats comparatifs entre des établissements hospitaliers, des services de santé et des plans d'assurance est un phénomène qui nous paraît irréversible pour plusieurs raisons : la rationalisation de l'utilisation des ressources dans les systèmes de santé passe par la réduction de l'asymétrie d'information entre les acteurs, au profit des patients et de leurs porte-parole que sont les assureurs, publics ou privés. Cette publicité se justifie autant dans le cadre d'un système concurrentiel à l'américaine que dans le cadre de systèmes publics, qui doivent rendre des comptes aux citoyens sur la qualité du service rendu grâce aux ressources collectives. Cette irréversibilité est renforcée par la facilité croissante d'accès à des informations générales sur la maladie et les traitements, même si cet accès est encore sans doute très inégalitaire. Mais pour que cette publicité soit effectivement profitable à la collectivité, elle doit répondre à des exigences 1) en termes de systèmes d'information ; 2) de partenariat avec les fournisseurs ; 3) de communication envers les personnes bénéficiaires des services de soins ; et 4) de formalisation de la relation qualité-coût avec les payeurs.

4.1. Les systèmes d'information

Tous les auteurs soulignent le travail considérable qu'il est nécessaire de faire pour s'assurer 1) de la standardisation de la définition des données recueillies, 2) de la régularité de ce recueil, 3) de l'exactitude des informations, 4) et enfin de la qualité du traitement statistique réalisé. Ce travail est de la responsabilité de chaque fournisseur mais aussi de l'organisation qui prend la responsabilité de constituer une base inter-établissement. Au lancement de l'expérience, il requiert une période importante de formation puis de rodage. Dans la routine, il est fait d'audits réguliers avec de fréquents retours au dossier médical et des discussions avec l'équipe soignante pour s'assurer de la validité de l'information. Chaque nouvel adhérent ajoute à la valeur de la base, mais il y a peu d'effets d'échelle lors de son adhésion, car si l'organisation gestionnaire des données a acquis une compétence, le nouveau venu a souvent tout à apprendre. La moindre dérive en termes de qualité remet en cause la valeur des conclusions et la crédibilité de la base de données.

Il existe un débat sur la manière de constituer ces bases. Pour diminuer les coûts de constitution et de gestion de ces bases, certains recommandent d'utiliser le plus possible les bases de données existantes, à caractère administratif ou réglementaire. Le principal outil connu est le résumé de sortie (*discharge abstract*), qui sert entre autre à identifier la production de l'hôpital en groupes homogènes de malades (les Diagnosis Related Groups ou tout autre système de classification des séjours). Cependant, ces données présentent l'inconvénient d'être produites à des fins tarifaires, et leur codification répond de ce fait à des consignes précises qui peuvent s'éloigner de la description rigoureuse d'un processus de pris en charge. De surcroît, les données médicales disponibles ne sont pas toujours assez détaillées pour mesurer le résultat d'un processus de soins. C'est pourquoi d'autres auteurs recommandent la constitution de bases spécifiques, s'appuyant sur des données cliniques fines recueillies au cours du séjour du patient. Dans ce cas, la production des données est plus coûteuse et sa qualité plus difficile à maintenir. En effet, pour les équipes de soins, il s'agit de recueillir de façon routinière des données qu'ils utilisent pour un court instant dans le processus de soins mais qu'ils n'ont pas besoin de mémoriser. L'utilité de ce recueil n'est perceptible qu'une fois l'analyse faite, bien après le recueil. Il est difficile de trancher entre ces deux écoles. On n'en connaît pas vraiment les conséquences en termes de coûts, comparés aux bénéfices attendus d'une information plus spécifique dans le deuxième cas. Pour des pays ayant une expérience moins longue que les États-Unis du recueil de « discharge abstracts », il serait faux de penser qu'ils peuvent très vite aboutir à des résultats grâce aux données existantes : les « discharge abstracts » nord-américains contiennent déjà une information plus riche et plus précise que leurs homologues dans les pays européens. Certains auteurs attendent beaucoup de la généralisation du dossier médical informatisé. Mais, là encore, il ne faut pas sous-estimer la difficulté d'aboutir à une structure commune de dossiers au niveau d'un pays, ainsi qu'à une définition standard de tous les items.

