

31 janvier 2012 / n° 5-6

Numéro thématique - Surveillance épidémiologique des cancers en France : utilisation des bases de données médico-administratives

Special issue - Cancer surveillance in France: the use of health administrative databases

p. 53 **Éditorial / Editorial**

p. 54 **Sommaire détaillé / Table of contents**

Coordination scientifique du numéro / *Scientific coordination of the issue*: Arlette Danzon et Agnès Rogel, Institut de veille sanitaire, Saint-Maurice, France, et pour le comité de rédaction : Juliette Bloch, Caisse nationale de solidarité pour l'autonomie, Paris, France ; et Bertrand Gagnière, Cellule de l'InVS en région Ouest, Rennes, France

Éditorial – Surveillance des cancers à partir des bases médico-administratives : des avancées prometteuses pour répondre aux attentes du terrain

Editorial – Cancer surveillance based on health administrative databases: promising steps to respond to realities

Amandine Cochet¹ et Hubert Isnard²

1/ Cellule de l'InVS en région Languedoc-Roussillon, Montpellier, France

2/ Cellule de l'InVS en régions Île-de-France – Champagne-Ardenne, Paris, France

Aujourd'hui, près de 20% de la population française est couverte par un registre des cancers. Si cette couverture est jugée suffisante pour la surveillance au niveau national et la production d'estimations nationales et régionales d'incidence des cancers, elle ne permet pas de répondre à l'ensemble des besoins sur le terrain.

Les risques environnementaux, auxquels la population est de plus en plus sensible, sont l'objet de nombreuses inquiétudes. Certains suspectent telle installation industrielle d'avoir un effet néfaste sur leur santé ; d'autres perçoivent un nombre inhabituel de cancers dans leur entourage : autant de situations dans lesquelles l'existence de données locales et récentes sur la survenue des cancers serait précieuse.

Les responsables de l'organisation de l'offre de soins ont eux aussi besoin de disposer de données de surveillance des cancers à un niveau infrarégional, par exemple dans les territoires de santé, afin de déployer au mieux la palette des soins en cancérologie.

Enfin, la question des inégalités sociales et territoriales de santé, qui est au centre des programmes régionaux de santé, concerne très fortement les cancers. Aujourd'hui, seule la mortalité par cancer permet de l'approcher. Mais quelle est l'incidence de certains cancers à un niveau infradépartemental ? Quelles hypothèses les inégalités éventuellement constatées peuvent-elles faire émerger ? Existe-t-il des différences dans le stade au diagnostic ? Y a-t-il des différences de survie entre territoires ? Comment ces données évoluent-elles dans le temps ? Autant de questions auxquelles il faudrait pouvoir répondre...

Pour apporter des réponses à ces questionnements, l'accent est mis aujourd'hui, en complément des données issues des registres, sur l'utilisation des données médico-administratives à des fins de surveillance épidémiologique. En effet, la mise en place à l'échelle nationale de registres de cancers fonctionnant avec la méthodologie des registres départementaux actuels (méthodologie qui en fait des dispositifs de référence – voir l'article d'A. Danzon et coll.) n'a pas été jugée faisable pour les cancers de l'adulte. C'est ainsi que le déploiement d'un système national multi-sources de surveillance des cancers (SMSC), basé sur le croisement passif automatisé de plusieurs sources de données anonymisées, a été envisagé et expérimenté.

Plusieurs pistes sont explorées dans ce numéro du BEH : le Programme de médicalisation des systèmes d'information (PMSI) pour l'estimation de l'incidence à l'échelon infranational (N. Bossard et coll.), la base des Affections de longue durée (ALD) pour l'estimation de l'incidence départementale ou la surveillance à court terme de l'évolution des cancers (Z. Uhry et coll.). Cependant, la comparaison des données de registres de cancer avec, d'une part, les ALD et d'autre part le PMSI sur la même population, montre les limites importantes en termes de sensibilité et de spécificité de ces deux sources, qui peuvent difficilement être utilisées seules pour la surveillance selon les auteurs (P. Grosclaude et coll. ; D. Lauzeille et coll.). Le croisement des données ALD/PMSI, déjà à l'étude au sein de l'Institut de veille sanitaire, est une piste intéressante. À terme, l'utilisation d'une troisième source complémentaire, constituée à partir des données anatomo-cytopathologiques dans le cadre du déploiement du SMSC, devrait permettre de tendre à l'exhaustivité des cas et de disposer de données plus complètes et précises sur les cancers pour tout le territoire et à une échelle géographique fine (Y. Kudjawu et coll.). Ce dispositif devrait donc mieux répondre aux besoins du terrain. La mise en place de cette troisième source est toutefois dépendante du déploiement du Dossier communicant de cancérologie, ce qui nous invite à la patience.

Au-delà des questions méthodologiques, il est aujourd'hui urgent de mieux identifier les besoins d'informations sur le cancer au niveau local et la façon de s'en servir pour lutter contre les inégalités de santé, qu'elles soient liées aux facteurs de risque, à l'environnement ou à l'accès aux soins. Cette réflexion est en partie portée par les Agences régionales de santé dans le cadre des Schémas régionaux de l'offre de soins. Elle bénéficierait d'une articulation forte avec les réflexions de l'Institut de veille sanitaire et de l'Institut national du cancer sur leurs orientations régionales, et pourrait se concrétiser par l'organisation de séminaires réunissant l'ensemble des partenaires concernés.

Sommaire détaillé / Table of contents

SURVEILLANCE ÉPIDÉMIOLOGIQUE DES CANCERS EN FRANCE : UTILISATION DES BASES DE DONNÉES MÉDICO-ADMINISTRATIVES *CANCER SURVEILLANCE IN FRANCE: THE USE OF HEALTH ADMINISTRATIVE DATABASES*

- p. 53 **Éditorial**
Editorial
-
- p. 54 **La surveillance épidémiologique des cancers en France : outils actuels et perspectives**
Cancer surveillance in France: current tools and perspectives
-
- p. 58 **Tendances récentes des données d'affections de longue durée (ALD) : intérêt pour la surveillance nationale de l'incidence des cancers, période 1997-2009, France**
Recent trends in long-term illness health insurance data: relevance for monitoring recent national trends in cancer incidence, France, 1997-2009
-
- p. 63 **Utilité des bases de données médico-administratives pour le suivi épidémiologique des cancers. Comparaison avec les données des registres au niveau individuel**
Relevance of health administrative databases in cancer surveillance. Comparison with registries records at individual level
-
- p. 68 **Les données d'exonération du ticket modérateur pour affections de longue durée (ALD), source potentielle pour la surveillance des cancers d'origine professionnelle ?**
Long-term illness health insurance data: a potential source for monitoring occupational cancer?
-
- p. 71 **Évaluation de l'utilisation des données d'affection de longue durée (ALD) pour estimer l'incidence départementale des cancers pour 24 localisations cancéreuses, France, 2000-2005**
Evaluation of the use of long-term disease health insurance data to estimate cancer incidence for 24 cancer sites at district level, 2000-2005
-
- p. 77 **Utilisation des données du PMSI pour estimer l'incidence des cancers en France à l'échelon infranational : la démarche du groupe Oncepi**
Using data from the French hospital database (PMSI) to estimate district-specific cancer incidence: the ONCEPI group approach
-
- p. 81 **Système multi-sources de surveillance des cancers (SMSC) : bilan de l'étude pilote**
Multi source surveillance system of cancers (MSSC): assessment of the pilot study
-

La surveillance épidémiologique des cancers en France : outils actuels et perspectives

Arlette Danzon (a.danzon@invs.sante.fr), Joëlle Le Moal, Laurence Chérié-Challine, Anne-Catherine Viso

Institut de veille sanitaire, Saint-Maurice, France

Résumé / Abstract

L'article décrit l'intérêt respectif des informations issues de différents dispositifs de surveillance des cancers, et de dispositifs dont ce n'est pas l'objectif premier, pour l'évaluation de la stratégie de lutte contre le cancer. Un éclairage spécifique pour les cancers liés à l'environnement et au milieu professionnel est apporté.

Le dispositif d'enregistrement des causes de décès, ancien et exhaustif, permet une analyse spatio-temporelle de la mortalité par cancer.

À partir des données des registres, qui couvrent environ 20% de la population, les tendances des estimations d'incidence nationale et régionales sont publiées régulièrement. Les études de survie et les publications de l'incidence selon les facteurs pronostiques contribuent à l'évaluation de l'impact du dépistage des cancers.

Les données issues du Programme de médicalisation des systèmes d'information (PMSI) et des inscriptions en affection longue durée (ALD) pour cancer permettent de compléter le dispositif précédent par l'identification de ruptures récentes de tendances ainsi que par la production d'estimations d'incidence à l'échelle départementale.

Cancer surveillance in France: current tools and perspectives

This article describes the interest of information issued respectively from cancer surveillance systems, and systems whose primary objectives are not cancer surveillance, to assess cancer control strategy. It sheds light specifically on environmental and occupational cancers.

The registration system of causes of death is old and exhaustive, and allows analyses on time trends and spatial distribution.

National registry coverage is around 20%. Data from registries are used to produce regular national and regional incidence estimates. They also produce survival data and incidence by prognostic factors to contribute to the evaluation of the impact of cancer screening.

Cancer-related data from the French Hospital Information System (PMSI) and health insurance for long-term illness (LTI) complement the registries system through the identification of recent trend ruptures and the production of incidence estimates at infra-regional level.

Les perspectives portent sur le développement au niveau national, en complément des registres qui restent la référence, d'un système multi-sources de surveillance des cancers (SMSC), avec croisement automatisé des données issues des ALD, du PMSI et des données des dossiers communicants en cancérologie. En attendant le fonctionnement de ce système (d'ici au moins 5 ans), le croisement des bases PMSI et ALD est à l'étude.

While the registry system will remain the gold-standard, perspectives are based on the development of a national automated system (multi-source cancer system) by linking data from PMSI, health insurance, and medical records. This system should be operational in at least 5 years. During that period, the opportunity of cross-referencing the PMSI and health insurance databases is under consideration.

Mots clés / Key words

Surveillance, cancer, mortalité, registres, PMSI, ALD / *Epidemiological surveillance, cancer, registries, hospital discharge records, LTI*

Introduction

Dans le domaine de la lutte contre les cancers, les informations issues de la surveillance épidémiologique sont essentielles pour contribuer à la prise de décision. Les deux Plans Cancer gouvernementaux (2003-2008 puis 2009-2013) soulignent ce rôle par des mesures destinées à renforcer les dispositifs de surveillance¹.

L'analyse de la situation épidémiologique relative au cancer permet d'identifier des risques nouveaux (risques comportementaux, environnementaux, professionnels ou sociaux) ou de suivre des risques connus. Elle repose sur des indicateurs d'incidence et de mortalité. La mesure de l'ampleur et le suivi des évolutions spatio-temporelles de ces indicateurs ont pour but d'identifier d'éventuelles ruptures de tendance et de repérer des agrégats spatio-temporels pour des populations exposées à des risques particuliers (par exemple des risques environnementaux dans des zones exposées à des polluants connus ou suspectés ou des risques professionnels).

Les informations issues de la surveillance contribuent aussi à l'élaboration et à l'évaluation des politiques et actions de prévention primaire, de dépistage et de soins des patients atteints de cancer. L'impact du dépistage est mesuré par l'analyse de l'évolution de l'incidence des cancers selon des facteurs pronostiques tels que le stade, et de l'évolution de celle des cancers non invasifs (cancers *in situ*). Une politique de dépistage devrait conduire à l'augmentation de l'incidence des cancers de bon pronostic (cancers *in situ*, cancer de stade I) et à la diminution de l'incidence des cancers de mauvais pronostic (stade IV). La mesure de la survie et de son évolution permet d'évaluer l'impact des prises en charge des patients, qu'elles soient diagnostiques ou thérapeutiques. La prévalence, enfin, est utile pour évaluer les besoins de soins d'une population.

Le système actuel de surveillance des cancers repose sur les registres, dispositif de référence (voir encadré) et qui ont également une mission de recherche, ainsi que sur les données de mortalité du Centre d'épidémiologie sur les causes médicales de décès (CépiDC-Inserm) [1;2]. Alors que les données de mortalité ont une couverture nationale, les registres de cancer français couvrent actuellement environ 20% de la population (à l'exception des registres pédiatriques qui ont une couverture nationale jusqu'à l'âge de 18 ans depuis le 1^{er} janvier 2011). Afin de compléter la surveillance à partir du dispositif des registres et de la mortalité, l'apport des bases de données médico-administratives (BDMA), dont

l'objectif premier n'est pas épidémiologique, a été étudié.

Cet article présente dans un premier temps les indicateurs produits ou susceptibles d'être produits par les outils actuels (mortalité, registres, BDMA). Puis il se focalise sur les champs de la surveillance des cancers liés à l'environnement et des cancers d'origine professionnelle, qui ont des besoins de surveillance spécifiques. Il présente enfin les développements de nouveaux outils à court et à plus long terme.

Surveillance à partir des données de mortalité

L'ancienneté du dispositif d'enregistrement des causes médicales de décès par le CépiDC et sa couverture nationale permettent une analyse de la situation épidémiologique sur de longues séries temporelles et à différents échelons géographiques (national, régional, départemental voire infradépartemental) [1-3]. Bien que l'indicateur « mortalité » ne suffise pas à lui seul pour mesurer le risque de cancer, il est utile pour la formulation d'hypothèses sur des variations du risque de cancer. L'analyse de la mortalité nécessiterait toutefois d'être confrontée aux données d'incidence et de survie, mais celles-ci n'ont pas de couverture nationale.

Surveillance à partir des registres de cancer

Les données issues des registres sont rassemblées au sein d'une base de données commune (base du réseau français des registres de cancer Francim) gérée par les Hospices civils de Lyon (HCL), dans le cadre d'un partenariat qui réunit l'Institut de veille sanitaire (InVS), l'Institut national du cancer (INCa), le réseau Francim et les HCL (voir encadré page suivante). À partir des données de cette base, différents types d'indicateurs sont publiés selon une méthodologie utilisant la mortalité comme corrélat de l'incidence [4;5] :

- tous les 5 ans, estimations nationales de l'incidence et de la mortalité et leur évolution depuis 1980 [6] (mise à jour en 2012 pour les estimations 2010) ;
- tous les 5 ans, estimations régionales de l'incidence et son évolution depuis 1980 (mise à jour en 2012 pour les estimations 2010) ;
- tous les ans, projections de l'incidence et de la mortalité pour l'année en cours. Celles-ci reposent sur des scénarios d'évolution pour les années les plus récentes, puisque le délai de publication des données des registres est de 3 à 4 ans [7].

Ces indicateurs permettent à ceux qui en ont la charge de piloter et d'évaluer au niveau national et régional les actions de prévention et de prise en

charge des patients. Ils servent aussi à apprécier l'évolution dans le temps de l'impact de certains risques connus (ex : l'augmentation importante du cancer du poumon chez la femme en lien avec la consommation de tabac). Leur communication au grand public sert à sensibiliser les populations et les professionnels à la prévention primaire et au dépistage. Cependant, la méthodologie utilisée ne peut pas être déclinée au niveau départemental car les hypothèses qui sous-tendent la méthode sont peu solides à cette échelle géographique [5].

Dans les départements couverts par un registre de cancer, les données permettent de répondre à la majorité des besoins locaux du département (connaissance des risques, évaluation des actions de santé publique et des soins) y compris à une échelle géographique fine (Iris²). La géolocalisation des cas de cancer, possible en raison de l'enregistrement de l'adresse des patients au diagnostic, est essentielle pour la surveillance des cancers liés à l'environnement ou l'investigation de *clusters*. La possibilité de retour aux dossiers médicaux dans des études ciblées permet d'évaluer de façon plus précise les risques.

