

Correction du nombre de cancers incidents identifiés à partir des bases de données médico-administratives

Corrected Estimates of the Number of Incident Cancers Identified from Claims Databases

Couris CM (*), Colin C (**), Rabilloud M (***), Schott AM (****), Ecochard R (*****)

Résumé

Objectifs : Les estimations du nombre de cancers incidents hospitalisés obtenues à partir des bases de données médico-administratives nécessitent correction car elles sont soumises à plusieurs types d'erreurs : cancers faussement incidents par manque de spécificité de la base ou défaut d'identification de cancers réellement incidents par manque de sensibilité.

Méthode : La méthode statistique présentée tient compte de la sensibilité et de la spécificité et permet de corriger ces estimations en acceptant une certaine imprécision quantifiée. Cette méthode est basée sur une étude en deux phases. La première phase identifie les cancers potentiellement incidents alors que la deuxième phase estime la sensibilité et la spécificité à partir de données externes. L'imprécision autour de l'estimation corrigée est déterminée par un intervalle de crédibilité calculé selon un principe bayésien utilisant une méthode de Monte Carlo.

Résultats : A partir de la population hospitalisée aux Hospices Civils de Lyon en 1997, 334 femmes ont été identifiées dans la base du PMSI comme potentiellement atteintes d'un cancer du sein incident. La sensibilité et la spécificité de la base de données de *Medicare* pour l'identification de cancers du sein incidents étaient respectivement de 57 % et de 99,9 %. L'estimation corrigée par notre méthode était de 565 avec un intervalle de crédibilité de 550 à 580.

Conclusions : Les bases de données médico-administratives apparaissent très utiles pour documenter l'incidence hospitalière lorsqu'une méthode de correction adaptée est utilisée. Cette approche peut être utilisée pour d'autres maladies à condition de disposer d'estimations adaptées de sensibilité et de spécificité pour l'identification de ces maladies dans les bases de données médico-administratives.

Rev Med Ass Maladie 2001;32,3:257-263

Mots clés : erreur de mesure, bases de données médico-administratives, méthode de correction, cancer, incidence, épidémiologie.

Summary

Objectives: Estimations of the number of hospitalized incident cancer cases have biases when claims databases are used because a lack of specificity can result in an excessive number of incident cancers while truly incident cancers can go unrecorded due to a lack of sensitivity.

Methods: The statistical method presented, which provides corrected estimates, was based on a two-phase study design using an external data set for sensitivity and specificity estimates. Inaccuracy in the corrected number of hospitalized incident cancer cases was assessed by the credibility interval which was calculated with a Bayesian method using a Monte Carlo algorithm.

Results: By using the population hospitalized in a large group of French university hospitals in 1997, we were able to identify 334 women as potential incident cases of breast cancer through the French claims database. The corrected estimate by our method was 565 (credibility interval: 550-580).

Conclusions: Claims databases can be useful for documenting cancer incidence when a correction method is used. This approach has many potential applications for hospital planning and can be used for a number of different diseases.

Rev Med Ass Maladie 2001;32,3:257-263

Key words: measurement error, claims databases, biostatistics, cancer, incidence, epidemiology.

* Doctorante en épidémiologie, Département d'information médicale des Hospices Civils de Lyon.

** Professeur d'épidémiologie, économie de la santé, Département d'information médicale des Hospices Civils de Lyon.

*** Praticien hospitalier universitaire, Service de biostatistiques des Hospices Civils de Lyon.

**** Praticien hospitalier, Unité d'information médicale en cancérologie, Département d'information médicale des Hospices Civils de Lyon.

***** Professeur de biostatistiques, Service de biostatistiques des Hospices Civils de Lyon.

Adresse pour correspondance : Chantal Marie Couris, Département d'information médicale des Hospices Civils de Lyon, 162, avenue Lacassagne, F-69424 Lyon cedex 03.