4.2. Le partenariat avec les professionnels

Ces bases de données servent à comparer les performances d'équipes de professionnels. Ceux-ci sont certes prêts à participer à des actions pour améliorer leurs résultats, mais redoutent les sanctions qui peuvent en découler, tant économiques que juridiques. Ils sont particulièrement sensibles à l'image que peuvent donner les média d'un établissement, suite à la publication de résultats négatifs. Or ils sont une ressource critique incontournable pour que de tels travaux voient le jour. Ils produisent les données de base du système. Ils peuvent donc en saper les fondements simplement en diminuant leur vigilance sur la qualité des informations produites. Ce comportement peut d'ailleurs se justifier si les données publiées servent aux usagers pour mener des actions en justice contre les professionnels (Liang, 2000). Ils sont aussi les principaux acteurs pour interpréter les résultats et les traduire dans leurs pratiques. Enfin, leur collaboration est souvent nécessaire sur le moyen terme, afin de produire des résultats significatifs. Les règles du jeu en matière de communication des résultats et des décisions qui en résultent doivent donc être négociées avant et respectés tout au long de la gestion de la base de données. Ainsi, dans l'expérience de Cleveland, la coopération a failli s'interrompre lorsque l'un des partenaires, le payeur, contrairement aux engagements pris, a décidé de se servir des résultats du travail pour sélectionner des établissements sans en référer auparavant à l'ensemble du groupe. Ce n'est pas la décision elle-même qui était contestée, mais le non-respect des règles écrites du groupe. Cette précaution peut être comprise comme un plaidoyer pour le maintien d'une certaine impunité des professionnels : elle résulte du constat lucide que même s'il est possible aujourd'hui de mieux observer les comportements des professionnels et de leur demander des comptes, il leur reste un formidable pouvoir discrétionnaire dans la justification *a posteriori* de leurs décisions. Il y a réduction de l'asymétrie d'information, pas disparition.

4.3. La communication aux patients

L'impact de la publication par la presse (ou par des instances autres) de palmarès comparant les résultats cliniques des hôpitaux est peu connu. Il serait opportun de mener des études sur ce thème. L'hypothèse la plus plausible est celle d'un effet d'attention à court terme, qui s'effacerait au cours du

temps. Ceci est logique : l'information produite est surtout utile si on est malade, en passe de recevoir un traitement et que le palmarès donne effectivement des informations sur cette maladie et son traitement. Ce point n'est sans doute pas le plus important. Il ressort que l'information sur les résultats cliniques est une information complexe, surtout si elle est donnée avec les précautions de méthode que réclament les statisticiens. Elle peut être mal comprise, et même à contresens. Elle ne correspond pas aux attentes principales des personnes profanes, qui sont plus intéressées par des indicateurs portant sur la qualité relationnelle de la prise en charge et la satisfaction des patients. Cette information est d'autant moins accessible que le niveau d'éducation du récepteur est faible, et elle ne couvre pas encore toutes les pathologies.

Il est donc nécessaire de trouver des modalités efficaces de communication de ces informations. Ceci peut se faire en travaillant sur la forme des documents mis à la disposition du public, qui doivent inclure plusieurs niveaux de communication et d'explication (*layered information*) de façon à ce que chacun puisse approfondir les thèmes qui les intéressent. Il faut aussi contextualiser l'information, c'est-à-dire expliquer dans quel espace cognitif elle a un sens. Il ne suffit pas de donner des résultats, mais aussi d'expliquer le processus de traitement. Le recours à des médiateurs – associations de patients, par exemple – peut être efficace pour répondre aux questions que se posent les personnes. Enfin, il faut adapter le mode de communication aux différents publics visés. On peut parler de l'initiation d'un processus d'acculturation du grand public aux informations relatives aux résultats cliniques des hôpitaux.