L'actualisation du statut vital des cas incidents des registres permet la réalisation d'études de survie. Par rapport aux études de survie réalisées à partir de séries hospitalières, les études de survie à partir des données des registres reflètent l'impact des prises en charge des patients atteints de cancer en population générale. En complément de la publication des estimations d'incidence, le programme de travail partenarial 2011-2013 relatif à la surveillance des cancers à partir des registres (encadré) prévoit une publication régulière d'indicateurs de survie et de prévalence. La première publication de la survie relative à 1, 3 et 5 ans par localisation et sous-localisation de cancer pour les cas incidents de 1989 à 1997 réalisée à partir des données des registres a été publiée en 2007 [8]. Une nouvelle publication de la survie actualisée (1989-2004) est prévue pour le 2^e trimestre 2012. Une publication relative à la prévalence est également prévue pour fin 2012.

Enfin, les registres sont des outils essentiels pour contribuer à l'évaluation de l'impact des actions de prévention primaire (par exemple la vaccination contre le papillomavirus pour les lésions précancéreuses et les cancers du col de l'utérus, les campagnes de prévention contre l'exposition aux rayonnements ultra-violet pour le mélanome cutané) et de dépistage (dépistage des cancers du sein, du côlon-rectum, du col de l'utérus, du mélanome). En effet, les registres ont la possibilité de déterminer de façon précise, à partir d'un recueil standardisé et

¹ <http://www.plan-cancer.gouv.fr>

² Ilots regroupés pour l'information statistique – <http://www.insee.fr/fr/methodes/default.asp?page=definitions/iris.htm>

Encadré La surveillance des cancers à partir des registres : description du dispositif et de la méthodologie, partenariat entre le réseau Francim, le service de biostatistique des Hospices civils de Lyon (HCL), l'Institut national du cancer (INCa) et l'Institut de veille sanitaire (InVS) [1] / *Box Cancer surveillance based on registries in France: description of the system, partnership between the FRANCIM Network, the Biostatistics Unit of the Hospices Civils de Lyon, the National Cancer Institute (INCa), and the Institute for Public Health Surveillance (InVS) [1]*

Les premiers registres de cancer ont été créés au milieu des années 1970. La couverture des registres s'est progressivement étendue. Avant le premier Plan Cancer gouvernemental, elle était de 14%. Le premier Plan Cancer (2003-2008) a permis l'extension de cette couverture qui s'est concrétisée par la qualification de deux registres en zone urbaine. Le 2e Plan Cancer (2009-2013) a prévu un renforcement du dispositif existant (mesure 7.1) sans extension de la couverture, jugeant que celle-ci était suffisante pour la surveillance nationale des cancers.

La couverture actuelle de la population par les registres est d'environ 20% (elle varie selon la localisation des cancers). Au 1er janvier 2011, le nombre de registres généraux métropolitains et ultramarins qualifiés par le Comité national des registres était de 14, et celui de registres spécialisés d'organe (appareil digestif, sein et organes gynécologiques, hémopathies malignes, système nerveux central) de 8. Le dispositif des registres généraux et spécialisés est complété par 2 registres nationaux pédiatriques : le registre des tumeurs solides et celui des hémopathies malignes. Ils sont au nombre de 2 pour des raisons historiques, mais leur base de données est commune. Leur couverture est étendue depuis le 1er janvier 2011 jusqu'à l'âge de 18 ans. Enfin, le registre multicentrique du mésothéliome couvre 22 départements. L'activité des registres est évaluée sur les plans de la surveillance et de la recherche tous les 4 ans par le Comité national des registres (<http://www.invs.sante.fr/Espace-professionnels/Comite-national-des-registres>).

La méthodologie d'enregistrement vise à atteindre l'exhaustivité des nouveaux diagnostics de cancer de la population d'un territoire donné par le croisement de plusieurs sources de données (principalement laboratoires d'anatomie et de cytologie pathologiques, établissements de soins et réseaux de cancérologie, assurance maladie). Cette méthodologie, qui suit des règles nationales et internationales, doit permettre une standardisation et une exhaustivité des informations recueillies pour chaque cas. C'est pourquoi les registres sont considérés comme l'outil de référence.

Les données des registres généraux et spécialisés d'organe sont rassemblées dans une base de données commune administrée par le service de biostatistiques des HCL. Avant d'être intégrées à cette base, les données font l'objet de contrôles de cohérence. La période d'incidence couverte est variable selon les registres, certains ayant débuté l'enregistrement en 1975 et d'autres en 2002. Au 31 décembre 2011, la dernière année disponible était l'année 2008 pour la majorité des registres. Le délai entre la dernière année enregistrée et l'année en cours est dû aux nécessaires validations des données, ce qui a pour conséquence un délai de 3 à 4 ans dans la publication des incidences observées dans la zone couverte par les registres. Disposer en France d'une base de données commune pour les données issues des registres est un atout important qui contribue notamment à un meilleur contrôle qualité des données, et à une production harmonisée d'indicateurs ou de résultats lors d'interrogations sur des thèmes spécifiques.

Le réseau des registres de cancer Francim, le service de biostatistique des HCL, l'INCa et l'InVS ont élaboré un programme de travail partenarial, coordonné par l'InVS en lien avec l'INCa, afin de répondre aux mesures du Plan Cancer relatives à la surveillance à partir des registres. Les principales actions destinées à renforcer le dispositif des registres portent en particulier sur la réduction d'un an du délai de publication des données (ceci devrait être atteint par tous les registres en 2012), sur l'enregistrement d'informations complémentaires (comme le stade des cancers), sur l'actualisation du statut vital pour réaliser des études de survie ainsi que sur la facilitation de l'accès aux données. Ce dernier point est essentiel pour les registres afin d'assurer l'exhaustivité et la qualité des données. Malheureusement, en France, la contribution des sources de données au fonctionnement des registres n'est pas toujours effective, ce qui contribue à la fragilisation du dispositif. Un comité de suivi et un conseil scientifique encadrent le déroulement des actions du programme partenarial qui s'étend sur la période 2011-2013. Le programme est annexé à un accord-cadre qui définit les modalités du partenariat. Les registres reçoivent une subvention de fonctionnement de l'INCa et de l'InVS pour assurer leur activité de surveillance.

exhaustif, le stade des cancers invasifs, et d'enregistrer les cancers *in situ* ou les états précancéreux (par exemple, les lésions de haut grade pour le cancer du col de l'utérus). Cet enregistrement est toutefois complexe car il nécessite des définitions précises de cas et repose sur des comptes-rendus qui doivent comporter précisément l'information recherchée. L'enregistrement en routine de cas supplémentaires (cancer *in situ*, lésions précancéreuses), ou d'informations complémentaires pour les cas invasifs (stade ou autres facteurs pronostiques tel que l'indice de Breslow et le niveau de Clark pour le mélanome, le PSA pour le cancer de la prostate) sont des objectifs à atteindre par les registres dans le cadre du programme de travail partenarial (encadré). Cet enregistrement est déjà effectif pour le mélanome, le cancer de la prostate, le cancer de la thyroïde (taille de la tumeur). Pour les cancers du sein, du côlon et du rectum, les résultats d'une étude de faisabilité de l'enregistrement en routine du stade seront communiqués début 2012. Pour le col de l'utérus, une étude sur la faisabilité de l'enregistrement des lésions de haut grade et des cancers *in situ* est en cours.

Surveillance à partir des bases de données médico-administratives

Les BDMA utilisées pour la surveillance des cancers sont, d'une part, les données hospitalières du Programme de médicalisation des systèmes

d'information (PMSI) et, d'autre part, les données d'inscription en affection de longue durée pour cancer (ALD 30). Par rapport aux registres, ces bases ont un intérêt particulier pour la surveillance car leur couverture est nationale et les données sont disponibles assez rapidement (1 à 2 ans). Néanmoins, elles ont des objectifs propres, non épidémiologiques et ne peuvent être considérées comme exhaustives. En effet, en ce qui concerne le PMSI, certains cancers comme par exemple le mélanome sont pris en charge essentiellement en ambulatoire et échappent à l'enregistrement. De même, certains cas peuvent ne pas faire l'objet d'une demande d'ALD [9;10]. Par ailleurs, les BDMA sont soumises à des évolutions réglementaires qui doivent être prises en compte dans l'interprétation des résultats d'études menées à des fins épidémiologiques [11].

Les BDMA sont utilisées pour la surveillance des cancers selon trois modalités qui dépendent de l'objectif recherché :

- utilisation des BDMA sans correction par l'incidence observée dans la zone « registre », pour identifier les ruptures de tendances récentes de l'incidence [9;12] ;
- utilisation des BDMA avec modélisation de l'incidence observée dans la zone « registre » pour produire des estimations de l'incidence à l'échelle infranationale [10;13] ;

- croisement avec les données des registres pour étudier la capacité des BDMA à repérer les cas incidents [14;15].

Les exemples ci-dessus concernent les bases PMSI ou ALD utilisées séparément. L'utilisation d'une base croisée ALD/PMSI par chaînage des données individuelles à partir d'un identifiant commun est à l'étude à l'InVS (cf. § Perspectives).

L'utilisation des BDMA est d'intérêt limité pour l'évaluation des actions de prévention, de dépistage et de soins, car elles n'apportent pas d'informations sur les facteurs pronostiques tels que le stade, ni sur le statut vital. De plus, le codage des cancers y est effectué selon la Classification internationale des maladies 10^e révision (CIM10), moins précise que la Classification internationale des maladies appliquée à l'oncologie (CIMO 3) utilisée dans les registres.

Surveillance des cancers liés à l'environnement

Les registres de cancer de l'enfant sont des outils de choix pour la surveillance des cancers liés à l'environnement car ils ont une couverture nationale. De plus, la plupart des cancers pédiatriques sont suspectés d'être liés à des expositions environnementales précoces, fœtales, post-natales voire pré-conceptionnelles, et leur latence est faible. Deux registres recensent ces cancers : le registre national des hémopathies malignes de l'enfant depuis 1990 et le registre national des tumeurs solides de l'enfant depuis 2000. En 2011, la couverture a été étendue jusqu'à 18 ans. La géolocalisation exhaustive des cas, réalisée à partir de l'adresse précise des patients, permet d'étudier et potentiellement de détecter les agrégats spatio-temporels aux niveaux national et local (projet Géocap de l'unité Inserm 754 « Épidémiologie environnementale des cancers » en lien avec l'InVS). Elle permet également de rechercher des associations entre la survenue de ces cancers et des expositions environnementales géographiquement déterminées (lignes à haute tension, proximité d'installations nucléaires, radon, trafic routier).

Les registres généraux et spécialisés du réseau Francim étudient la faisabilité d'un système centralisé de géolocalisation des cas, qui permettrait de répondre également chez les adultes aux questions relatives au risque de cancer lié à l'environnement.

Les besoins de la santé environnementale dans le domaine du cancer sont importants et, malgré des progrès notables et continus, le système actuel n'est souvent pas en mesure d'y répondre. Les besoins sont pressants autour d'installations industrielles ou émettrices particulières (installations nucléaires, projets de grandes infrastructures comme une ligne à très haute tension, etc.). Ils justifieraient la mise en place d'une surveillance à couverture nationale telle que le Système multi-sources de surveillance des cancers - SMSC (cf. § Perspectives) avec une résolution à échelle infradépartementale. Pour répondre à de telles préoccupations dans des zones non couvertes par les registres, l'InVS est actuellement amené à mettre en place des études *ad hoc* [16].

En ce qui concerne l'utilisation des BDMA pour la santé environnementale, les besoins seraient au minimum de disposer par individu d'une

géolocalisation de sa résidence au moment du diagnostic, ce qui n'est pas le cas. De plus, les BDMA n'enregistrent pas de manière exhaustive les cas, ce qui est essentiel. Par ailleurs, il peut s'avérer nécessaire pour certains cancers rares, notamment ceux considérés comme sentinelles d'une exposition, de mettre en place une surveillance avec retour au cas et enquête individuelle. C'est ce qui motive le passage en déclaration obligatoire (DO) du mésothéliome (voir ci-après), mais également la mise en place de la surveillance des angiosarcomes du foie en lien avec l'exposition au chlorure de vinyle monomère d'origine environnementale [17].

Surveillance des cancers d'origine professionnelle

Concernant la surveillance des cancers d'origine professionnelle, le système actuel ne permet pas la production des taux d'incidence par profession et secteur d'activité. En effet, d'une part la profession ne peut être recueillie par les registres car elle ne figure pas dans les dossiers médicaux comme l'a montré une étude de faisabilité [18] et, d'autre part, l'Insee ne fournit pas de données de population par profession et secteur d'activité.

Seul le mésothéliome pleural fait l'objet d'une surveillance particulière par le Programme national de surveillance des mésothéliomes (PNSM). En fonctionnement depuis 1998, le PNSM permet la surveillance des mésothéliomes pleuraux dans 23 départements métropolitains. Il présente la particularité d'allier en routine 1) les objectifs d'expertises diagnostiques (anatomopathologique, clinique), 2) les expositions, notamment professionnelles, à l'amiante (questionnaires d'exposition) et 3) les informations médico-sociales. En termes de surveillance, il permet d'estimer l'incidence nationale des mésothéliomes et d'étudier la proportion des cas attribuables à une exposition professionnelle à l'amiante [19]. Le PNSM va être renforcé par un système réglementaire déclaratif obligatoire, la DO des mésothéliomes, suite au test réalisé au cours du 1^{er} semestre 2011 et jugé satisfaisant (98 notifications pour 120 attendues). La DO va étendre le champ de la surveillance à tous les sites anatomiques de mésothéliome, sur tout le territoire national métropolitain et ultramarin, avec une vigilance particulière pour les mésothéliomes survenant chez les femmes, les moins de 50 ans et les mésothéliomes du péritoine. La DO devrait contribuer à améliorer la connaissance des cas survenant en dehors d'une exposition professionnelle à l'amiante pour mieux comprendre le rôle possible d'expositions environnementales [20].

Perspectives

Le développement à l'échelle nationale du SMSC constitue l'un des enjeux majeurs de la surveillance des cancers en France. Ce projet a fait l'objet d'une étude pilote [21;22]. Le SMSC repose sur le croisement passif et automatisé des données individuelles de trois bases nationales grâce à un identifiant unique. Les deux premières bases sont celles issues du PMSI et des ALD. La troisième base sera constituée des données issues des Dossiers communicants de cancérologie (DCC). Le SMSC devrait permettre le repérage de cas incidents, leur géolocalisation, ainsi que de disposer d'informations plus

précises sur les cancers (morphologie, stade...). Il pourra aussi être croisé avec d'autres bases de données individuelles à condition de disposer d'identifiants communs (par exemple avec des bases de l'assurance vieillesse pour connaître le parcours professionnel des patients ou avec des cohortes) ou encore des bases de données agrégées (par exemple pour réaliser des études de type écologique).

Le développement du SMSC est toutefois complexe et dépendant de l'avancement de la mise en place des DCC/DMP (dossier médical partagé) au niveau des régions. Le projet, piloté par l'INCa et l'Agence des systèmes d'information partagés de santé (Asip-Santé), est actuellement expérimenté dans 7 régions pilotes. Le SMSC ne fonctionnera donc pas avant plusieurs années (délai estimé au minimum à 5 ans).