E-mail : chantal.couris@chu-lyon.fr

INTRODUCTION

Les bases de données médico-administratives englobent pour chaque séjour d'un patient dans un établissement de soins des informations standards et obligatoires sur le patient, ses maladies et sur les soins reçus [1]. Les bases de données médico-administratives des établissements de soins en France ont été constituées suite à la mise en place du « Programme de Médicalisation du Système d'Information » (PMSI). De par leur couverture et compte tenu de la procédure de chaînage des séjours récemment mis en place, ces bases de données peuvent potentiellement fournir des estimations nationales du nombre de cancers incidents hospitalisés. Cependant, l'identification de faux cancers incidents due à un manque de spécificité et le défaut d'enregistrement des cancers incidents lié à un manque de sensibilité peuvent biaiser l'estimation du nombre de cancers incidents [2-5]. En conséquence, les bases de données médico-administratives sont imparfaites pour fournir des indications sur l'activité hospitalière liée à la prise en charge de cancers incidents. Compte tenu du rôle majeur de cette information dans le choix des orientations de la politique hospitalière [6, 7], les estimations issues des bases médico-administratives doivent impérativement être corrigées.

Quelques études ont évalué la fiabilité des bases médico-administratives hospitalières pour l'identification des cancers incidents [8, 9]. Dans la détection des cancers du sein incidents, la plus fréquente cause d'erreur était liée à la présence d'un code diagnostique de cancer du sein et d'un code acte de biopsie qui s'est par la suite révélé négative. Dans cette étude, les dossiers médicaux étaient considérés comme le « gold standard » [8]. En dehors des limites liées aux problèmes de codage de l'information médicale, la principale difficulté était liée à la différenciation des cancers incidents et des cancers prévalents. Pour pallier au mieux cette difficulté, des algorithmes composés de codes diagnostics et de codes actes spécifiques du cancer ont été utilisés [8-10]. Avec des registres du cancer comme « gold standard », la sensibilité des bases médico-administratives hospitalières pour l'identification de cancers invasifs incidents de la prostate, du poumon, du colon-rectum, du sein, du pancréas, et de l'endomètre, variait de 63,6 % (cancers invasifs de la prostate) à 86,1 % (cancers colo-rectaux invasifs) [9]. D'autre part, le fait que les bases médico-administratives aient pour objectif principal d'évaluer les dépenses de santé, plutôt que de fournir des informations sur l'état de santé des patients et leur prise en charge, peut également s'avérer être à l'origine de biais potentiels dans la détection des cancers incidents hospitalisés. Un diagnostic peut avoir été présent sans être enregistré dans la base médico-administrative au profit de maladies plus

sérieuses ou simplement pour ne pas avoir été considéré comme le diagnostic contribuant le plus aux soins fournis au patient. Cependant, dans le cas particulier du cancer, cette limite se présente comme mineure.

Globalement, au vu de l'ensemble des imperfections connues, il apparaît clairement qu'une mesure vraie ou exacte du nombre de cancers incidents hospitalisés ne peut être obtenue directement. Les mesures obtenues à partir des bases de données médico-administratives doivent impérativement être considérées comme des mesures de substitution (imparfaite). Cette mesure imparfaite est égale à la mesure exacte plus un écart qu'il est nécessaire d'estimer [11, 12]. Notre objectif est donc de présenter une des méthodes statistiques permettant de fournir des estimations corrigées du nombre de cancers du sein incidents hospitalisés assorties d'un intervalle de crédibilité. Cette méthode repose sur un schéma d'étude en deux phases [13, 14], avec l'utilisation au cours de la deuxième phase de données externes qui fournissent des estimations de sensibilité et de spécificité. Le nombre de femmes hospitalisées avec un cancer du sein incident aux Hospices Civils de Lyon en 1997 a été calculé à partir de la base de données médico-administratives française. Cette estimation a été corrigée en utilisant la sensibilité et la spécificité pour l'identification de cancers du sein incidents calculées à partir de la base de données médico-administratives du programme d'assurance américain, *Medicare* [10]. Ces estimations de sensibilité et de spécificité ont été choisies du fait de l'absence de données française équivalentes.