La critique des palmarès et des « league tables » a déjà été faite plus haut. Ce type de classement ne reflète que très grossièrement la réalité des différences de performance entre les établissements, et beaucoup d'établissements seraient en droit de demander réparation pour atteinte à leur réputation. Il est tentant de reprocher aux media la recherche du sensationnel dans la publication de données sous la forme peu valide du palmarès. Comme on ne connaît pas l'effet de telles publications sur la renommée et l'attractivité des établissements, il est difficile d'en mesurer les effets négatifs. Elles ont de toute évidence un effet sur les établissements qui s'estiment mal traités par le palmarès, et dont le droit de réponse n'a pas la même audience que la publication initiale. Trois solutions sont envisageables pour aboutir à des publications qui soient plus rigoureuses. Premièrement, un support concurrent peut être désireux de produire une contre-évaluation en faisant valoir sa plus grande rigueur, ou en montrant les progrès par rapport aux premiers classements. Deuxièmement, un dialogue peut s'instaurer entre experts et journalistes, aboutissant progressivement à la publication de données plus rigoureuses, reflétant les limites statistiques des travaux (Chassin *et al.*, 1996 ; Rosenthal *et al.*, 1998). Troisièmement, les organisations qui produisent ces travaux peuvent en organiser la diffusion eux-mêmes, en utilisant l'aide de spécialistes de la communication.

4.4. Les arbitrages qualité-coût

Que ce soit dans un système concurrentiel ou dans un système à dominante publique, le lien entre niveau de performance et financement est encore peu formalisé. Ceci n'est pas surprenant pour deux raisons. D'une part, malgré le progrès des travaux sur la mesure de la performance, l'activité hospitalière présente une variété qu'il est difficile de résumer par quelques indicateurs. La relation entre le niveau de ces indicateurs et le coût des prises en charge est encore peu étudiée. C'est cette relation qui devrait faire l'objet de travaux en utilisant entre autre les méthodes dites de « benchmarking ». Pour que ce « benchmarking » soit effectif, la collaboration des établissements est indispensable. Il faut donc également trouver les incitations efficaces à la collaboration, sous la forme d'un partage des risques et des bénéfices des actions entreprises pour mieux connaître la performance des établissements. D'autre part, formaliser cette relation qualité-prix revient pour un payeur à expliciter le prix qu'il est prêt à payer pour une unité supplémentaire de qualité. Un payeur en situation concurrentielle obtiendrait en théorie cette information par son marché, un payeur public doit en principe s'appuyer sur le débat démocratique pour fixer son prix. Or, cet affichage des préférences collectives est l'une des tâches les plus difficiles d'un gouvernement, que ce soit en santé ou pour d'autres actions collectives.

NOTES

1. Le seul indicateur de résultat clinique de Dr Foster est la mortalité. Le site publie également par hôpital une fiche descriptive, des résultats comparatifs principalement centrés sur les délais d'attente par spécialité, le nombre de plaintes et le nombre de plaintes traitées.
2. Pour comparer les performances des plans de santé, le NCQA a développé une base de données comparative intitulée HEDIS (Health Plan Employer Data and Information Set) alimentée de façon volontaire par les plans qui ont obtenu l'accréditation du NCQA.
3. Pour des raisons stylistiques, on utilisera dorénavant de façon équivalente dans le texte les termes performance et qualité des soins, pour désigner d'une part des indicateurs centrés sur la mesure des résultats à partir des bases biologiques ou cliniques, et le niveau atteint par un établissement sur ces indicateurs. Il ne s'agit donc pas ici de mesure de l'efficacité, qui rendrait compte aussi de la satisfaction du besoin du patient sur d'autres dimensions, rapportée aux coûts.
4. Il se peut aussi que la préférence pour la bonne qualité soit relativement insensible au niveau de risque perçu par la clientèle elle-même. Par exemple, en France, un obstacle important à la mise en place de réseaux de maternités hiérarchisées selon leurs capacités à prendre en charge des grossesses à risque est que les femmes à très bas risque, qui pourraient accoucher dans des maternités dites de niveau 1 préfèrent néanmoins la sécurité offerte par la technicité des maternités de niveau 3 alors même qu'elles n'en ont pas besoin.