Compte tenu des délais de développement du SMSC, l'intérêt d'utiliser la base croisée ALD/PMSI a été étudié. Par rapport aux bases utilisées séparément, le chaînage des données individuelles ALD et PMSI à partir d'un identifiant commun a pour objectif d'augmenter la probabilité d'enregistrement des cas susceptibles d'être incidents et la probabilité de repérage des cas prévalents à éliminer. Des travaux exploratoires ont été conduits à l'échelle nationale pour 8 localisations de cancers (sein, côlon-rectum, système nerveux central, col de l'utérus, corps de l'utérus, poumon, rein, thyroïde) [23] et se poursuivront pour l'ensemble des cancers en 2012. Les utilisations possibles de la base croisée sont identiques à celles des bases utilisées séparément, à savoir l'étude des tendances pour identifier des ruptures de tendances récentes et la production d'estimations d'incidence à l'échelle infra nationale. La plus-value d'utiliser la base croisée par rapport aux bases séparées devra être évaluée.

Conclusion

Cet article décrit l'intérêt, pour la surveillance des cancers, des données de mortalité, de celles issues des registres de cancer et des bases médico-administratives (ALD et PMSI). La stratégie française en matière de surveillance des cancers (2^e Plan Cancer 2009-2013) consiste, d'une part, à renforcer le dispositif des registres existants, considérés comme la référence en raison de leur méthodologie, sans en étendre la couverture qui a été jugée suffisante pour la surveillance nationale (les registres pédiatriques constituant un cas particulier puisqu'ils ont une couverture nationale de 0 à 18 ans) et, d'autre part, à développer un système à couverture nationale automatisé, le SMSC. Par ailleurs, d'autres dispositifs complémentaires (DO et surveillance de cancers rares) visent à répondre à des questions plus ciblées sur les cancers d'origine environnementale ou professionnelle. Le dispositif des registres constitue un outil essentiel pour la surveillance des cancers non seulement pour l'intérêt des informations produites mais aussi pour l'étalonnage des autres dispositifs utilisés actuellement (ALD, PMSI) et, dans l'avenir, du SMSC.

Remerciements

À Pascale Grosclaude, présidente du réseau français des registres de cancer Francim et directrice du registre des cancers du Tarn, pour sa relecture. À tous les acteurs du système de surveillance des cancers.

Références

- [1] Dossier thématique « Surveillance épidémiologique des cancers ». Site Internet de l'Institut de veille sanitaire. Disponible à : <http://www.invs.sante.fr/Dossiers-thematiques/Maladies-chroniques-et-traumatismes/Cancers/Surveillance-epidemiologique-des-cancers>
- [2] Site Internet du CépiDC-Inserm : <http://www.cepidc.vesinet.inserm.fr/inserm/html/index2.htm>
- [3] Site Internet de la Fédération nationale des Observatoires régionaux de la santé : <http://www.score-sante.org/score2008/contextes.html>
- [4] Remontet L, Belot A, Bossard N. Tendances de l'incidence et de la mortalité par cancer et projections pour l'année en cours : méthodes d'estimation et rythme de production. Bull Epidemiol Hebd. 2009;(38):405-8.
- [5] Colonna M, Bossard N, Mitton N, Remontet L, Belot A, Delafosse P, et al. ; le réseau Francim. Éléments d'interprétation des estimations régionales de l'incidence du cancer en France sur la période 1980-2005. Rev Epidemiol Santé Publique. 2008;(56):434-40.
- [6] Belot A, Velten M, Grosclaude P, Bossard N, Launoy G, Remontet L, et al. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2005. Saint-Maurice: Institut de veille sanitaire; 2008.132p. Disponible à : http://www.invs.sante.fr/publications/2009/estimation_cancer_1980_2005/index.html
- [7] Projection de l'incidence et de la mortalité par cancer en France en 2011. Rapport technique. Saint-Maurice: Institut de veille sanitaire; 2011. 78 p. Disponible à : <http://www.invs.sante.fr/Publications-et-outils/Rapports-et-syntheses/Maladies-chroniques-et-traumatismes/2011/Projection-de-l-incidence-et-de-la-mortalite-par-cancer-en-France-en-2011>
- [8] Réseau Francim. Survie des patients atteints de cancer en France. Étude des registres de cancer du réseau Francim. Paris: Springer-Verlag France; 2007.
- [9] Uhry Z, Remontet L, Grosclaude P, Belot A, Marc Colonna M, Nadine Bossard N, et al., Tendances récentes des données d'affections de longue durée (ALD) : intérêt pour la surveillance nationale de l'incidence des cancers, période 1997-2009, France. Bull Epidemiol Hebd. 2012;(5-6):58-63.
- [10] Uhry Z, Remontet L, Colonna M, Belot A, Grosclaude P, Mitton N, et al. Évaluation de l'utilisation des données d'affection de longue durée (ALD) pour estimer l'incidence départementale des cancers pour 24 localisations cancéreuses, France, 2000-2005. Bull Epidemiol Hebd. 2012;(5-6):71-7.
- [11] Uhry Z, Remontet L, Grosclaude P, Velten M, Mitton N, Kudjawan Y, et al. Intérêt des données hospitalières 2002-2008 pour la surveillance de l'incidence du cancer de la thyroïde en France. Saint-Maurice: Institut de veille sanitaire; 2011. 25 p. Disponible à : http://www.invs.sante.fr/publications/2011/incidence_cancer_thyroïde/index.html
- [12] Uhry Z, Remontet L, Grosclaude P, Belot A, Colonna M, Boussac-Zarebska M, et al. Tendances récentes des données d'affections de longue durée : intérêt pour la surveillance nationale de l'incidence des cancers - Période 1997-2009. Saint-Maurice: Institut de veille sanitaire; 2011. 40 p. Disponible à : <http://www.invs.sante.fr/Publications-et-outils/Rapports-et-syntheses/Maladies-chroniques-et-traumatismes/2011/Tendances-recentes-des-donnees-d-affections-de-longue-duree-interet-pour-la-surveillance-nationale-de-l-incidence-des-cancers>
- [13] Bossard N, Gomez F, Remontet L, Uhry Z, Olive F, Mitton N, et al. Utilisation des données du PMSI pour estimer l'incidence des cancers en France à l'échelon infranational : la démarche du groupe Oncepi. Bull Epidemiol Hebd. 2012;(5-6):77-80.
- [14] Grosclaude P, Dentan C, Trétarre B, Velten M, Fournier E, Molinié F. Utilité des bases de données médico-administratives pour le suivi épidémiologique des cancers. Comparaison avec les données des registres au niveau individuel. Bull Epidemiol Hebd. 2012;(5-6):68-71.

[15] Lauzeille D, Dentan C, Grosclaude P, Marchand JL, Chérié-Challine L. Les données d'exonération du ticket modérateur pour affections de longue durée (ALD), source potentielle pour la surveillance des cancers d'origine professionnelle ? Bull Epidemiol Hebd. 2012;(5-6):68-71.

[16] Incidence des cancers à proximité des usines d'incinération d'ordures ménagères. Exposition aux incinérateurs pendant les années 1970-1980. Résultats définitifs. Saint-Maurice: Institut de veille sanitaire; 2008. 4 p. Disponible à : http://www.invs.sante.fr/publications/2008/plaquette_resultats_uiom/index.html

[17] Galey C, Le Moal J. Chlorure de vinyle monomère. Note de position. Saint-Maurice: Institut de veille sanitaire; juillet 2010. 2 p. Disponible à : [http://www.invs.sante.fr/pmb/invs/\(id\)/PMB_586](http://www.invs.sante.fr/pmb/invs/(id)/PMB_586)

[18] Lauzeille D. Recueil d'informations professionnelles auprès de patients atteints de cancer. Résultats

de deux études pilotes menées avec les registres dans les départements de l'Hérault et de Vendée. Saint-Maurice: Institut de veille sanitaire; 2011. 26 p. Disponible à : <http://www.invs.sante.fr/Publications-et-outils/Rapports-et-syntheses/Travail-et-sante/2011/Recueil-d-informations-professionnelles-aupres-de-patients-atteints-de-cancer>

[19] Gilg Soit Ilg A, Chamming's S, Rolland P, Ducamp S, Brochard P, Galateau-Sallé F, et al. Programme national du mésothéliome : principaux résultats, France, 1998-2004. Bull Epidemiol Hebd. 2007;(41-42):350-4.

[20] Chérié-Challine L, Imbernon E. Déclaration obligatoire des mésothéliomes : une phase pilote satisfaisante. Info respiration n°104. Août-septembre 2011:8-9. Disponible à <http://www.splf.org/s/spip.php?article1766>

[21] Caserio-Schönemann C, Kudjawa Y, Chérié-Challine L, Guillet A, Musset A, Nicolau J, et al. Système multi-sources de surveillance des cancers

(SMSC). Étude pilote sur le cancer de la thyroïde en Ile-de-France et dans le Nord-Pas-de-Calais. Rapport final d'étude, octobre 2010. Saint-Maurice: Institut de veille sanitaire; 2011. 68 p. Disponible à : <http://www.invs.sante.fr/Publications-et-outils/Rapports-et-syntheses/Maladies-chroniques-et-traumatismes/2011/Systeme-multi-sources-de-surveillance-des-cancers-SMSC>

[22] Kudjawa Y, Caserio-Schönemann C, Guillet A, Nicolau J, Musset A, Rudant J. Système multi-sources de surveillance des cancers (SMSC) : bilan de l'étude pilote. Bull Epidemiol Hebd. 2012;(5-6):81-4.

[23] Kudjawa Y, Rudant J, Decool E, Danzon A, Grémy I. Système multi-sources de surveillance des cancers (SMSC), croisement des données ALD et PMSI : description des données 2006-2008 et construction d'indicateurs. Saint-Maurice: Institut de veille sanitaire; 2012. (sous presse).

Tendances récentes des données d'affections de longue durée (ALD) : intérêt pour la surveillance nationale de l'incidence des cancers, période 1997-2009, France

Zoé Uhry (z.uhry@invs.sante.fr)¹, Laurent Remonet², Pascale Grosclaude^{3,4}, Aurélien Belot^{1,2}, Marc Colonna^{3,5}, Nadine Bossard², Marjorie Boussac-Zarebska¹, Solène Delacour-Billon^{3,6}, Julie Gentil^{3,7}, Michelle Altana⁸, François Frete⁹, Alain Weill¹⁰, Agnès Rogel¹

1/ Institut de veille sanitaire, Saint-Maurice, France

3/ Réseau français des registres de cancer Francim, France

5/ Registre du cancer de l'Isère, Grenoble, France

7/ Registre des cancers gynécologiques de Côte-d'Or, Dijon, France

8/ Direction de la gestion des risques et de l'action sociale, Régime social des indépendants, La Plaine-Saint-Denis, France

9/ Échelon national du contrôle médical, Mutualité sociale agricole, Bagnolet, France

10/ Direction de la stratégie, des études et des statistiques, Caisse nationale d'assurance maladie des travailleurs salariés, Paris, France

2/ Service de biostatistique, Hospices civils de Lyon, Lyon, France

4/ Registre des cancers du Tarn, Toulouse, France

6/ Registre des cancers de Loire-Atlantique et de Vendée, Nantes, France

Résumé / Abstract

Introduction – Cette étude présente les évolutions nationales du taux d'admission en affection de longue durée (ALD) de 1997 à 2009 pour 24 localisations cancéreuses. Son objectif était de confronter les évolutions entre taux d'admissions (taux d'ALD) et taux d'incidence nationale sur la période 1997-2006 et d'identifier les ruptures de tendances récentes des taux d'ALD.

Matériel et méthode – Les évolutions des taux d'ALD sont présentées de 1997 à 2009. Elles sont comparées graphiquement à celles du taux d'incidence nationale estimé, issu d'une modélisation à partir de données observées jusqu'en 2006 et d'un modèle de projection au-delà jusqu'en 2011.

Résultats – Sur la période 1997-2006, les évolutions du taux d'ALD et du taux d'incidence modélisé étaient relativement similaires, sauf pour les cancers de la vessie et, chez l'homme, du larynx. Les taux d'ALD présentaient des ruptures de tendance, à partir du milieu des années 2000, pour les cancers du sein, de la prostate et de la thyroïde, après plusieurs années de croissance soutenue.

Conclusion – Les évolutions des taux d'ALD doivent être interprétées avec prudence car elles peuvent être affectées par des évolutions réglementaires ou par une évolution de la fréquence des demandes d'ALD parmi les personnes atteintes d'un nouveau cancer. Ces données sont toutefois disponibles rapidement et, de ce fait, très utiles à la surveillance nationale de l'incidence des cancers. Elles permettent d'identifier des ruptures de tendances avec réactivité, comme illustré pour les cancers du sein, de la prostate et de la thyroïde. Pour les cancers du sein et de la prostate, les évolutions sont cohérentes avec les connaissances épidémiologiques et les données internationales. Pour le cancer de la thyroïde, les données internationales ne permettaient pas d'anticiper cette évolution.

Recent trends in long-term illness (LTI) health insurance data: relevance for monitoring recent national trends in cancer incidence, France, 1997-2009

Introduction – This study presents the national trends in the rate of long-term illness (LTI) agreements issued from the health insurance database from 1997 to 2009 for 24 cancer sites. The study aimed at confronting trends in LTIs agreements and national incidence rates on the period 1997-2006, and identifying recent changes in the trends of LTI rates.

Material and method – Trends in LTI rates are presented from 1997 to 2009. They are compared graphically with the trends in national incidence rates, issued from a modelling based on observed data up to 2006 and a projection model beyond up to 2011.

Results – Over the period 1997-2006, trends in ALD and incidence rates were relatively consistent, except for bladder cancers and, for men, larynx cancers. Trends in LTI rates recently changed for breast, prostate and thyroid cancers, in the mid 2000, after several years of sustained increase.

Conclusion – The trends in LTI rates should be interpreted with caution, since they may be affected by regulatory changes or variations in the frequency of ALD requests among persons with a new cancer. These data are however available rapidly and thus useful for national cancer surveillance. They provide clues to recent changes in cancer incidence trends with reactivity, as illustrated by breast, prostate, and thyroid cancers. For breast and prostate cancer, these trends are consistent with epidemiological knowledge and international data. For thyroid cancer, this evolution could not be anticipated from international data.

Mots clés / Key words

Cancer, affection de longue durée, incidence, tendances, France / Cancer, long-term illness, health insurance data, incidence, trends, France

Introduction

La surveillance des cancers en France est assurée par les registres départementaux de cancer, qui couvrent actuellement environ 20% de la population. Leurs données sont disponibles dans un délai de trois à quatre ans. Les estimations nationales de l'incidence sont réalisées en extrapolant à la France le rapport entre incidence et mortalité observé sur la zone couverte par l'ensemble des registres [1]. De plus, des projections à court terme sont nécessaires pour fournir des estimations pour l'année en cours [2;3].

Le dispositif des affections de longue durée (ALD) des caisses d'assurance maladie permet d'améliorer la prise en charge financière des soins liés à une maladie chronique lourde. Les données d'admissions en ALD sont disponibles rapidement (délai d'un an) et peuvent apporter une information sur les tendances récentes de l'incidence des cancers. Cette étude s'intéresse aux tendances et non au niveau des taux d'admissions en ALD (taux d'ALD) : la différence entre le niveau des taux d'ALD et des taux d'incidence est sans conséquence pour suivre les tendances, dès lors qu'elle est constante dans le temps.

Cette étude avait deux objectifs : d'une part, confronter les évolutions des taux d'ALD et des taux d'incidence sur la période 1997-2006 et, d'autre part, identifier les ruptures de tendances des taux d'ALD sur la période 2006-2009. Les résultats de cette étude sont mis à jour chaque année [4].