MÉTHODES

1. Population d'étude

Cette analyse englobe toutes les femmes de 20 ans et plus vivant dans le département du Rhône et hospitalisées aux Hospices Civils de Lyon en 1997, soit 44 507 femmes. Les Hospices Civils de Lyon constituent le second plus gros groupes d'hôpitaux publics en France. Ils englobent 13 hôpitaux, soit un total de 118 pavillons médicaux et chirurgicaux et 4 800 lits de court séjour.

a) Données

Les données pour ce projet sont issues de la base de données du PMSI. Cette base est composée de résumés standards comportant des informations obligatoires sur le patient, les diagnostics (jusqu'à 8 codes suivant la CIM-10 depuis 1997) et les soins reçus (jusqu'à 7 codes issus du Catalogue des Actes Médicaux) pour chaque séjour d'un patient dans un établissement de soins. Le diagnostic principal est le

motif de prise en charge qui a mobilisé l'essentiel de l'effort médical et soignant au cours du séjour du malade dans une unité médicale [15]. Aux Hospices Civils de Lyon un identifiant unique et permanent est attribué à chaque nouveau patient hospitalisé.

b) Identification des patientes hospitalisées avec un cancer du sein incident

Parmi les 44 507 femmes de la population d'étude, celles ayant une première occurrence de code de cancer du sein en diagnostic principal (code C50.0-C50.9, D05.0-D05.9 d'après la CIM-10) en 1997, ont été identifiées et définies comme de potentiels cas de cancers du sein incidents.

Parmi ces cas potentiels ont été identifiés des cas de cancers prévalents correspondant aux femmes qui avaient déjà eu un code de cancer du sein (code 174.0-174.9, 233.0 d'après la CIM-9) ou un code antécédent personnel de cancer du sein (code Z85.3 d'après la CIM-10, code V10.3 d'après la CIM-9) au cours d'un séjour entre 1991 (première année de disponibilité des données) et 1996. Ces cas de cancers prévalents ont été exclus.

2. Base externe de données

La base de données médico-administratives hospitalière du programme d'assurance *Medicare*, maintenue par la *Health Care Financing Administration* aux États-Unis, est similaire à celle issue du PMSI. Jusqu'à 10 codes de diagnostic et d'acte sont enregistrés d'après la ICD-9-CM pour les personnes âgées d'au moins 65 ans, les personnes handicapées et les personnes souffrant d'insuffisance rénale chronique. Dans l'étude que nous avons utilisée [10], les femmes de 65 ans et plus, vivant dans l'un des 5 États d'Amérique couverts par un registre du cancer du programme « *Surveillance, Epidemiology, and End Results* » (SEER) (Connecticut, Hawaii, Iowa, New Mexico, et Utah) ont été sélectionnées en 1992. L'échantillon final incluait 659 260 femmes. Les femmes hospitalisées avec une première occurrence de code de cancer du sein en diagnostic principal (code 174-174.9, 233.0 selon ICD-9-CM) ont été identifiées. Parmi ces potentiels cas de cancers incidents ont été identifiés des cas de cancers prévalents correspondant aux femmes ayant un code diagnostic de cancer ou antécédent personnel de cancer (code V10.3) lors d'une hospitalisation *Medicare* entre 1984 et 1991. Ces cas de cancers prévalents ont été exclus.

3. Sensibilité et spécificité

Les registres de population du programme SEER ont été utilisés comme « gold standard » de vrais cas incidents. Warren *et al.* [10] ont défini la sensibilité (*Se*) comme le pourcentage de cas de cancers du sein enregistrés dans les registres de population

du programme SEER en 1992 identifié dans la base de données médico-administratives du programme *Medicare* la même année. La spécificité (*Sp*) a été définie comme le pourcentage de femmes identifiées comme indemnes de cancer du sein incident d'après la base de données *Medicare* en 1992 parmi les femmes non retrouvées dans l'un des cinq registres la même année.