RÉFÉRENCES

- ANSARI, M.Z., ACKLAND, M.J., JOLLEY, D.J., CARSON, N. et MCDONALD, I.G. (1999),
« Inter-hospital comparison of mortality rates », *International Journal of Quality in Health Care*, vol. 2, n° 1, pp. 29-35.
- ARON, D.C., HARMER, D.L., SHEPARDSON, L.B. et ROSENTHAL, G.E. (1998),
« Impact of risk-adjusting Cesarean delivery rates when reporting hospital performance », *JAMA*, vol. 279, n° 24, pp. 1968-1972.
- BUCKNALL, C.E., RYLAND, I., COOPER, A., COUTTS, H., CONNOLLY, C.K. et PEARSON, M.G. (2000),
« National benchmarking as a support system for clinical governance », *Journal of the Royal College of Physicians of London*, vol. 34, n° 1, pp. 52-56.
- CARON, A. et NEUHAUSER, D. (1999),
« The effect of public accountability on hospital performance : trends in rates for Cesarean section rates and vaginal births after Cesarean section in Cleveland, Ohio », *Quality Management in Health Care*, Vol. 7, n° 2, pp. 1-10.
- CHALÉ, J.J. et NAIDITCH, M. (1999),
« Retour sur une enquête controversée », *La Recherche*, n° 324, pp. 71-74.
- CHASSIN, M.R., HANNAN, E.L. et DE BUONO, B.A. (1996),
« Benefits and hazards of reporting medical outcomes publicly », *NEJM*, vol. 334, n° 6, pp. 394-398.
- CIHI (2001),
Hospital Report 2001. www.cihi.ca/HospitalReport/
- COMAROW, A. (2000a),
« Higher volumes, fewer deaths », *USNews*, 17 juillet.
- COMAROW, A. (2000b),
« Picking the top centers. Reputation, outcomes and technology all count », *USNews*, 17 juillet.
- COOPER, J.K., KOHLMANN, T., MICHAEL, J.A., HAFFER, S.C. et STEVIC, M. (2001),
« Health outcomes. New quality measure for Medicare », *International Journal for Quality in Health Care* vol. 13, n° 1, pp. 9-16.
- COULTER, A. et CLEARY, P. (2001),
« Patients' experience with hospital care in five countries », *Health Affairs*, vol. 20 (3), pp. 244-252.
- DREES (1999),
« Les éléments de la performance hospitalière », *Études et Résultats*, n° 42, Paris.
- DREES (2001),
« Éléments pour évaluer la performance des établissements hospitaliers », *Dossiers solidarité et santé*, n° 2, Paris.
- EDDY, D.M. (1998),
« Performance measurement : problems and solutions », *Health Affairs*, vol. 17, n° 4, pp. 7-25.
- EPSTEIN, M. et KURTZIG, B. (1994),
« Statewide health information : a tool for improving hospital accountability », *The Joint Commission Journal of Quality Improvement*, vol. 20, n° 7, pp. 370-375.
- HANNAN, E.L., KILBURN, J.F., O'DONNELL, J.F. et al. (1990),
« Adult open heart surgery in New York State. An analysis of risk factors and hospital mortality rates », *JAMA*, pp. 2768-2774.
- HANNAN, E.L., KILBURN, J.F., RACZ, M. et al. (1994),
« Improving the outcomes of Coronary Bypass Surgery in New York State », *JAMA*, n° 271, pp. 761-766.
- HANNAN, E.L., SIU, A.L., KUMAR, D. et al. (1995),
« The decline in Coronary Bypass Graft Surgery Mortality in New York State. The role of surgeon volume », *JAMA*, n° 273, pp. 209-213.
- HIBBARD, J.H., SLOVIC, P. et JEWETT, J.J. (1997),
« Informing consumer decisions in health care : implications from decision-making research », *The Milbank Quarterly*, vol. 75, n° 3, pp. 395-414.