Matériel et méthode

Données ALD

L'admission en ALD pour cancer permet l'exonération du ticket modérateur pour les soins liés au cancer. Cette exonération est accordée pour cinq ans renouvelables. Le diagnostic est codé sur 3 caractères selon la Classification internationale des maladies 10^e révision (CIM10) depuis 2000, et était codé selon la 9^e révision auparavant (CIM9). Depuis 1997, les principales évolutions du dispositif des ALD ont été : l'instauration de la majoration du maintien à domicile pour les personnes âgées de plus de 75 ans (2000) ; la rémunération de 40 € par patient en ALD pour le médecin traitant (2005) et surtout, le nouveau parcours de soins pour l'admission en ALD (2006). Dans ce cadre, le médecin traitant doit établir un protocole de soins, qui définit les actes et prestations remboursés à 100% au titre de l'ALD (auparavant, tout médecin soignant pouvait faire une demande d'admission en ALD). Enfin, à partir de 2010, une modification des critères de renouvellement exclut le maintien de l'ALD du cancer au delà de cinq ans si aucun traitement n'est indiqué.

Les données d'admission en ALD proviennent de la Caisse nationale d'assurance maladie des travailleurs salariés (CnamTS), qui couvrait environ 84% de la population française en 2001 [5]. Les taux annuels nationaux d'admission en ALD ont été calculés de 1997 à 2009 pour 24 localisations cancéreuses (tous âges) [4]. Les nombres d'admissions en ALD pour le régime général (CnamTS) sont rapportés à l'ensemble de population (personnes-années calculées à partir des données de l'Insee). Les leucémies ne sont pas

étudiées car les différents types (leucémies lymphoïdes chroniques, leucémies aigües) ne peuvent être différenciés dans les données ALD analysées.

Estimations nationales de l'incidence des cancers

Les estimations nationales de l'incidence sont issues des projections 2011 [2;3]. Ces projections reposaient sur une analyse des tendances de l'incidence et de la mortalité depuis 1975 jusqu'à la dernière année disponible, et sur un modèle de projections pour les années ultérieures. Les données analysées pour les projections 2011 concernaient la période 1975-2006 pour l'incidence, et 1975-2008 pour la mortalité. Trois scénarios de projections sont explorés : (A) poursuite de la tendance récente ; (B) taux constant depuis 2005¹ ; (C) poursuite de la tendance long terme. Des épidémiologistes des registres de cancers privilégient un scénario parmi les trois, d'après les connaissances épidémiologiques sur l'évolution des facteurs de risque, les évolutions observées dans d'autres pays et les informations externes disponibles, et en particulier les données ALD (de 1997 à 2009 pour les projections 2011).

Méthode

Les tendances des taux d'ALD sont comparées graphiquement aux tendances des taux d'incidence estimés et projetés, en distinguant la période d'observation des données de registres (1997-2006) et la période de projection (après 2006). Les taux sont standardisés sur la structure d'âge de la population mondiale (TSM). Les effectifs correspondants ne sont pas présentés ici mais sont disponibles et publiés [3;4].

Résultats

La figure 1 présente les tendances des taux d'ALD et les taux d'incidence estimés et projetés, chez l'homme (figure 1a) et la femme (figure 1b). Les évolutions de ces deux indicateurs sont relativement cohérentes sur la période 1997-2006 pour la plupart des localisations cancéreuses, à l'exception des cancers de la vessie et du larynx chez l'homme. Trois localisations cancéreuses présentent une nette rupture de tendance du taux d'ALD à partir du milieu des années 2000 (sein, prostate, thyroïde), après des années de croissance soutenue.

La figure 2 présente, pour ces trois cancers, les tendances des taux d'ALD par groupe d'âge (20-49 ans, 50-74 ans, ≥75 ans). Une inversion de tendance est observée pour le cancer du sein chez les femmes âgées de 50 à 74 ans, ainsi que pour le cancer de la prostate chez les hommes âgés de plus de 75 ans depuis 2005 et, plus récemment, chez les hommes âgés de 50 à 74 ans depuis 2008. Pour le cancer de la thyroïde, on observe, pour les deux sexes, un ralentissement de la croissance chez les personnes âgées de 50 à 74 ans, plus net chez les femmes, et une stabilisation pour les personnes âgées de 20 à 49 ans.

¹ L'année 2005 a été choisie comme référence pour le scénario B, pour assurer une cohérence avec les analyses détaillées des tendances de l'incidence et de la mortalité, qui sont actualisées tous les 5 ans et concernent actuellement la période 1980-2005.

Discussion

Cette étude a présenté le taux d'ALD de 1997 à 2009 et comparé sur la période 1997-2006 les évolutions du taux d'ALD et du taux d'incidence modélisé. Les taux d'ALD présentent des ruptures de tendance, à partir du milieu des années 2000, pour les cancers du sein, de la prostate et de la thyroïde. Les données d'ALD permettent d'identifier ces ruptures avec réactivité, dans un délai de 1 à 2 ans. Les taux pour les trois principaux régimes (général, agricole et indépendants) sont disponibles depuis 2000 et décrits dans un rapport plus détaillé [4]. Seuls les taux du régime général sont présentés ici car le recul est plus important et leurs évolutions sont similaires à celles du taux d'ALD interrégimes [4]. L'écart entre taux d'ALD et taux d'incidence est très variable selon la localisation cancéreuse mais, comme évoqué en introduction, cette étude s'intéresse aux évolutions et non au niveau des taux d'ALD.

Cette étude repose sur de simples comparaisons graphiques et seuls les phénomènes majeurs sont interprétables. La prochaine mise à jour sera étoffée par une analyse statistique de la stabilité dans le temps du rapport entre taux d'ALD (interrégimes) et taux d'incidence, directement observés dans les départements avec registre.

Il faut rester prudent dans l'interprétation des évolutions du taux d'ALD, car elles peuvent être affectées par des évolutions réglementaires ou par une évolution de la fréquence de demande d'ALD parmi les personnes atteintes d'un nouveau cancer. Certaines évolutions réglementaires ont entraîné des *artefacts* visibles, comme l'augmentation du taux d'ALD chez les personnes âgées de plus de 75 ans en 2000-2001, observée pour plusieurs localisations cancéreuses [4]. Les demandes d'ALD ne coïncident que partiellement avec les cancers incidents [6]. La fréquence des demandes d'admission en ALD, parmi les cas incidents de cancers, peut varier notamment avec les évolutions de la gravité des cancers au moment du diagnostic et de la lourdeur des traitements.

Les évolutions des taux d'ALD et des taux d'incidence modélisés sont relativement similaires sur la période 1997-2006 pour la plupart des localisations cancéreuses, à l'exception des cancers de la vessie, et, dans une moindre mesure, des cancers du larynx chez l'homme. Toutefois, elles diffèrent légèrement pour quelques localisations cancéreuses (côlon-rectum, col de l'utérus, foie, maladie de Hodgkin). Par ailleurs, pour les lymphomes malins non-hodgkiniens et les cancers du corps de l'utérus, le taux d'ALD augmente brusquement entre 1997 et 1999. Cette brusque augmentation ne correspond pas une évolution réelle de l'incidence. La divergence d'évolution entre taux d'ALD et taux d'incidence des cancers de la vessie est probablement liée à l'inclusion de tumeurs non infiltrantes (pTa, *in situ*) dans les ALD alors que l'incidence est restreinte aux tumeurs infiltrantes. Les données d'ALD incluent aussi des tumeurs *in situ* ou bénignes pour d'autres localisations cancéreuses (système nerveux central, sein, col de l'utérus, etc.), mais ce phénomène est particulièrement marqué pour les tumeurs de la vessie [6]. Il est en effet parfois difficile au début de la prise en charge d'une

Figure 1a Taux annuels d'incidence 1997-2011 (source : projections 2011, Francim-Hospices civils de Lyon) et d'admissions en ALD 1997-2009, standardisés sur l'âge de la population mondiale, hommes, échelle logarithmique, France / Figure 1a Annual incidence rates 1997-2011 (Source: projections Francim-HCL 2011) and LTI rates 1997-2009, age-standardised to the world population, men, logarithmic scale, France

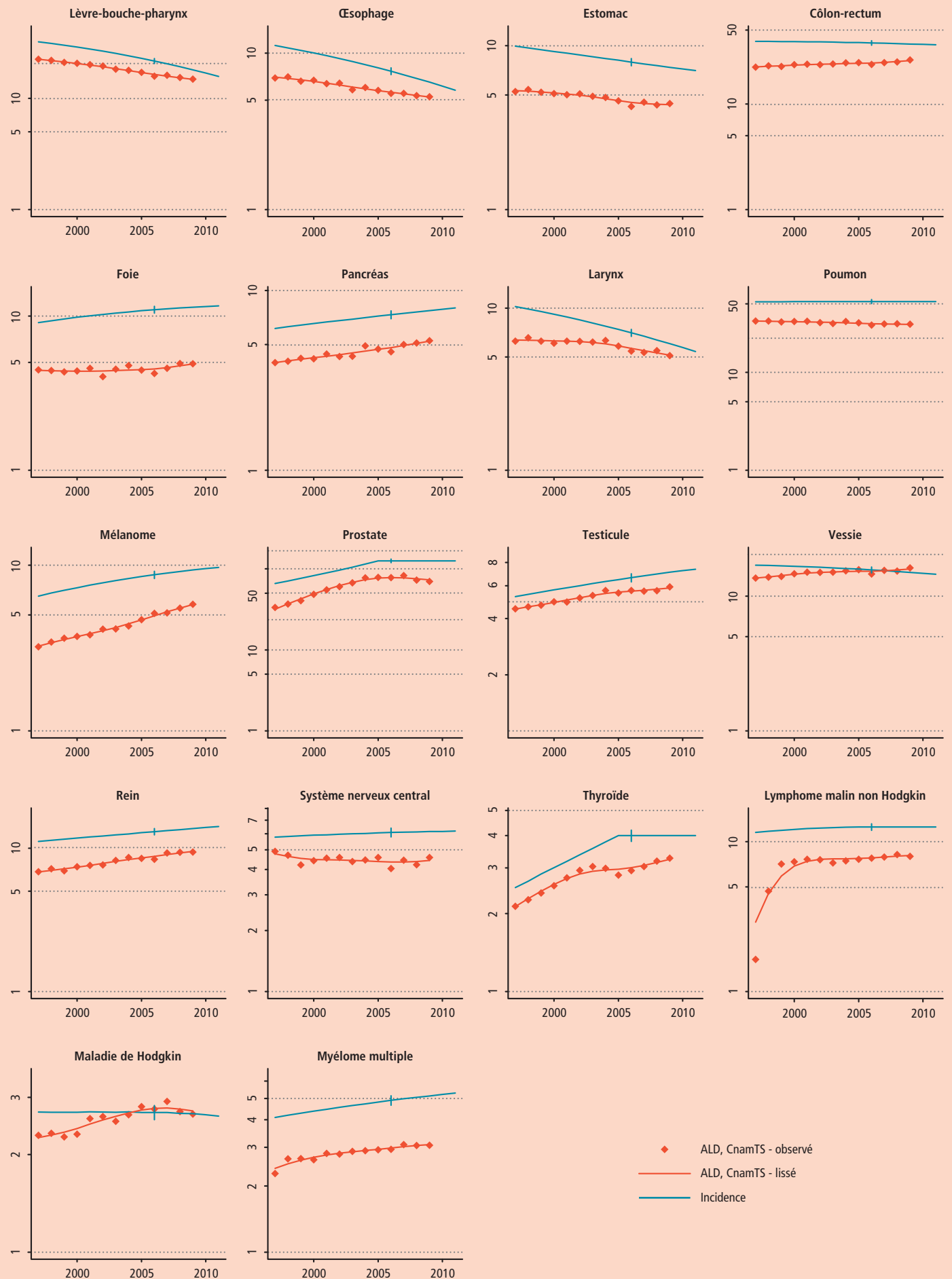


Figure 1b Taux annuels d'incidence 1997-2011 (source : projections 2011, Francim-Hospices civils de Lyon) et d'admissions en ALD 1997-2009, standardisés sur l'âge de la population mondiale, femmes, échelle logarithmique, France / Figure 1b Annual incidence rates 1997-2011 (Source: projections Francim-HCL 2011) and LTI rates 1997-2009, age-standardised to the world population, women, logarithmic scale, France

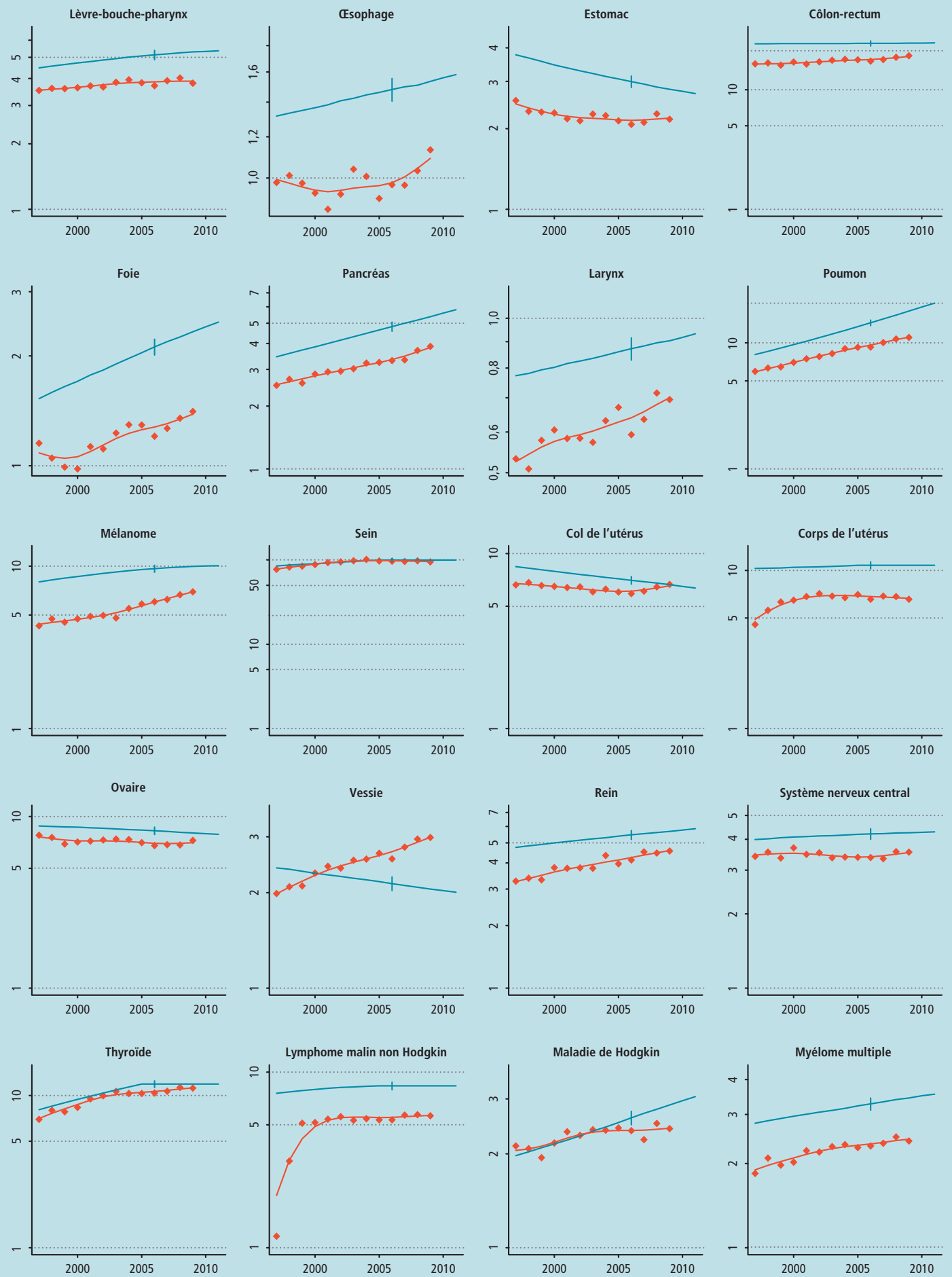
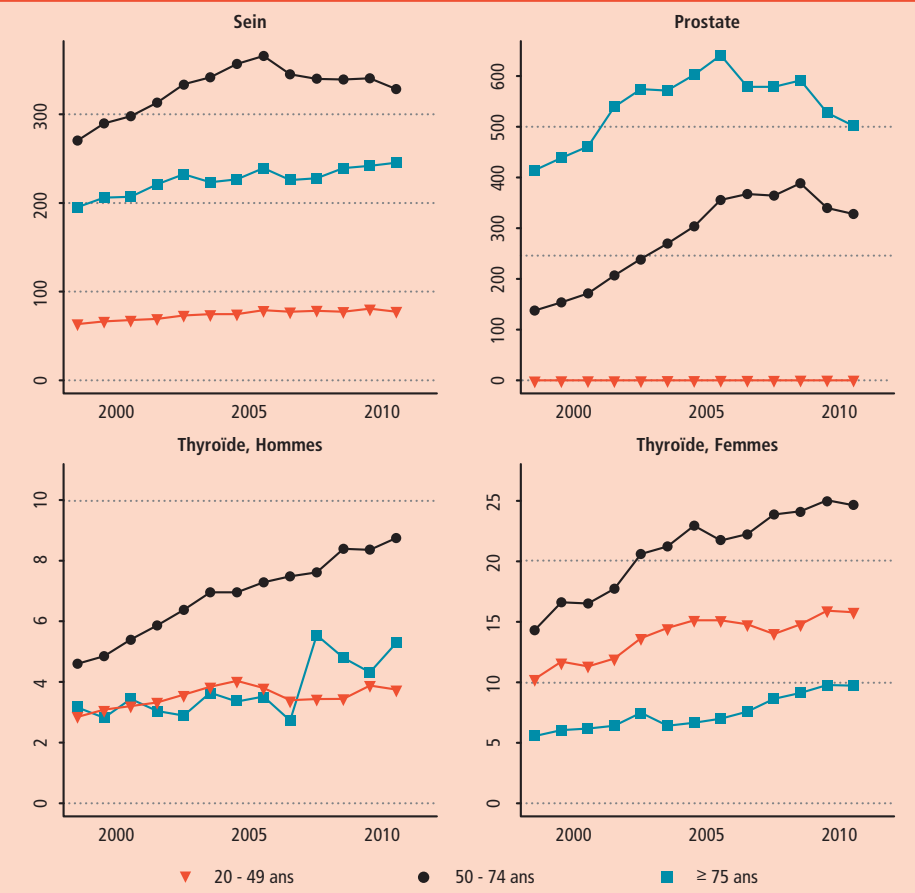