4. Analyse statistique

L'analyse statistique a été développée par ailleurs [16]. Brièvement, l'estimation corrigée du nombre de cas de cancers incidents (Kc_{reel}) a été obtenue en utilisant la relation suivante :

$$Kc_{reel} = \frac{Kc_{enr} - (1 - Sp)N}{Se + Sp - 1}$$

Dans cette relation, *N* représente le nombre de femmes de 20 ans et plus vivant dans le département du Rhône et hospitalisées aux Hospices Civils de Lyon en 1997 ; Kc_{enr} symbolise le nombre de femmes ayant potentiellement un cancer du sein incident d'après la base de données médico-administratives française parmi *N*. L'imprécision autour de Kc_{reel} a été estimée en utilisant un intervalle de crédibilité déterminé selon un principe bayésien. Dans cette approche, la loi de probabilité *beta* a été choisie pour reconstituer les distributions (probabilité subjective) de *Se* et *Sp*. Il a alors été possible de fournir une estimation de la distribution *a posteriori* du nombre corrigé de cancers du sein incidents en utilisant une méthode de Monte Carlo. La moyenne de cette distribution (densité de probabilité) a été utilisée comme estimateur du nombre corrigé de cancers du sein incidents. Les 2,5^e et 97,5^e percentiles de la distribution ont servi d'estimateurs pour définir les limites de l'intervalle de crédibilité. Un graphe de la densité de probabilité *a posteriori* du nombre corrigé de cancers du sein incidents a été obtenu en utilisant une méthode non paramétrique d'estimation de la densité de probabilité (fonction Stata : *kdensity*).

RÉSULTATS

Parmi les 44 507 femmes vivant dans le département du Rhône et hospitalisées aux Hospices Civils de Lyon en 1997, 35,5 % étaient âgées de plus de 64 ans. 420 ont été identifiées avec au moins un séjour comportant un diagnostic principal de cancer du sein. Parmi elles, 86 (20 %) avaient déjà eu un code de diagnostic de cancer du sein ou d'antécédent personnel de cancer du sein entre 1991 et 1996. Le nombre observé de femmes atteintes de cancers incidents du sein était donc égal à 334. Parmi ces femmes, 34,8 % avaient plus de 64 ans.

Tableau I
Sensibilité et spécificité de la base de données médico-administratives hospitalière du programme *Medicare* pour identifier les femmes atteintes d'un cancer du sein incident en 1992

MEDICARE :		Registres (SEER), 1992			
diagnostic principal de cancer du sein	Oui	Non	Total		
Oui	1 841	176	2 017		Sensibilité = 57,0
Non	1 389	655 854	657 243		Spécificité = 99,9
Total	3 230	656 030	659 260		

Reproduit avec l'autorisation des éditeurs [10]

La sensibilité et la spécificité de la base de données médico-administratives hospitalière du programme *Medicare* pour l'identification de cancers du sein incidents étaient de 57 % et de 99,9 %, respectivement en 1992 (Tableau I) [10].

Le nombre estimé de cancers du sein incidents a été augmenté de 41 % après correction selon notre méthode. L'estimation corrigée du nombre de cancers du sein incidents hospitalisés était finalement égale à 565 (547-584).

La distribution *a posteriori* du nombre corrigé de cancers du sein incidents a une allure en forme de cloche (Figure 1).