- HIBBARD, J.H., SOFAER, S. et JEWETT, J.J. (1996),
« Condition-specific performance information : assessing salience, comprehension, and approaches for communicating quality », *Healthcare Financing Review*, vol. 18, n° 1, pp. 95-111.
- HILLNER, B.E., SMITH, T.J. et DESCH, C.E. (2000),
« Hospital and physician volume or specialization and outcomes in cancer treatment : importance in quality of cancer care », *J Clin Oncol*, Vol. 18, n° 11, pp. 2327-2340.
- HUMPHREYS, H. et EMMERSON, A.M. (1993),
« Control of hospital acquired infection : accurate data and more resources, not league tables », *Journal of Hospital Infection*, n° 25, pp. 75-78.
- IEZZONI, L. (1997),
« The risks of risk adjustment », *JAMA*, Vol. 278, n° 19, pp. 1600-1607.
- IEZZONI, L., SCHWARTZ, M., ASH, A.S., HUGHES, J.S., DALEY, J. et MACKIERNAN, Y.D. (1995),
« Using severity-adjusted stroke mortality rates to judge hospitals », *International Journal for Quality in Health Care*, vol. 7, n° 2, pp. 81-94.
- JEWETT, J.J. et HIBBARD, J.H. (1996),
« Comprehension of quality indicators : differences among privately insured, publicly insured and uninsured », *Healthcare Financing Review*, vol. 18, n° 1, pp. 75-94.
- Joint Commission on Accreditation of Health Care Organizations (2001),
The Self-assessment Checklist : Hospitals, Oakbrook terrace, Illinois, États-Unis.
- KRITSCHESKY, S.B., BRAUN, B.I., GROSS, P.A., NEWCOMB, C.S., KELLEHER, C.A. et SIMMONS, B.P. (1999),
« Definition and adjustment of Cesarean section rates and assessments of hospital performance », *International Journal for Quality in Health Care*, vol. 2, n° 4, pp. 283-291.
- LANDON, B., IEZZONI, L., ASH, A., SHWARTZ, M., DALEY, J., HUGUES, J.S. et MACKIERNAN, Y.D. (1996),
« Judging hospitals by severity adjusted mortality rates : the case of CABG surgery », *Inquiry*, n° 33, pp. 155-166.
- LEYLAND, A.H. et BODDY, F.A. (1998),
« League tables and acute myocardial infarction », *The Lancet*, Vol. 351, pp. 555-558.
- LIANG, B.A. (2000),
« Risks of reporting sentinel events », *Health Affairs*, vol. 19, n° 5, pp. 113-120.
- LOMBRAIL, P. et NAIDITCH, M. (1999),
« Comment partager l'information ? », *La Recherche*, n° 324, pp. 75-78.
- MARSHALL, R. (2001),
« Where We Get the Best Health Care », *Macleans*, n° 114, pp.31-36.
- MCCORMACK, L.A., GARFINKEL, S.A., SCHNAIER, J.A., LEE, A.J. et SANGL, J.A. (1996),
« Consumer information development and use », *Healthcare Financing Review*, vol. 18, n° 1, pp. 15-31.
- MCKEE, M. et HUNTER, D. (1995),
« Hospital league tables : do they inform or mislead? », *Quality in Health Care*, n° 4, pp. 5-12.
- MCLAUGHLIN, C.G. et GINSBURG, P.B. (1998),
« Competition, quality of care and the role of the consumer », *The Milbank Quarterly*, vol. 76, n° 4, pp. 737-743.
- MC MULLAN, M. (1996),
« HCFA's consumer information commitment », *Healthcare Financing Review*, vol. 18, n° 1, pp. 9-15.
- MINVIELLE, E. (1996),
« Gérer la singularité à grande échelle », *Revue française de gestion*, n° 109, pp. 119-124.
- NHS (1995),
The NHS Performance Guide 1994-1995, Londres.
- NAIDITCH, M. (1999),
« L'exemple de la chirurgie cardiaque new-yorkaise », *La Recherche*, n° 324, pp. 66-70.
- NATHENS, A.B., JURKOVICH, G.J., GROSSMAN, D.C., MACKENZIE, E.J. et RIVARA, F.P. (2001),
« Relationship between trauma center volume and outcomes », *JAMA*, vol. 285, n° 9, pp. 1164-1171.
- PHILLIPS, K.A., LUFT, H.S. et RITCHIE, J.L. (1995),
« The association of hospital volumes of percutaneous transluminal coronary angioplasty with adverse outcomes, length of stay, and charges in California », *Medical Care*, vol. 33, n° 5, pp. 502-514.
- Le Point (2001),
« Hôpitaux. Le Palmarès 2001 », n° 1151, 31 août.
- RITCHIE, J.L., MAYNARD, C., CHAPKO, M.K., EVERY, N.R. et MARTIN, D.C. (1999),
« Association between percutaneous transluminal coronary angioplasty volumes and outcomes in the Health Care Cost and Utilization Project 1993-1994 », *AM J Cardiol*, vol. 83, n° 4, pp. 493-497.