Figure 2 Taux annuels d'admissions en ALD, standardisés sur l'âge de la population mondiale, par groupe d'âge (20-49 ans, 50-74 ans, ≥75 ans)*, pour les cancers du sein, de la prostate et de la thyroïde, échelle arithmétique, 1997-2009, France / Figure 2 Annual ALD rates, age-standardised on the world population, by age group (20-49, 50-74, ≥75)* for breast, prostate, and thyroid cancers, arithmetic scale, 1997-2009, France



* Les taux d'ALD ne doivent pas être comparés entre les différents groupes d'âges ; seules les tendances sont interprétables.

tumeur de la vessie, au moment de la demande d'ALD, de distinguer entre infiltrant et non infiltrant. Cette distinction dépend à la fois des moyens diagnostiques et des définitions nosologiques qui évoluent dans le temps. Les effectifs pour la maladie de Hodgkin et le cancer du foie chez la femme, qui sont des cancers rares, ne permettent pas d'interpréter les différences d'évolutions. Pour les autres localisations pour lesquelles les évolutions diffèrent légèrement, les raisons ne sont pas clairement identifiées, mais on peut supposer une augmentation des demandes d'ALD parmi les personnes atteintes de cancers. Les tendances observées pour les cancers du col et du corps de l'utérus doivent être interprétées avec prudence, car la fréquence des ALD pour « cancer de l'utérus sans précision » a varié au cours du temps [4].

Des ruptures de tendances des taux d'ALD ont été observées sur la période d'étude pour les cancers du sein, de la prostate et de la thyroïde.

La CnamTS avait publié, dès janvier 2008, un article faisant état d'une stabilisation du taux de cancer du sein [7], soulignant l'intérêt et la réactivité des données ALD pour la surveillance nationale des cancers. Cette rupture de tendance a d'abord été observée aux États-Unis en 2003 [8;9] et ensuite dans plusieurs pays [4]. Elle a également été constatée dans plusieurs registres français [10;11]. La diminution rapide de l'utilisation des traitements hormonaux

substitutifs de la ménopause (THS) a été la principale explication évoquée dans la littérature internationale.

Le taux d'ALD pour cancer de la prostate, après des années de forte augmentation, a diminué pour la première fois en 2008 et 2009. Cette diminution est cohérente avec les connaissances épidémiologiques et les données internationales [8;12]. Le développement intensif du dépistage par dosage du *Prostate Specific Antigen* (PSA) a principalement été évoqué pour expliquer les augmentations majeures de l'incidence du cancer de la prostate dans plusieurs pays occidentaux (+8,5% par an en France entre 2000 et 2005 [1]). Une baisse brutale des taux d'incidence avait été observée aux États-Unis [7], et plus récemment en Suède et en Finlande [12], après avoir atteint des taux standardisés (population mondiale) compris entre 115 et 140 pour 100 000 selon le pays. En France, le taux d'incidence atteignait déjà 120 pour 100 000 en 2005 [1] et un arrêt de la croissance de l'incidence du cancer de la prostate, voire une diminution, était ainsi attendu à court terme. Le taux d'incidence du cancer de la prostate se stabilise également au Royaume-Uni [13].

Le taux d'ALD pour cancer de la thyroïde présente pour les deux sexes un ralentissement de la croissance chez les personnes âgées de 50 à 74 ans, et une stabilisation chez les plus jeunes. L'incidence du cancer de la thyroïde était

en forte augmentation depuis plus de 25 ans en France (+6% par an) [1]. Une analyse des tendances dans les départements avec registre sur la période 1982-2006 semblait toutefois indiquer une stabilisation en toute fin de période, en moyenne sur l'ensemble des registres [14]. Les évolutions différaient cependant fortement selon le département [14]. L'incidence de ce cancer est en augmentation dans la plupart des pays [15], et à notre connaissance, aucun pays n'a encore observé un ralentissement de la croissance de l'incidence de ce cancer, à l'exception de la Finlande [13]. Une étude en France dans la région Rhône-Alpes avait toutefois observé un ralentissement de la croissance de ce cancer après 2001 [16]. L'incidence du cancer de la thyroïde est très sensible aux pratiques médicales et aux techniques diagnostiques (diagnostic précoce). Il est possible qu'une saturation des effets liés aux pratiques médicales ou une modification des pratiques ait débuté dans certaines classes d'âge. Les évolutions à court et moyen terme pour ce cancer sont toutefois incertaines.

Ainsi, pour ces 3 localisations, les projections d'incidence 2011 ont fait l'hypothèse d'un taux constant depuis 2005, sur la base des arguments épidémiologiques et des évolutions des taux d'ALD observées pour le cancer du sein et de la prostate, et sur la base des tendances observées dans les registres entre 2000 et 2006 et des données ALD pour le cancer de la thyroïde.

Il serait par ailleurs intéressant de confronter les tendances de taux d'ALD à celles des taux de patients hospitalisés ou opérés, issus des données hospitalières, en particulier pour les 3 localisations présentant une rupture de tendance. De plus, pour les cancers dont l'incidence est sensible aux pratiques diagnostiques et médicales, une surveillance des évolutions de ces pratiques serait également utile pour anticiper et interpréter les tendances de l'incidence.

Enfin, notons que la surveillance dans les registres est maintenant élargie au-delà des tumeurs malignes pour certains cancers (système nerveux central, vessie), conformément aux recommandations internationales. Mais dans les analyses de tendance de l'incidence, il est nécessaire d'adopter des définitions des localisations homogènes dans le temps, imposant des contraintes restrictives. Les tumeurs en hématologie sont également concernées par ce type de contraintes, leur classification ayant régulièrement évolué au cours du temps.

Conclusion

Les évolutions des taux d'ALD doivent être interprétées avec prudence. Ces données sont toutefois disponibles dans un délai de un à deux ans et, de ce fait, très utiles à la surveillance nationale de l'incidence des cancers, permettant d'identifier des ruptures de tendances avec réactivité. Cette étude est mise à jour tous les ans et apporte une contribution au choix du scénario de projection d'incidence pour l'année en cours.

Sur la période 1997-2009, les taux d'ALD présentent des ruptures des tendances pour 3 localisations cancéreuses (sein, prostate et thyroïde), à partir du milieu des années 2000. Ces évolutions sont cohérentes avec les connaissances

épidémiologiques et les données internationales pour les cancers du sein et de la prostate. Pour le cancer de la thyroïde, les données internationales ne permettaient pas d'anticiper cette évolution.

Références

[1] Belot A, Grosclaude P, Bossard N, Jouglu E, Benhamou E, Delafosse P, *et al.* Cancer incidence and mortality in France over the period 1980-2005. *Rev Epidemiol Santé Publique.* 2008;56(3):159-75.

[2] Remontet L, Belot A, Bossard N. Tendances de l'incidence et de la mortalité par cancer en France et projections pour l'année en cours : méthodes d'estimation et rythme de production. *Bull Epidemiol Hebd.* 2009;(38):405-8.

[3] Hospices civils de Lyon, Institut de veille sanitaire, Institut national du cancer, Francim, Institut national de la santé et de la recherche médicale. Projections de l'incidence et de la mortalité par cancer en France en 2011. Rapport technique. Saint-Maurice: Institut de veille sanitaire; 2011. 78 p. Disponible à : <http://www.invs.sante.fr/Publications-et-outils/Rapports-et-syntheses/Maladies-chroniques-et-traumatismes/2011/Projection-de-l-incidence-et-de-la-mortalite-par-cancer-en-France-en-2011>

[4] Uhry Z, Remontet L, Grosclaude P, Belot A, Colonna M, Boussac-Zarebska M, *et al.* Tendances récentes des données d'affections de longue durée : intérêt pour la surveillance nationale de l'incidence des cancers. Période 1997-2009. Saint-Maurice: Institut de veille sanitaire; 2011. 40 p. Disponible à : <http://www.invs.sante.fr/Publications-et-outils/Rapports-et-syntheses/Maladies-chroniques-et-traumatismes/2011/Tendances-recentes-des-donnees-d-affections-de-longue-duree-intere-t-pour-la-surveillance-nationale-de-l-incidence-des-cancers>

www.invs.sante.fr/Publications-et-outils/Rapports-et-syntheses/Maladies-chroniques-et-traumatismes/2011/Tendances-recentes-des-donnees-d-affections-de-longue-duree-intere-t-pour-la-surveillance-nationale-de-l-incidence-des-cancers

[5] Caisse nationale de l'assurance maladie des travailleurs salariés. Population protégée par les régimes de sécurité sociale en 2001. Dossier Études et Statistiques. 2003;(56).

[6] Grosclaude P, Dentan C, Trétarre B, Velten M, Fournier E, Molinié F. Utilité des bases de données médico-administratives pour le suivi épidémiologique des cancers Comparaison avec les données des registres au niveau individuel. *Bull Epidemiol Hebd.* 2012;(5-6):63-7.

[7] Allemand H, Seradour B, Weill A, Ricordeau P. Decline in breast cancer incidence in 2005 and 2006 in France: a paradoxical trend. *Bull Cancer.* 2008;95(1):11-5.

[8] Ravdin PM, Cronin KA, Howlader N, Berg CD, Chlebowski RT, Feuer EJ, *et al.* The decrease in breast-cancer incidence in 2003 in the United States. *N Engl J Med.* 2007;356(16):1670-4.

[9] SEER Cancer Statistics Review, 1975-2008 [Internet]. Bethesda, MD: National Cancer Institute. Disponible à : http://seer.cancer.gov/csr/1975_2008/

[10] Fontenoy AM, Leux C, Delacour-Billon S, Allieux C, Frenel JS, Campone M, *et al.* Recent trends in breast cancer incidence rates in the Loire-Atlantique, France: a decline since 2003. *Cancer Epidemiol.* 2010;34(3):238-43.

[11] Daubisse-Marliac L, Delafosse P, Boitard JB, Poncet F, Grosclaude P, Colonna M. Breast cancer incidence and time trend in France between 1990 and 2007: a population-based study from two French cancer registries. *Ann Oncol.* 2011;22(2):329-34.

[12] Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff M, Klint A, *et al.* NORDCAN: Cancer incidence, mortality, prevalence and prediction in the Nordic countries. Version 3.5 [Internet]. Association of the Nordic Cancer Registries. Danish Cancer Society. Disponible à : <http://www.ancreu.net>

[13] Cancer Research UK. Cancer Statistics for the UK [Internet]. Disponible à : <http://info.cancerresearchuk.org/cancerstats>

[14] Rogel A, Colonna M, Uhry Z, Lacour B, Schwartz C, Pascal L, *et al.* Évolution de l'incidence du cancer de la thyroïde en France métropolitaine. Bilan sur 25 ans. Saint-Maurice: Institut de veille sanitaire; 2010. 55 p. Disponible à : <http://www.invs.sante.fr/fr/Publications-et-outils/Rapports-et-syntheses/Maladies-chroniques-et-traumatismes/2011/Evolution-de-l-incidence-du-cancer-de-la-thyroide-en-France-metropolitaine.-Bilan-sur-25-ans>

[15] Kilfoy BA, Zheng T, Holford TR, Han X, Ward MH, Sjodin A, *et al.* International patterns and trends in thyroid cancer incidence, 1973-2002. *Cancer Causes Control.* 2009;20(5):525-31.

[16] Sassolas G, Hafdi-Nejjari Z, Remontet L, Bossard N, Belot A, Berger-Dutrieux N, *et al.* Thyroid cancer: is the incidence rise abating? *Eur J Endocrinol.* 2009;160(1):71-9.

Utilité des bases de données médico-administratives pour le suivi épidémiologique des cancers. Comparaison avec les données des registres au niveau individuel

Pascale Grosclaude¹ (registre81@orange.fr), Christine Dentan¹, Brigitte Trétarre², Michel Velten³, Evelyne Fournier⁴, Florence Molinié⁵

1/ Registre des cancers du Tarn, Albi, France,

3/ Registre bas-rhinois des cancers, Strasbourg, France

5/ Registre des cancers de Loire-Atlantique, Nantes, France

2/ Registre des cancers l'Hérault, Montpellier, France

4/ Registre des cancers du Doubs, Besançon, France

Résumé / Abstract

Objectif – En France, les registres des cancers couvrent seulement une partie des départements. Pour obtenir une couverture exhaustive en termes de morbidité, l'utilisation des bases de données médico-administratives (BDMA) a été proposée. L'objectif de ce travail était d'étudier les caractéristiques des BDMA nécessaires pour permettre de les utiliser non seulement comme outil de suivi épidémiologique des cancers, mais aussi pour identifier des cas dans le cadre d'un suivi de cohorte ou pour réaliser une surveillance hors d'une zone couverte par un registre.