DISCUSSION

Cette étude a permis d'estimer le nombre de femmes hospitalisées aux Hospices Civils de Lyon avec un cancer du sein incident en 1997 grâce à l'utilisation d'une approche bayésienne. Cette approche tient compte de l'erreur de mesure liée à l'utilisation de la base de données du PMSI. Dans le contexte d'utilisation de notre méthode, il apparaît à première vue que la correction porte essentiellement sur le manque de sensibilité. Cependant, la sensibilité et la spécificité ont été considérées. Ces estimations étaient respectivement de 57 % et de 99,9 % dans l'étude américaine que nous avons utilisée [10]. Dans ce contexte, si la méthode de correction avait uniquement tenu compte de la sensibilité, l'estimation corrigée du nombre de cancers du sein incidents hospitalisés aurait atteint 585 (334/0,57) au lieu de 565. Si un autre algorithme de sélection avait été choisi avec une sensibilité de 76,2 % et une spécificité de 99,3 % (annexe 1), l'estimation corrigée aurait atteint 8 902 (6 784/0,762) au lieu de 3 230. Cette dernière estimation aurait ainsi été fortement biaisée. Ces deux éléments montrent clairement que les faux positifs doivent nécessairement être pris en compte même si la spécificité apparaît élevée. L'influence du manque de spécificité est fortement liée à la proportion élevée de femmes indemnes de cancers du sein.

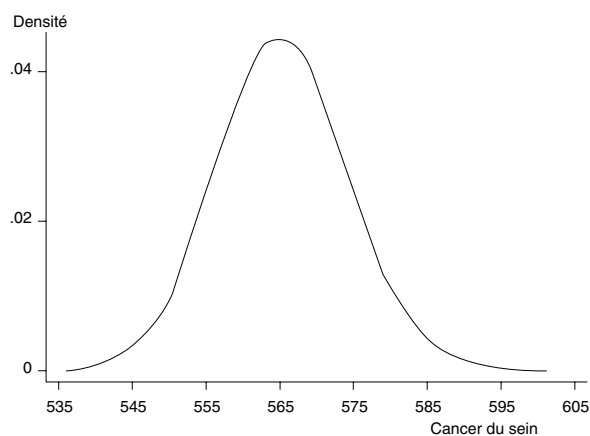


Figure 1. Densité de probabilité du nombre de cancers du sein incidents hospitalisés aux Hospices Civils de Lyon en 1997, estimation obtenue à partir d'une approche bayésienne.

Le fait que la base de données médico-administratives fournisse des valeurs sous-estimées est en accord avec les résultats publiés par Warren *et al.* [10]. Dans ces deux études, le même algorithme de sélection des patients atteints de cancers du sein incidents (femmes hospitalisées avec une première occurrence de diagnostic principal de cancer du sein) a été appliqué. Cet algorithme de sélection a d'ores et déjà été décrit comme incomplet pour identifier tous les patients hospitalisés avec un cancer incident [8-10]. Cependant, Warren *et al.* [10] a trouvé une valeur prédictive positive de 91,3 % pour les femmes de 65 ans et plus avec cet algorithme d'identification. Bien que la probabilité d'identifier une femme atteinte d'un cancer du sein incident soit plus importante dans la base de données *Medicare* que dans la base française, à cause de la différence d'âge de la population couverte par ces deux bases, cet algorithme a été considéré comme offrant le meilleur compromis pour obtenir un taux de faux positif bas. Par ailleurs, dans cette étude, nous avons choisi d'utiliser un seul algorithme de sélection pour illustrer notre méthode de correction. Toutefois, différents algorithmes ont été décrits dans la littérature [8-10]. Ces différents algo-

Annexe 1
Sensibilité et spécificité de l'ensemble des bases de données médico-administratives du programme Medicare (hôpital, ambulatoire et médicale) pour l'identification des cancers du sein incidents, 1992

D'après les données publiées par Warren *et al.* [10], le tableau suivant a été obtenu pour les femmes identifiées avec un code de diagnostic principal ou associé de cancer du sein dans les bases de données médico-administratives hospitalière et médicale du programme Medicare. Ces bases de données couvrent l'ensemble du système de soins (ville et hôpital).