- ROMANO, P.S., ZACH, A., LUFT, H.S., RAINWATER, J., RELY, L.L. et CAMPA, D. (1995),
« The California Hospital Outcomes Project : using administrative data to compare hospital performance », *The Joint Commission Journal on Quality Improvement*, vol. 21, n° 12, pp. 668-685.
- ROSENTHAL, G. (1997),
« Weak association between hospital mortality for individual diagnoses : implications for profiling hospital quality », *American Journal of Public Health*, vol. 87, n° 3, pp. 429-433.
- ROSENTHAL, G.E., QUINN, L. et HARPER, D.L. (1997),
« Declines in Hospital Mortality associated with a regional initiative to measure hospital performance », *American Journal of Medical Quality*, vol. 12, n° 2, pp. 103-112.
- ROSENTHAL, G.E., HAMMAR, P.J., WAY, L.E., SHIPLEY, S.A., DONER, D., WOJTALA, B., MILLER, J. et HARPER, D.L. (1998),
« Using hospital performance data in quality improvement : the Cleveland Health Quality Choice Experience », *The Joint Commission Journal on Quality Improvement*, vol. 24, n° 7, pp. 347-360.
- SHAW, C.D. (1997),
« Health-care league tables in the United Kingdom », *J. Qual. Clin. Practice*, n° 17, pp. 215-219.
- SILBER, J.H. et ROSENBAUM, P.R. (1997),
« A spurious correlation between hospital mortality and complication rates : the importance of severity adjustment », *Medical Care*, n° 35(suppl.), pp. OS77-OS92.
- SIRIO, C.A., ANGUS, D.C. et ROSENTHAL, G.E. (1994),
« Cleveland Health Quality Choice (CHQC) – An ongoing collaborative, community based outcomes assessment program », *New Horizons*, vol. 2, n° 3, pp. 321-325.
- SISK, J. (1998),
« Increased competition and the quality of health care », *The Milbank Quarterly*, vol. 76, n° 4, pp. 687-707.
- THOMAS, J.W. et HOFER, T.P. (1999),
« Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care », *Medical Care*, vol. 37, n° 1, pp. 83-92.
- VOELKER, R. (1995),
« Why has historic' disclosure of hospital performance data attracted so little attention? », *JAMA*, vol. 273, n° 9, pp. 689-690.

TABLE DES MATIÈRES

Préface de l'éditeur par Peter Smith	7
---	---

Partie I Aperçu des enjeux et des défis

Chapitre 1. Mesurer la performance des systèmes de santé : problèmes et possibilités à l'heure de l'évaluation et de l'imputabilité par C. David Naylor, Karen Iron et Kiren Handa.....	13
Chapitre 2. Mesure et amélioration de la performance dans les systèmes de santé de l'OCDE : aperçu des enjeux et des défis par Jeremy Hurst.....	39

Partie II La mesure et la gestion des performances : bilan des expériences conduites dans quelques pays

Chapitre 3. Améliorer l'efficacité du Service national de santé au Royaume-Uni : la mesure et l'amélioration de la performance dans un système centralisé par Clive H. Smee	63
Chapitre 4. Pour une plus grande transparence dans un système décentralisé : le point de vue de la Suède par Nina Rehnqvist.....	95
Chapitre 5. Les performances du service de santé : l'information du consommateur aux États-Unis par Thomas Reilly, Gregg Meyer, Carla Zema, Christine Crofton, David Larson, Charles Darby et Katherine Crosson.....	105
Chapitre 6. Peut-on transformer une tulipe en rose ? L'autoréglementation guidée d'un réseau de santé intégré axé sur la communauté : l'exemple néerlandais par Niek Klazinga, Diana Delnoij et Isik Kulu-Glasgow.....	129
Chapitre 7. Vers des systèmes d'information sur la santé intégrés et cohérents pour évaluer la performance : l'exemple du Canada par Michael Wolfson et Richard Alvarez.....	147

Partie III Les initiatives de mesure des performances au niveau international : comment les comparaisons internationales peuvent-elles guider l'élaboration des politiques nationales ?