Méthode – Nous avons analysé la concordance, au niveau individuel, entre les données des « affections longue durée » (ALD) d'une part, les données du Programme de médicalisation des systèmes d'information (PMSI) d'autre part et les données des registres de cancers départementaux, utilisées dans ce travail comme la référence en matière d'incidence. Ce travail a porté sur les cas diagnostiqués en 2004.

Résultats – Pour les ALD 2004, les 5 registres inclus dans l'étude ont recensé 22 455 cancers, qui ont été comparés à 17 133 ALD. Pour le PMSI 2004, les 4 registres inclus dans l'étude ont recensé 16 066 cancers, qui ont été comparés à 26 294 diagnostics de cancer fournis par les Départements d'information médicale de 68 établissements.

Les résultats montrent que pour repérer des cas de cancer incidents, la sensibilité des ALD (0,56) est plus mauvaise que celle du PMSI (0,70). En revanche la valeur prédictive positive (VPP) d'un signalement par une ALD est meilleure que celle d'un signalement par le PMSI (0,74 versus 0,43), car le PMSI inclut beaucoup plus de cas prévalents que les ALD. Ces résultats varient en fonction des cancers étudiés car les facteurs de sous-déclaration sont multiples.

Relevance of health administrative databases in cancer surveillance. Comparison with registries records at individual level

Objective – In France, cancers registries cover only part of the districts. For an exhaustive coverage in terms of morbidity, the use of health administrative databases (ADB) was proposed. The aim of this work was to study the characteristics of ADBs, and use them both as epidemiological indicators of cancer, and as tools allowing to identifying cases in the monitoring of cohorts or carrying out surveillance outside an area covered by cancer registries.

Method – The correlation between long-term illness (LTI) data at individual level, data from the French Hospital Information System (PMSI), and data from population based cancer registries - used as the gold standard in this study - were analysed. This work includes cancer cases diagnosed in 2004.

Results – Regarding LTIs recorded in 2004, the five registries included in the study recorded 22,455 cancer cases which were compared with 17,133 LTI cases. Regarding the 2004 PMSI data, the four registries included in the study recorded 16,066 cancer cases, which were compared with 26,294 diagnoses of cancer provided by 68 medical information units. Our results show that in order to detect incident cancers cases, LTI sensitivity (0.56) is lower than that of PMSI (0.70). On the other hand, the positive predictive value (PPV) of a LTI is better than the one provided by PMSI (0.74 versus 0.43), as PMSI includes more prevalent cases than LTIs. These results vary according to the types of cancer studied as underreporting factors are multiple.

Conclusion – Les ALD semblent difficiles à utiliser seules pour effectuer un recensement exhaustif des cas. En revanche, elles sont un bien meilleur indicateur du fait qu'une personne a effectivement un cancer prévalent. Le PMSI semble plus exhaustif, mais son utilisation conduirait à devoir contrôler beaucoup de cas du fait de sa mauvaise VPP. Des critères d'extraction des enregistrements plus complexes que ceux que nous avons utilisés, permettant de différencier les cas incidents des prévalents, sont nécessaires. Toutefois, une sélection trop importante risque de dégrader la capacité des bases PMSI à identifier les cas.

Conclusion – To make an exhaustive inventory of cancer cases, LTIs seem difficult to use on their own. They are, however, a much better indicator of cancer prevalence. Even though PMSI data seems more exhaustive, its use would mean checking more cases due to its poor PPV. More complex criteria of record extraction than the ones that we used in the present study are needed to distinguish incident cases from prevalent cases. However, too large a selection could reduce the PMSI capacity to detect cancer cases.

Mots clés / Key words

Cancer, affection de longue durée, PMSI, registre / Cancer, long-term illness, hospital discharge records, registries

Introduction

En France, les registres des cancers couvrent moins de 20% de la population, ce qui ne permet pas une surveillance territoriale fine, par exemple celle nécessaire à l'étude de foyers de cancers éventuellement associés à une exposition environnementale. Pour obtenir une couverture exhaustive en termes de morbidité, l'utilisation des bases de données médico-administratives (BDMA), notamment le registre des assurés de l'assurance maladie exonérés du ticket modérateur pour les soins liés à une affection de longue durée (ALD), ou les bases des résumés de sortie standardisé (RSS) répertoriés dans le cadre du Programme de médicalisation des systèmes d'information (PMSI), est envisagée. Ces données possèdent deux atouts majeurs : une couverture nationale et une disponibilité relativement rapide. En revanche, elles ne sont pas nécessairement adaptées à une mesure précise d'indicateurs épidémiologiques, même si elles compilent des données utilisées pour le calcul de l'incidence (localisation du cancer codée selon la Classification internationale des maladies (CIM10), lieu de résidence, date de diagnostic).

L'objectif de ce travail était d'étudier la possibilité d'utiliser les données fournies par ces BDMA pour l'estimation d'indicateurs de suivi épidémiologique des cancers hors d'une zone couverte par un registre.

Matériel et méthode

Nous avons analysé la concordance, au niveau individuel, entre les données ALD d'une part, les données du PMSI d'autre part, et les données des registres de cancers départementaux, utilisées ici comme la référence pour le calcul de l'incidence. Ce travail a été centré sur les cas diagnostiqués en 2004.

Données recueillies

Les registres recensent de façon exhaustive tous les cas de cancers diagnostiqués dans la population résidant dans leur département. Nous avons utilisé l'ensemble des tumeurs recensées jusqu'en 2004, dernière année disponible au moment de l'étude. La base des registres est constituée grâce au recoupement et la vérification de toutes les informations disponibles (y compris des informations issues des sources de signalement que sont les ALD et le PMSI). Les registres impliqués dans cette étude couvrent les départements du Tarn, du Doubs, du Bas-Rhin, de l'Hérault et de Loire-Atlantique (étude des ALD uniquement pour ce dernier département).

Les bases étudiées étaient : les ALD des principales caisses d'assurance maladie (Caisse primaire d'assurance maladie, Mutualité sociale agricole,

Régime social des indépendants, Caisse nationale militaire de sécurité sociale, Caisse autonome nationale de la sécurité sociale dans les mines) des départements couverts par les registres, et les données du PMSI (de court séjour) fournies par les structures de soins fréquentées par la population de ces départements.

Les données demandées aux caisses de chaque département comportaient toutes les premières demandes enregistrées au répertoire des ALD pour « tumeur maligne, affection maligne du tissu lymphatique ou hématopoïétique » durant l'année 2004, ainsi que pour les années 2003 et 2005 afin d'avoir des données avant et après période d'étude. Ces fichiers ne comportaient normalement qu'un signalement pour chaque cancer d'un sujet donné. Dans chaque département, le fichier ALD a été constitué par la compilation des fichiers des différentes caisses.

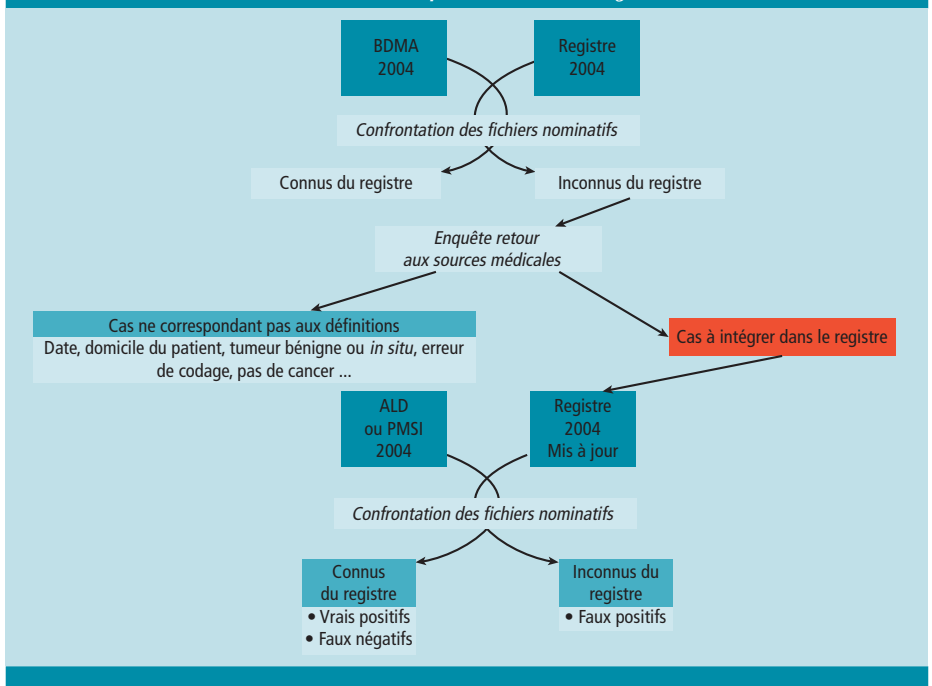
Pour le PMSI, le registre a demandé aux Départements d'information médicale (DIM) des établissements de soins les RSS de tous les séjours ayant un diagnostic principal et/ou un diagnostic relié avec un code CIM10 de cancer durant l'année 2004, ainsi que pour les années 2003 et 2005 (même raison que ci-dessus) pour les patients résidant dans sa zone de surveillance. Ainsi, le fichier initial pouvait contenir plusieurs enregistrements par cas de cancer correspondant aux

différentes hospitalisations d'un patient (parfois dans des structures différentes). Des règles ont donc été définies pour ne conserver des fichiers PMSI qu'un seul enregistrement par patient, par année et par type de cancer (seul le premier séjour de chaque année a été conservé). Le chaînage des RSS a été fait par chaque registre, grâce aux informations nominatives.

Croisement des données

Les fichiers ainsi constitués ont ensuite été confrontés à la base de données de chaque registre. Cette confrontation, sur la base des données nominatives présentes dans les différents fichiers, a été réalisée en deux temps (figure 1). Après un premier croisement, tous les cas signalés par les BDMA qui ne figuraient pas dans le registre ont fait l'objet d'une enquête. Cette enquête a été réalisée par les enquêteurs des registres auprès des différentes sources médicales, pour vérifier si leur absence était légitime ou non. S'il s'agissait d'une erreur du registre (liée à une mauvaise transmission des données de signalement), les cas ont été enregistrés dans la base registre ; s'il s'agissait de cas ne relevant pas de l'enregistrement, le motif de cette exclusion a été noté. Après cette étape, un second croisement des fichiers (registre « enrichi »/ALD ou PMSI) a été réalisé. En se limitant aux cas de 2004, il a permis de classer les cas fournis par

Figure 1 Procédure de confrontation des données 2004 des bases de données médico-administratives (BDMA) et des registres départementaux de cancer / **Figure 1** Procedure for matching the 2004 data from health and medical databases and from departmental cancer registries



les BDMA et l'ensemble de ceux recensés par les registres dans deux tableaux où, par analogie avec un test de dépistage, un cas signalé par la BDMA et figurant dans le registre était considéré comme un vrai positif (VP) et un cas signalé par la BDMA non retrouvé par le registre comme un faux positif (FP). Les cas présents dans le registre ne figurant pas dans les BDMA sont classés en faux négatifs (FN). Il est alors possible d'étudier pour différents sous-groupes (sexe, âge, type de cancer) :

- la sensibilité, qui est le rapport des vrais positifs sur le nombre total de cas incidents en 2004 recensés par le registre ;
- la valeur prédictive positive (VPP) qui est le rapport des VP sur le nombre total de cas de cancers dans la BDMA en 2004.

La définition des VP dépend du niveau de concordance attendu. Pour cette première partie de l'étude, la concordance n'était considérée comme bonne que si la localisation du cancer et le département de résidence étaient les mêmes et que l'année de signalement par les BDMA correspondait à l'année du diagnostic.

En étant moins restrictif sur la date de diagnostic et en considérant que tous les cas incidents en 2004 retrouvés dans les BDMA de l'année précédente (2003) et de l'année suivante (2005) sont des VP, il est possible de calculer une « pseudo-sensibilité ». Cette nouvelle valeur ne peut être considérée comme une vraie sensibilité car il n'existe plus de cohérence entre numérateur (cas signalés par les BDMA de 2003 à 2005 correspondant à des cas incidents en 2004) et le dénominateur (ensemble des cas incidents en 2004) (figure 2).

Il est également possible de calculer une « pseudo-VPP » en considérant comme VP non seulement les cancers diagnostiqués en 2004 mais aussi ceux diagnostiqués avant 2004 (des cas prévalents signalés par les registres). Cette nouvelle valeur n'est pas la VPP pour les cas prévalents car, comme précédemment, il n'existe pas de cohérence entre numérateur (cas recensés par les registres jusqu'en 2004) et le dénominateur (ensemble des cas signalés par la BDMA en 2004) (figure 2).

Résultats

Pour les ALD et l'année 2004, les 5 registres inclus dans l'étude ont recensé 22 455 cancers qui ont été comparés à 17 133 ALD. Il était possible de calculer le pourcentage de la population correspondant aux caisses d'assurance maladie pour lesquelles les ALD avaient été obtenues. La couverture variait de 99% pour le Tarn à 93% pour l'Hérault, qui n'a pu obtenir les données du Régime social des indépendants.

Pour l'année 2004, la sensibilité était de 0,56 pour l'ensemble des tumeurs (tableau 1). Mais les sensibilités variaient de façon importante d'un département à l'autre. Elles étaient de 0,62 pour le Tarn, 0,48 pour le Bas-Rhin, 0,55 pour l'Hérault, 0,62 pour le Doubs et 0,60 pour la Loire-Atlantique. La sensibilité diminuait avec l'âge et était un peu plus élevée chez les femmes que chez les hommes. La variabilité liée au sexe était en partie liée à la variabilité observée en fonction des cancers. En effet, dans tous les départements, la meilleure sensibilité était celle observée pour les cancers du sein (0,77). En revanche, les

Figure 2 Définition des pseudo-sensibilité et pseudo-VPP / Figure 2 Definition of pseudo-sensitivity and pseudo-PPV

	BDMA 2003	BDMA 2004	BDMA 2005		BDMA 2003	BDMA 2004	BDMA 2005		BDMA 2003	BDMA 2004	BDMA 2005
Registre antérieur				Registre antérieur				Registre antérieur			
Registre 2004		VP		Registre 2004		VP		Registre 2004		VP	

Définition des vrais positifs (VP) utilisée dans le calcul de la sensibilité et de la VPP

Définition des vrais positifs (VP) utilisée dans le calcul de la pseudo-sensibilité (VP / registre 2004)

Définition des vrais positifs (VP) utilisée dans le calcul de la pseudo-VPP (VP / BDMA 2004)

Tableau 1 Confrontation des données de la base des affections longue durée (ALD) avec celle des registres de cancer / Table 1 Comparison of data from LTI and cancer registries databases