	Cancer (+)	Cancer (-)	Total
Medicare (+)	2 461	4 323	6 784
Medicare (-)	769	651 707	652 476
	3 230	656 030	659 260

Medicare (+) et Medicare (-) étaient respectivement, le nombre de femmes identifiées avec un code diagnostique de cancer du sein dans les bases de données du programme Medicare et le nombre de femmes qui n'ont pas de codes diagnostique de cancers du sein dans ces bases.

Cancer (+) et cancer (-) étaient respectivement, le nombre de femmes ayant ou non un cancer du sein incident d'après les registres du programme du « *Surveillance, Epidemiology and End Results* ».

Dans cette situation, les valeurs de sensibilité et de spécificité étaient estimées à 76,2 % et 99,3 %.

rithmes cherchent à identifier les patients atteints d'un cancer incident en regardant la position des codes diagnostics de cancer et/ou les codes actes spécifiques du cancer. La méthode de correction présentée est adaptée pour tous types d'algorithmes d'identification dans la mesure où des estimations de sensibilité et de spécificité sont disponibles pour chaque algorithme utilisé.

Notre équation de prise en compte de l'erreur de mesure a été développée à partir d'un schéma d'étude en deux phases en utilisant des données externes au cours de la deuxième phase. L'utilisation de données issues d'une étude indépendante pour estimer l'erreur de mesure peut introduire un biais dans l'estimation finale [12]. Pour cette raison, quelques différences entre la population française et la population américaine méritent d'être considérées afin d'évaluer le biais éventuellement introduit. Ces populations sont différentes en terme de distribution d'âge, et potentiellement selon le type de cancer, les stratégies de traitements et la politique d'enregistrement de la base de données médico-administratives.

En considérant la distribution d'âge de notre population d'étude, 35,5 % des femmes avaient plus de 64 ans. Ainsi pour ces femmes, notre estimation corrigée peut être considérée comme plus exacte car les estimations de sensibilité et de spécificité ont été obtenues à partir d'une population de femmes âgées de plus de 64 ans. Cela est d'autant plus vrai si une variation de la sensibilité et de la spécificité avec l'âge est admise. En effet, pour une population âgée, le nombre d'hospitalisations peut être supposé plus élevé à cause de l'existence, la plupart du

parativement à une population plus jeune. En conséquence, la chance de détecter un nouveau cas de cancer est aussi probablement plus importante au cours d'une hospitalisation.

En considérant le type de cancer, la seule condition pour être détecté est d'avoir été hospitalisé avec un diagnostic principal de cancer du sein dans les deux bases de données. La différence potentielle entre le type de cancer du sein hospitalisé en France et aux Etats-Unis n'apparaît donc pas comme une limite à l'utilisation des estimations de sensibilité et de spécificité issues des bases américaines.

En considérant les stratégies de traitements, dans les deux études, seule la première occurrence d'un code diagnostique principal de cancer du sein a été retenue pour définir un potentiel cancer incident. Ainsi pour un patient sélectionné, il n'y avait aucune mention de code diagnostique concernant le cancer du sein avant le séjour sélectionné et aucun autre séjour postérieur n'a été examiné. Le séjour sélectionné correspondait majoritairement à un séjour pour chirurgie. Le traitement chirurgical étant pratiqué dans la majorité des cas, dans les deux pays, nous admettons que même si les stratégies de traitements peuvent varier avec l'âge [17], et plus particulièrement le nombre de chimiothérapies et de radiothérapies, cela influence peu les estimations de sensibilité et de spécificité.