Chapitre 8. Ouvrir la boîte noire : que peut-on apprendre d'une approche par pathologie ? par Stéphane Jacobzone, Pierre Moise et Lynelle Moon	175
---	-----

<i>Chapitre 9.</i> Évolution de l'approche de l'Organisation mondiale de la santé en matière d'évaluation de la performance des systèmes de santé par <i>David B. Evans</i>	213
<i>Chapitre 10.</i> Évaluer l'expérience des patients : comment peut-on améliorer les réseaux de santé ? par <i>Angela Coulter</i> et <i>Paul Cleary</i>	227
<i>Chapitre 11.</i> L'équité en matière d'utilisation des visites médicales dans les pays de l'OCDE : a-t-on atteint l'équité de traitement à besoin équivalent ? par <i>Eddy van Doorslaer</i> , <i>Xander Koolman</i> et <i>Frank Puffer</i>	243

Partie IV

**Mesurer et améliorer différents aspects de la performance des systèmes de santé :
les pratiques exemplaires**

<i>Chapitre 12.</i> La mesure de la qualité des soins à l'hôpital : l'état de l'art par <i>Gérard de Pourville</i> et <i>Étienne Minvielle</i>	275
<i>Chapitre 13.</i> Évaluer la qualité des soins de longue durée en institution et à domicile par <i>Naoki Ikegami</i> , <i>John P. Hirdes</i> et <i>Iain Carpenter</i>	303
<i>Chapitre 14.</i> Construction d'indicateurs composites pour l'évaluation de l'efficacité des systèmes de santé par <i>Peter Smith</i>	323

Partie V

**Utiliser les indicateurs de performance
pour améliorer les systèmes de santé**

<i>Chapitre 15.</i> Utiliser les indicateurs de performance pour améliorer les systèmes de santé par <i>Sheila Leatherman</i>	349
--	-----

Partie VI

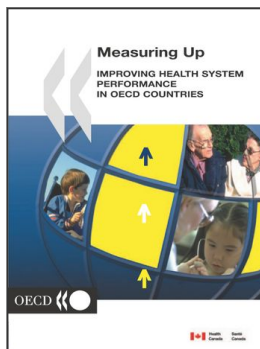
Résumé et conclusions

<i>Chapitre 16.</i> Savoir évaluer : enseignements et perspectives par <i>Jean-François Girard</i> et <i>Étienne Minvielle</i>	369
---	-----

Partie VII

Table ronde ministérielle : leadership, succès et défis

<i>Chapitre 17.</i> Le point de vue des ministres des pays de l'OCDE sur la mesure de la performance des systèmes de santé par <i>Julio Frenk</i> (<i>ministre de la Santé, Mexique</i>), <i>John Hutton</i> (<i>ministre d'État à la Santé, Royaume-Uni</i>), <i>Bernard Kouchner</i> (<i>ministre délégué à la Santé, France</i>), <i>Allan Rock</i> (<i>ministre de la Santé, Canada</i>), <i>Edward Sondik</i> (<i>directeur, Centre national de statistiques sur la santé, États-Unis</i>) au nom de <i>Tommy Thompson</i> , <i>secrétaire à la Santé et aux Services sociaux, États-Unis</i>	383
---	-----



Extrait de :

Measuring Up

Improving Health System Performance in OECD Countries

Accéder à cette publication :

<https://doi.org/10.1787/9789264195950-en>

Merci de citer ce chapitre comme suit :

OCDE (2003), « La mesure de la qualité des soins à l'hôpital : L'état de l'art - quelle information donner au public ? », dans *Measuring Up : Improving Health System Performance in OECD Countries*, Éditions OCDE, Paris.

DOI: <https://doi.org/10.1787/9789264295957-14-fr>

Ce document, ainsi que les données et cartes qu'il peut comprendre, sont sans préjudice du statut de tout territoire, de la souveraineté s'exerçant sur ce dernier, du tracé des frontières et limites internationales, et du nom de tout territoire, ville ou région. Des extraits de publications sont susceptibles de faire l'objet d'avertissements supplémentaires, qui sont inclus dans la version complète de la publication, disponible sous le lien fourni à cet effet.

L'utilisation de ce contenu, qu'il soit numérique ou imprimé, est régie par les conditions d'utilisation suivantes :

<http://www.oecd.org/fr/conditionsdutilisation>.