	Registres 2004	ALD 2004	Vrais positifs	Sensibilité	Pseudo-sensibilité	VPP ^a	Pseudo-VPP
Total	22 454	17 133	12 624	0,56	0,60	0,74	0,83
Sexe							
Homme	13 127	9 396	7 135	0,54	0,58	0,76	0,85
Femme	9 327	7 737	5 489	0,59	0,63	0,71	0,80
Âge							
<50 ans	2 686	2 491	1 723	0,64	0,68	0,69	0,79
50-74 ans	13 034	10 473	7 904	0,61	0,65	0,75	0,84
≥75 ans	6 734	4 169	2 997	0,45	0,48	0,72	0,80
Type de tumeur (Code CIM-10)							
Lèvre - bouche - pharynx (C.00 - C.14 ; C.30 - C.31)	932	644	481	0,52	0,56	0,75	0,84
Larynx (C.32)	221	188	103	0,47	0,50	0,55	0,66
Œsophage (C.15)	302	214	167	0,55	0,58	0,78	0,82
Estomac (C.16)	410	244	175	0,43	0,45	0,72	0,80
Grêle - côlon - rectum - anus (C.17 - C.21)	2 688	1 852	1 485	0,55	0,58	0,80	0,89
Foie (C.22)	529	267	150	0,28	0,30	0,56	0,66
Pancréas (C.25)	492	333	248	0,50	0,54	0,74	0,79
Appareil digestif autres (C.23 - C.24. C.26)	148	75	32	0,22	0,22	0,43	0,51
Poumon - bronche - trachée (C.33 - C.34)	2 090	1 359	1 122	0,54	0,57	0,83	0,88
Mésothéliome (C.45)	73	28	25	0,34	0,38	0,89	0,89
Mélanome cutané (C.43)	572	286	182	0,32	0,34	0,64	0,80
Sarcome (C.37 - C.41 C.46 - C.49)	233	252	100	0,43	0,45	0,40	0,47
Sein (C.50)	3 437	3 484	2 633	0,77	0,81	0,76	0,83
Col utérin (C.53)	181	160	95	0,52	0,55	0,59	0,69
Corps utérin (C.54)	403	277	216	0,54	0,58	0,78	0,87
Utérus sans précision (C.55)	17	48	2	0,12	0,18	0,04	0,17
Ovaire, autres organes génitaux féminins (C.51. C.52 C.56 - C.58)	351	309	209	0,60	0,62	0,68	0,77
Prostate (C.61)	4 621	3 575	3 035	0,66	0,71	0,85	0,95
Testicule, autres organes génitaux masculins (C.60. C.62 - C.63)	170	146	107	0,63	0,68	0,73	0,79
Rein, autres organes urinaires (C.64 - C.66. C.68)	701	521	369	0,53	0,57	0,71	0,82
Vessie (C.67)	620	670	311	0,50	0,55	0,46	0,53
Système nerveux central - Malin et bénin (C.70 - C.72. D.32 - D.33. D.42 - D.43)	546	359	212	0,39	0,42	0,59	0,67
Œil (C.69)	34	34	17	0,50	0,59	0,50	0,59
Thyroïde (C.73)	334	254	178	0,53	0,58	0,70	0,84
Autre glande endocrine (C.74 - C.75)	15	25	8	0,53	0,53	0,32	0,36
Lymphome malin (C.81 à C.85)	836	548	422	0,50	0,55	0,77	0,87
Leucémie (C.91 à C.95)	613	461	292	0,48	0,51	0,63	0,78
Maladie immunoproliférative (C.88. C.90. C.96)	327	289	150	0,46	0,49	0,52	0,59
Siège mal défini (C.76 - C.80. C.97)	558	231	98	0,18	0,18	0,42	0,54

^a Valeur prédictive positive.

résultats étaient plus mauvais pour les cancers qui représentent une forte proportion des cancers chez l'homme : cancer de la prostate (0,66), des poumons (0,54) et de la sphère ORL (0,52). Si l'on prenait en compte des ALD attribuées de 2003 à 2005 pour identifier les cas diagnostiqués en 2004, la pseudo-sensibilité était de 0,60.

Pour l'ensemble des tumeurs, la VPP était de 0,74. Les valeurs observées étaient beaucoup plus homogènes entre les différents départements : 0,79 pour le Doubs, 0,76 pour la Loire-Atlantique, 0,74 pour le Tarn, 0,70 pour le Bas-Rhin et 0,71 pour l'Hérault. Elles étaient aussi plus semblables dans les calculs par sous-groupes. Contrairement à ce que l'on observait pour la sensibilité, les VPP variaient peu avec l'âge et étaient plus élevées chez les hommes. Cette observation ne peut être dissociée de ce que l'on observait par organe. La VPP des mises en ALD pour cancer de la prostate était de 0,85, celle du cancer du poumon de 0,83 alors que celle du cancer du sein était de 0,76. Si l'on ne tenait pas compte de la date de diagnostic pour la concordance (inclusion des cas prévalents), la pseudo-VPP était de 0,83.

Pour le PMSI et l'année 2004, les 4 registres inclus dans l'étude ont recensé 16 066 cancers qui ont été comparés à 26 294 diagnostics de cancer fournis par les DIM de 68 établissements. Contrairement à ce qui se passait pour les ALD, il n'a cependant pas été possible de calculer le pourcentage de la population fréquentant les structures de soins pour avoir un taux de couverture.

Pour l'ensemble des tumeurs invasives, la sensibilité était de 0,70 (tableau 2). Elle variait peu d'un département à l'autre : 0,65 pour le Tarn, 0,70 pour l'Hérault et le Doubs, et 0,73 pour le Bas-Rhin. Dans tous les départements, la sensibilité diminuait avec l'âge et elle était un peu plus élevée chez les femmes que chez les hommes. Les meilleurs résultats étaient observés pour les cancers du côlon-rectum (0,83) et du pancréas (0,84) mais aussi pour ceux de la sphère ORL (0,81). Pour les cancers du sein, la valeur était moyenne (0,75) et c'était pour les cancers de la prostate et pour les mélanomes que la sensibilité était la moins bonne (0,53 et 0,49). Si l'on utilisait les RSS de 2003 à 2005 pour identifier les cas diagnostiqués en 2004, la pseudo-sensibilité était de 0,76.

La VPP était de 0,43 pour l'ensemble des tumeurs (0,39 pour le Doubs et le Bas-Rhin, 0,44 pour le Tarn, et 0,49 pour l'Hérault). La VPP était mauvaise chez les jeunes puis s'améliorait avec l'âge. Contrairement aux résultats de la sensibilité, la VPP était légèrement plus mauvaise chez les femmes que chez les hommes. Cette observation ne pouvait être dissociée de ce que l'on observait par organe. On constatait notamment que la VPP du PMSI pour cancer de la prostate était toujours meilleure que celle du cancer du sein (0,63 *versus* 0,47).

En revanche, si l'on ne tenait pas compte de la date de diagnostic pour la concordance (inclusion des cas prévalents), la pseudo-VPP était de 0,80.

Discussion

En France, plusieurs études ont décrit les données d'ALD pour cancer et les ont utilisées pour estimer l'incidence ou son évolution, en limitant

Tableau 2 Confrontation des données du Programme de médicalisation des systèmes d'information (PMSI) avec celles des registres de cancer / Table 2 Comparison of data from PMSI and cancer registries databases

	Registres 2004	PMSI 2004	Vrai Positifs	Sensibilité	Pseudo-sensibilité	VPP ^a	Pseudo-VPP
Total	16 066	26 294	11 286	0,70	0,76	0,43	0,80
Sexe							
Homme	9 380	14 358	6 370	0,68	0,74	0,44	0,81
Femme	6 686	11 936	4 916	0,74	0,79	0,41	0,78
Âge							
<50 ans	1 872	4 413	1 451	0,78	0,84	0,33	0,78
50-74 ans	9 236	15 915	6 753	0,73	0,80	0,42	0,81
≥75 ans	4 958	5 966	3 082	0,62	0,67	0,52	0,78
Type de tumeur (Code CIM10)							
Lèvre - bouche - pharynx (C.00 - C.14 ; C.30 - C.31)	603	1 093	488	0,81	0,86	0,45	0,82
Larynx (C.32)	142	293	100	0,70	0,79	0,34	0,68
Œsophage (C.15)	191	351	154	0,81	0,86	0,44	0,79
Estomac (C.16)	289	472	219	0,76	0,79	0,46	0,76
Grêle - côlon - rectum - anus (C.17 - C.21)	1 922	3 194	1 588	0,83	0,85	0,50	0,86
Foie (C.22)	327	562	250	0,76	0,82	0,44	0,76
Pancréas (C.25)	382	639	321	0,84	0,87	0,50	0,77
Appareil digestif autres (C.23 - C.24, C.26)	119	181	77	0,65	0,65	0,43	0,59
Poumon - bronche - trachée (C.33 - C.34)	1 576	2 747	1 318	0,84	0,88	0,48	0,86
Mésothéliome (C.45)	36	55	23	0,64	0,67	0,42	0,78
Mélanome cutané (C.43)	372	506	184	0,49	0,54	0,36	0,85
Sarcome (C.37 - C.41 C.46 - C.49)	158	522	104	0,66	0,70	0,20	0,41
Sein (C.50)	2 470	3 960	1 845	0,75	0,83	0,47	0,83
Col utérin (C.53)	144	270	98	0,68	0,72	0,36	0,73
Corps utérin (C.54)	319	398	233	0,73	0,79	0,59	0,81
Utérus sans précision (C.55)	13	18	0	0,00	0,00	0,00	0,17
Ovaire, autres organes génitaux féminins (C.51. C.52 C.56 - C.58)	262	675	215	0,82	0,86	0,32	0,77
Prostate (C.61)	3 386	2 867	1 810	0,53	0,64	0,63	0,95
Testicule, autres organes génitaux masculins (C.60. C.62 - C.63)	133	183	111	0,83	0,87	0,61	0,86
Rein, autres organes urinaires (C.64 - C.66, C.68)	524	641	403	0,77	0,79	0,63	0,84
Vessie (C.67)	499	1 622	372	0,75	0,80	0,23	0,47
Système nerveux central - Malin et bénin (C.70 - C.72, D.32 - D.33, D.42 - D.43)	407	625	282	0,69	0,72	0,45	0,69
Œil (C.69)	22	68	14	0,64	0,68	0,21	0,47
Thyroïde (C.73)	207	304	150	0,72	0,75	0,49	0,87
Autre glande endocrine (C.74 - C.75)	11	39	9	0,82	0,82	0,23	0,44
Lymphome malin (C.81 à C.85)	546	1 133	404	0,74	0,81	0,36	0,84
Leucémie (C.91 à C.95)	408	889	211	0,52	0,59	0,24	0,76
Maladie immunoproliférative (C.88, C.90 , C.96)	220	602	140	0,64	0,68	0,23	0,83
Siège mal défini (C.76 - C.80, C.97)	378	1 385	163	0,43	0,46	0,12	0,74

^a Valeur prédictive positive.

la validation à des comparaisons globales avec des données d'incidence fournies par des registres [1-3]. Il en va de même de l'exploitation des données du PMSI, qui est faite le plus souvent à partir de données globales et anonymes, pour estimer l'incidence, et ne dépend donc pas d'une concordance au niveau individuel [4-6].

Notre objectif est différent car nous souhaitons savoir s'il est possible d'utiliser les BDMA pour repérer les cas de cancers incidents. La première question est alors : quelle proportion de cas échappe à cet enregistrement ? La sensibilité répond à cette question et reflète la capacité des BDMA à recenser les cas de cancers. Une sensibilité élevée traduit une bonne exhaustivité de l'enregistrement. Mais pour que le recensement soit correct, il faut qu'il ne comporte pas trop de FP. Il est bien connu que la sensibilité s'améliore souvent au détriment de la spécificité

par une augmentation du nombre des FP, mais nous n'avons pas la possibilité de mesurer la spécificité dans cette étude. Toutefois, la VPP, qui dépend à la fois des VP et des FP, permet de répondre à une seconde question très opérationnelle, à savoir : quelle est la fréquence des erreurs si l'on considère que tout patient signalé par la BDMA est un cas incident ?

Les résultats montrent que, pour repérer des cas de cancer incidents, les performances des ALD sont moins bonnes que celles du PMSI. En revanche, la VPP d'un signalement par une ALD est meilleure que celle d'un signalement par le PMSI, car le PMSI comprend plus de cas prévalents que les ALD (10% environ pour les ALD et 46% pour le PMSI comme l'atteste le calcul de la pseudo-VPP).

Les faibles sensibilités observées dans ce travail peuvent être liées au fait que nous n'avons

pas eu accès à l'ensemble des bases. Dans le cas des ALD, toutes les caisses d'assurance maladie n'ont pas communiqué leurs données. Toutefois, il semble nous manquer au plus 7% des assurés dans un des départements et moins dans les autres. Pour le PMSI, nous ne pouvions pas calculer le taux de couverture des fichiers qui nous ont été transmis. Nous avons exclu de l'étude le département de la Loire-Atlantique, car un établissement majeur n'avait pas transmis son fichier ; dans les autres départements, les défauts de transmission ont été considérés comme marginaux.

Certains FN correspondent à des divergences de date : 4% des cas manquants ont fait l'objet d'une mise en ALD dans l'année précédant ou suivant le diagnostic. Pour le PMSI, 6% des cas n'avaient pas été hospitalisés en 2004 pour le cancer étudié mais l'avaient été soit en 2003 soit en 2005. Ces cas auraient donc pu être identifiés par l'utilisation des BDMA sur une période plus longue, mais l'année de diagnostic n'aurait pas été exacte.

Les chiffres de sensibilité traduisent donc une sous-déclaration dans ces bases. Les causes de cette sous-déclaration sont connues. Pour les ALD, malgré l'avantage financier que confère la prise en charge à 100% des soins, tous les patients ou les médecins traitants de patients atteints d'un cancer n'en font pas la demande, soit qu'ils ne souhaitent pas faire état de cette pathologie, soit qu'ils n'en aient pas eu le temps du fait d'une évolution rapide. Il est aussi très fréquent que cette demande ne soit pas faite car le patient bénéficie déjà d'un remboursement complet des frais médicaux (mutuelle, CMU, acte chirurgical important).

Pour le PMSI, la sous-déclaration peut correspondre au fait qu'il n'y a pas d'hospitalisation car le traitement est réalisé en ambulatoire. C'est notamment le cas pour les mélanomes cutanés (traités par chirurgie sans hospitalisation) ou pour les cancers de la prostate dont le traitement est médical (hormonothérapie) ou traités par radiothérapie dans les centres privés pour lesquels il n'existait pas de PMSI. Le fait que les causes de sous-déclarations dépendent de la prise en charge (nature, lieu ou coût) explique la variabilité des résultats en fonction des cancers mais aussi en fonction de l'âge des patients.

La VPP dépend de la sensibilité mais aussi de la spécificité (que nous ne pouvons pas mesurer dans ce travail) et de la fréquence de l'événement observé. En revanche, elle ne dépend pas de l'exhaustivité des bases de données qui nous ont été transmises car nous utilisons pour dénominateur les signalements qui nous sont parvenus.

Si 74% des cas identifiés par les ALD sont effectivement des cas incidents, cette proportion passe à 83% lorsqu'on recherche aussi les cas prévalents (diagnostiqués dans des années antérieures).

La VPP du PMSI pâtit plus encore de la présence des cas prévalents, qu'il est difficile de séparer des cas incidents. Le critère de sélection du séjour retenu pour cette étude, et basé uniquement sur la présence d'un code diagnostic de cancer, a l'avantage d'être extrêmement simple mais, en retour, il diminue la spécificité en recrutant des cas prévalents. Tenir compte de l'existence d'un

traitement, notamment d'une chirurgie durant l'hospitalisation, dans la sélection des enregistrements permettrait de résoudre une partie de ce problème en sélectionnant des séjours correspondant à la phase initiale de la maladie. Toutefois, il faut adapter la sélection à chaque tumeur étudiée. De plus, pour des tumeurs chroniques dont le traitement initial n'est pas très spécifique comme les hémopathies, cette sélection serait très difficile.

Nous n'avons trouvé aucune étude sur la concordance entre ALD et registre ; en revanche, il existe des travaux sur le PMSI (ou ses équivalents dans d'autres pays). L'objectif de ces travaux est souvent de comparer des algorithmes d'extraction de données pour optimiser soit la VPP, soit la sensibilité.