En considérant la politique d'enregistrement des bases de données médico-administratives, la base française est hautement similaire à la base américaine. En effet, la construction de la base française a été directement influencée par celle de Medicare. Dans chacun des pays, les deux bases ont été constituées pour optimiser l'allocation de ressources aux

établissements de santé. Les deux bases utilisent sensiblement la même classification basée sur la constitution de groupes homogènes de malades (*Diagnosis-Related Groups*). Un point supplémentaire à évaluer était la comparabilité des deux bases sur le plan de la qualité du codage concernant le cancer. Les contrôles de qualité sont obligatoires dans les deux pays, il est donc possible d'émettre l'hypothèse de comparabilité de la qualité du codage pour cancer entre les deux bases. La qualité de l'information médicale d'un des principaux hôpitaux des Hospices Civils de Lyon, l'hôpital Édouard-Herriot, a été évaluée en 1990. L'erreur de codage était définie au troisième digit des codes de la CIM 9. Globalement, le taux d'erreur de concordance entre résumés de séjours et dossiers médicaux était de 11,4 % pour les codes diagnostics et actes. Cette estimation était au moins équivalente à celles publiées pour les autres bases de données médico-administratives telles que celles issues de *Medicare* ou de l'administration des Vétérans aux Etats-Unis d'Amérique. Le codage pour cancer n'a pas été directement évalué dans cette étude. Cependant, l'hôpital évalué a une forte activité de cancérologie. De plus, il n'apparaît aucun argument permettant d'envisager des résultats différents pour le codage du cancer comparativement au codage d'autres maladies [18].

Une autre différence entre les deux études est à souligner. Cette différence concerne la longueur de la période de recherche des cancers prévalents, celle-ci étant plus longue dans l'étude américaine. Cette différence influence probablement les estimations de sensibilité et de spécificité, car plus la période de recherche de cancers prévalents est longue et plus la probabilité de les exclure est élevée. Cette notion est d'autant plus vraie pour le cancer du sein qu'il peut présenter des récives au moins jusqu'à 10 ans après le diagnostic initial. Néanmoins, les deux périodes comparées étaient limitées par la disponibilité des données. Concernant les données issues du programme *Medicare*, la principale limite est liée à l'impossibilité de rechercher des cancers prévalents parmi les femmes âgées de moins de 65 ans. Cet élément est probablement la limite principale pour la transposabilité des estimations de sensibilité et de spécificité américaines dans notre étude.

Malgré l'ensemble des limites énoncées ci-dessus, l'utilisation de données externes présente certains avantages et se trouve particulièrement utile en l'absence d'information suffisante dans la population d'étude pour modéliser directement l'erreur de mesure [12]. Globalement, nous avons trouvé opportun d'utiliser les résultats d'une étude américaine offrant des estimations de sensibilité et de spécificité très précises compte tenu du grand nombre de femmes évaluées. Dans tous les cas,

nous n'excluons pas la nécessité d'utiliser des estimations de sensibilité et de spécificité basées sur une population ayant une plus large distribution d'âge. A partir du moment où des estimations de sensibilité et de spécificité sont disponibles pour une maladie, la validité de ces estimations peut être admise pour une durée de 5 ans. Cette période de validité est fortement liée aux variations de qualité des données au cours du temps des bases médico-administratives. Cette hypothèse n'est cependant pas très forte car les contrôles de qualité sont obligatoires dans la plupart des pays disposant de bases de données médico-administratives.

Notre étude ne fournit qu'une estimation globale du nombre de femmes hospitalisées avec un cancer du sein incident. Cependant, il est également possible, avec des estimations adaptées de sensibilité et spécificité, de fournir ces mêmes estimations en fonction de l'âge, du stade du cancer (*in situ*, invasif) ou du traitement dans la mesure où ces informations sont disponibles dans la base de données médico-administratives.

Il est sage de préciser que la méthode présentée n'a pas pour objectif de remplacer les registres du cancer. En revanche, les estimations fournies peuvent servir d'indicateurs aux différents registres lors de leur travail d'investigation sur les nouveaux cas. De plus, les estimations issues des bases de données médico-administratives peuvent être utilisées dans les régions dépourvues de registre.

La méthode peut être utile pour tous ceux qui souhaitent utiliser les bases de données médico-administratives pour fournir des estimations sur l'incidence hospitalière de certains cancers ou d'autres maladies. Ces estimations peuvent s'avérer indispensables pour définir la politique hospitalière notamment en terme d'équipements et/ou de qualité de prise en charge. Ainsi, une nouvelle perspective d'utilisation des bases de données médico-administratives apparaît. En France, l'environnement politique est de plus en plus favorable à cette perspective, notamment grâce au chaînage anonyme des différents séjours d'un même patient pour les admissions en établissement de santé postérieures au 31 décembre 2000.

RÉFÉRENCES

1. Connell FA, Diehr P, Hart LG. The use of large data bases in health care studies. *Annu Rev Public Health* 1987;8:51-74.
2. Dussaucy A, Viel JF, Mulin B, Euvorard J. The framework Prospective Payment Information Systems : bias, sources of errors and consequences. *Rev Epidemiol Sante Publique* 1994;42(4):345-58.
3. Fourquet F, Demont F. Utilisation du PMSI en épidémiologie. *Technologie Santé* 1999;39:55-64.

4. Greenland S, Robins JM. Confounding and misclassification. *Am J Epidemiol* 1985;122(3):495-506.
5. Lombrail P, Minvielle E, Comar L, Gottot S. Prospective Payment Information Systems and epidemiology : a difficult link to establish. *Rev Epidemiol Sante Publique* 1994;42(4):334-44.
6. Borella L, Peuvrel P, Sauvage M, Maraninchi D, Philip T. A study based on national DRG data to evaluate work load and practice relating to cancer patients in not-for-profit hospitals. *Rev Epidemiol Sante Publique* 2000;48(1):53-70.
7. Direction générale de la santé, direction des hôpitaux. Circulaire DGS/DH/AFS n° 98-213 du 24 mars 1998 relative à l'organisation des soins en cancérologie dans les établissements d'hospitalisation publics et privés.
8. Solin LJ, Legorreta A, Schultz DJ, Levin HA, Zatz S, Goodman RL. Analysis of a claims database for the identification of patients with carcinoma of the breast. *J Med Syst* 1994;18(1):23-32.
9. Cooper GS, Yuan Z, Stange KC, Dennis LK, Amini SB, Rimm AA. The sensitivity of Medicare claims data for case ascertainment of six common cancers. *Med Care* 1999;37(5):436-44.
10. Warren JL, Feuer E, Potosky AL, Riley GF, Lynch CF. Use of Medicare hospital and physician data to assess breast cancer incidence. *Med Care* 1999;37(5):445-56.
11. Bashir SA, Duffy SW. The correction of risk estimates for measurement error. *Ann Epidemiol* 1997;7(2):154-64.
12. Carroll RJ, Ruppert D, Stefanski LA. *Measurement error in non-linear models*. London : Chapman & Hall, 1995.
13. Zhao LP, Lipsitz S. Designs and analysis of two-stage studies. *Stat Med* 1992;11(6):769-82.
14. Holford TR, Stack C. Study design for epidemiologic studies with measurement error. *Stat Methods Med Res* 1995;4(4):339-58.
15. Ministère des Affaires sociales et de la Solidarité nationale. Circulaire n° 119 du 4 octobre 1985, relative à la mise en place dans les établissements hospitaliers de résumé de sortie standardisé (RSS). *Bulletin Officiel* ; n° 85-24 bis.
16. Couris C, Colin C, Rabilloud M, Schott A, Ecochard R. Method of correction to assess the number of hospitalized incident breast cancer cases based on claims databases. *J Clin Epidemiol* 2001 (In Press).
17. Ballard-Barbash R, Potosky AL, Harlan LC, Nayfield SG, Kessler LG. Factors associated with surgical and radiation therapy for early stage breast cancer in older women. *J Natl Cancer Inst* 1996;88(11):716-26.
18. Colin C, Ecochard R, Delahaye F, et al. Data quality in a DRG-based information system. *Int J Qual Health Care* 1994;6(3):275-80.