Dans un travail italien [7] portant spécifiquement sur 3 cancers fréquents, la sélection des résumés d'hospitalisation en tenant compte du traitement donne des résultats proches des nôtres en termes de sensibilité mais avec une meilleure VPP (pour le cancer du sein : VPP : 0,93 – Sensibilité : 0,77 ; pour le cancer recto-colique : VPP : 0,88 – Sensibilité : 0,72). Pour le cancer du poumon où la chirurgie est moins fréquente, c'est l'existence d'une radiothérapie ou d'une chimiothérapie qui est prise en compte (VPP : 0,79 – Sensibilité : 0,80).

La seule étude française comparant les enregistrements individuels porte sur les cancers thyroïdiens diagnostiqués en 1999-2000 [8]. Elle utilise un algorithme de sélection prenant en compte la chirurgie et trouve une sensibilité du PMSI très proche de la nôtre (0,73) et une VPP de 0,89, bien meilleure que la nôtre.

Pour le cancer de la prostate, qui contrairement à la thyroïde n'est traité chirurgicalement que dans un tiers des cas, une étude réalisée en Espagne [9] montre l'effet délétère sur la sensibilité de la prise en compte de la chirurgie dans l'algorithme de sélection des cas (diagnostic seul VPP : 0,60 – Sensibilité : 0,46 ; diagnostic + chirurgie VPP : 0,74 - Sensibilité : 0,26).

Les FP sont aussi dus à des erreurs de codage. Il s'agit rarement d'une pathologie non tumorale codée en cancer (2% des FP des ALD et 1% du PMSI). Le plus souvent, c'est une localisation tumorale qui est mal codée (environ 10% des FP des ALD et 9% du PMSI) ou une tumeur bénigne ou *in situ* qui n'aurait pas dû être codée comme un cancer (19% des FP des ALD et 10% du PMSI), mais cette proportion varie considérablement en fonction des cancers : elle est d'autant plus importante que ces tumeurs font l'objet d'une prise en charge lourde très proche de celle d'une tumeur maligne (vessie, sein). Les erreurs de domicile (département) ne représentent que 5% des FP dans les ALD et 3% dans le PMSI).

Cette étude n'a pas exploré les deux bases simultanément. Toutefois, elle fait apparaître des complémentarités. Les raisons qui expliquent le manque de sensibilité ou la faible VPP de ces deux BDMA ne sont pas toujours identiques. Ainsi, les cas dont la prise en charge est essentiellement chirurgicale, qui peuvent ne pas faire l'objet d'une demande d'ALD, seront signalés par le PMSI ; à l'inverse, les cas faisant l'objet d'une prise en charge essentiellement ambulatoire, qui échappent au PMSI, ont une forte probabilité de figurer dans les ALD.

Conclusion

Les ALD semblent difficiles à utiliser seules pour réaliser un recensement exhaustif des cas ; en revanche, elles sont un bien meilleur moyen de repérer qu'une personne a effectivement un cancer prévalent. Le PMSI semble plus exhaustif, mais son utilisation conduirait à devoir vérifier beaucoup de cas du fait de sa mauvaise VPP. Des critères d'extraction des enregistrements plus complexes que ceux que nous avons utilisés et permettant de différencier les cas incidents des prévalents, seraient nécessaires pour améliorer la VPP. Toutefois, une sélection trop large risque de dégrader la capacité des bases PMSI à identifier les cas. L'utilisation conjointe de ces deux types de base, qui sont partiellement complémentaires, devra être étudiée.

Remerciements

Ce travail a été réalisé avec la collaboration de toutes les caisses d'assurance maladie du Tarn, du Doubs, du Bas-Rhin, de l'Hérault et de la Loire-Atlantique et des DIM des structures de soins prenant en charge la population de ces départements.

Il a bénéficié d'un financement de l'Institut national du cancer et de l'Institut de veille sanitaire (pour la partie concernant les ALD).

Références

- [1] Chinaud F, Weill A, Vallier N, Ricordeau P, Fender P, Allemand H. Incidence du cancer du poumon en France métropolitaine de 1997 à 2002 : les données du régime général de l'assurance maladie. *Rev Med Ass Mal.* 2004;35(2):69-79.
- [2] Sérador B, Allemand H, Weill A, Ricordeau P. Changes by age in breast cancer incidence, mammography screening and hormone therapy use in France from 2000 to 2006. *Bull Cancer.* 2009;96(4):E1-E6.
- [3] Uhry Z, Remontet L, Colonna M, Belot A, Grosclaude P, Mitton N, *et al.* Évaluation de l'utilisation des données d'affection de longue durée (ALD) pour estimer l'incidence départementale des cancers pour 24 localisations cancéreuses, France, 2000-2005. *Bull Epidemiol Hebd.* 2012;(5-6):71-7.
- [4] Uhry Z, Colonna M, Remontet L, Grosclaude P, Carré N, Couris CM, *et al.* Estimating infra-national and national thyroid cancer incidence in France from cancer registries data and national hospital discharge database. *Eur J Epidemiol* 2007;9:607-14.
- [5] Remontet L, Mitton B, Couris CM, Iwaz J, Gomez F, Olive F, *et al.* Is it possible to estimate the incidence of breast cancer from medico-administrative databases? *Eur J Epidemiol.* 2008;23:681-8.
- [6] Mitton N, Colonna M, Trombert B, Olive F, Gomez F, Iwaz J, *et al.* A suitable approach to estimate cancer incidence in area without cancer registry. *J Cancer Epidemiol.* 2011;2011:418968.
- [7] Baldi I, Vicari P, Di Cuozzo D, Zanetti R, Pagano E, Rosato R, *et al.* A high positive predictive value algorithm using hospital administrative data identified incident cancer cases. *J Clin Epidemiol.* 2008;61(4):373-9.
- [8] Carré N, Uhry Z, Velten M, Trétarre B, Schvartz C, Molinié F, *et al.* Valeur prédictive et sensibilité du programme de médicalisation des systèmes d'information (PMSI) par rapport aux registres des cancers : application au cancer de la thyroïde (1999-2000). *Rev Epidemiol Santé Publique.* 2006;54:367-76.
- [9] Bernal-Delgado E, Martos C, Martínez N, Chirlaque MD, Márquez M, Navarro C, *et al.* Is hospital discharge administrative data an appropriate source of information for cancer registries purposes? Some insights from four Spanish registries. *BMC Health Serv Res.* 2010;10:9.

Les données d'exonération du ticket modérateur pour affections de longue durée (ALD), source potentielle pour la surveillance des cancers d'origine professionnelle ?

Delphine Lauzeille¹ (d.lauzeille@invs.sante.fr), Christine Dentan², Pascale Grosclaude², Jean-Luc Marchand¹, Laurence Chérié-Challine¹

1/ Institut de veille sanitaire, Saint-Maurice, France

2/ Registre des cancers du Tarn, Albi, France

Résumé / Abstract

Introduction – En France, 4,5% à 8% des nouveaux cas de cancers seraient liés à des expositions professionnelles. La production de taux d'incidence de cancer par professions et secteur d'activité n'est actuellement pas possible. L'utilisation des demandes de mise en affection de longue durée (ALD) est une possibilité à envisager pour combler ce manque. Mais avant de mettre en place un recueil complémentaire d'informations professionnelles à partir des ALD, il était essentiel de connaître leur taux d'exhaustivité, ainsi que le nombre de faux positifs.

Méthode – Une étude menée par le réseau Francim sur la concordance entre les données des ALD et celles des registres (gold standard) dans 5 départements apporte des réponses à cette problématique. La sensibilité et la valeur prédictive positive (VPP) des ALD ont été étudiées pour 3 localisations d'intérêt majeur dans le domaine des risques professionnels : poumon, vessie et leucémies.

Résultats – La sensibilité était d'environ 50% pour les 3 localisations étudiées. La VPP était variable selon les localisations : 46% pour la vessie, 63% pour les leucémies et 83% pour le poumon.

Conclusion – La faible sensibilité des ALD n'est pas en faveur de leur utilisation comme seule source de données pour produire des taux d'incidence de cancers par métiers et secteurs d'activité en France. De plus, il est probable que les demandes de mise en ALD soient liées à des facteurs socio-professionnels.

Long-term illness health insurance data: a potential source for monitoring occupational cancer?

Introduction – In France, 4.5% to 8% of new cancer cases are linked to occupational exposures. It is not currently possible to produce rates of cancer incidence by occupation. The use of the long-term illness health insurance database (LTI) is to be considered. But before setting up a complementary collection of occupation information from LTI, it is essential to know their level of completeness and the number of false positive cases.

Method – The Francim study related to the correlation between LTIs and registries data (gold standard) in five districts provides answers to this question. The sensitivity and positive predictive value (PPV) of LTIS were studied for three locations of major interest in the field of occupational risks: lung, bladder and leukemia.

Results – The sensitivity was about 50% for the three locations studied. The PPV was more variable depending on the tumor location (46% for bladder, 63% for leukemia, and 83% for lung).

Conclusion – The low sensitivity of LTIs does not support their use as a single data source to produce cancer incidence rates by occupation in France. The request for LTI agreements may also be related to socio-professional factors.

Mots clés / Key words

Cancers professionnels, surveillance épidémiologique, ALD / Occupational cancers, epidemiological surveillance, long-term illness

Contexte et justification

En France, 4,5 à 8% des cancers seraient liés à des expositions professionnelles, soit entre 14 000 et 26 000 nouveaux cas par an (sur les 320 000 estimés en 2005) [1;2]. Bien que toutes les localisations soient concernées, certaines le sont plus particulièrement. C'est le cas des cancers du poumon, pour lesquels la fraction attribuable à des expositions professionnelles chez les hommes en France est estimée entre 13 et 29% ; les cancers de la vessie, avec une fraction attribuable estimée entre 8 et 14%, et les leucémies, avec une fraction attribuable estimée entre 5 à 18% [3]. Le Code de la sécurité sociale¹ définit un cancer professionnel comme « tout cancer, en particulier des cancers primitifs, résultant d'une exposition professionnelle à certains produits ou procédés », et répondant aux critères définis dans un tableau de maladie professionnelle. C'est le cas pour ces 3 localisations pour lesquelles il existe un certain nombre de facteurs de risque professionnels identifiés et donnant droit à réparation au titre des maladies professionnelles. On citera en particulier l'amiante pour les cancers du poumon, le benzène pour les leucémies et des amines aromatiques pour les cancers de la vessie. Le nombre de ces cancers reconnus

et indemnisés reste toutefois très inférieur à la réalité.

Dans ce contexte, la surveillance des cancers d'origine professionnelle est primordiale. Plusieurs programmes fournissent des éléments d'information (mortalité par secteurs d'activité [4], programme national de surveillance du mésothéliome [5], cohortes prospectives multi-secteurs [6]) mais aucun ne permet actuellement de produire des taux d'incidence de cancers par métier et par secteur d'activité pour la France. On se heurte en effet à la difficulté de mettre en relation des données de morbidité et des données professionnelles car il n'existe pas de bases de données regroupant ces deux informations. Concernant le repérage des cas de cancers, les registres représentent la seule source exhaustive pour les départements couverts, mais ils ne recueillent pas en routine d'informations professionnelles, et les expérimentations réalisées en ce sens n'ont pas été concluantes [7]. Les demandes de mise en affection de longue durée (ALD) pour tumeur maligne représentent une autre source de données potentielle sur les cas de cancers. Elles présentent l'avantage d'avoir une couverture nationale et d'être disponibles rapidement chaque année [8]. La demande de mise en ALD ne contient pas d'informations professionnelles mais elle nécessite un contact direct avec le patient, ce qui permet d'envisager

la mise en place d'un recueil de ces informations sur le formulaire dédié.

Mais, en amont de la faisabilité même de mettre en place un tel recueil à partir des ALD, il était essentiel de connaître le taux d'exhaustivité des ALD par rapport aux cancers incidents repérés par les registres, ainsi que le nombre de cas présents dans les ALD qui ne sont pas des cas de cancers validés par les registres (« faux positifs ») et qui pourraient biaiser les résultats d'une analyse d'incidence par métiers.

L'étude de P. Grosclaude et coll. sur la concordance des cas entre les données des ALD et celles des registres présente un intérêt particulier pour répondre à cette question [9]. Trois localisations d'intérêt majeur dans le domaine des risques professionnels (poumon, vessie et leucémies) ont été étudiées de façon détaillée (sexe, âge, département) afin de déterminer s'il existe des situations favorables au calcul d'incidence des cancers par métiers.

Objectif

L'objectif de ce travail était d'analyser la concordance des cas de cancers entre les données issues des registres et des ALD, afin d'évaluer la possibilité d'utiliser les ALD comme source principale dans le cadre d'un système de surveillance des cancers d'origine professionnelle.

¹ Article L.461-1 du Code de la sécurité sociale.

Matériel et méthode

L'étude de Grosclaude et coll. portait sur 5 départements français et consistait en une analyse de concordance individuelle des cas de cancers entre les bases de données des registres et celles des ALD pour l'année 2004 [9]. Les cas des registres étaient considérés comme le « *gold standard* ».

Les indicateurs étudiés pour répondre à l'objectif de ce travail sont ceux décrits dans l'article de P Grosclaude et coll. : la sensibilité (Se) et la valeur prédictive positive (VPP). Ces indicateurs ont d'abord été étudiés pour chacune des 3 localisations d'intérêt (poumon, vessie, leucémie) selon les critères d'appariements stricts définis dans l'étude globale : identification du sujet, lieu de résidence et année de diagnostic identique dans les 2 bases (2004), tumeur de

même localisation anatomique (CIM10) dans les 2 bases. Ils ont été déclinés selon le sexe, l'âge et le département.

Ils ont ensuite été calculés selon des critères d'appariement élargis notamment pour l'année d'attribution de l'ALD (2003, 2004 et 2005). En effet, il était utile de savoir si les cas de l'année 2004 du registre étaient retrouvés plus globalement dans la base des ALD des autres années.

Une description des cas présents dans les ALD mais absents des registres (« faux positifs ») a également été réalisée pour ces localisations.

Résultats

Selon les règles d'appariement strict, la Se était d'environ 50% pour les 3 localisations étudiées (tableaux 1 à 3). Elle était plus élevée chez les

femmes que chez les hommes sauf pour les cancers du poumon. Elle dépassait 65% chez les moins de 50 ans, mais diminuait de façon importante avec l'âge, quelle que soit la localisation.

La VPP était plus variable selon les localisations (46% pour la vessie, 63% pour les leucémies et 83% pour le poumon). Elle était plus faible chez les femmes que chez les hommes sauf pour les leucémies où elle était équivalente. Elle variait selon l'âge, mais différemment selon les localisations.

Les cancers du poumon et de la vessie présentaient une différence interdépartementale de sensibilité, pouvant atteindre près de 20%. Pour le cancer du poumon, la Se était de 49% dans le Bas-Rhin alors qu'elle était de 67% dans le Tarn.

Tableau 1 Sensibilité (Se) et valeur prédictive positive (VPP) des ALD pour le cancer du poumon (C33-C34), critères d'appariements stricts / *Table 1 Sensitivity and positive predictive value (PPV) of the health insurance data for pulmonary cancer (linkage with strict criteria)*

	Base registres 2004 (nombre de cas)	Base ALD 2004 (nombre de cas)	Vrais Positifs ^a	Faux Négatifs ^b	Faux Positifs ^c	Sensibilité (Se)	Valeur prédictive positive (VPP)	Incidence nationale 2005* (nombre de cas)
Total	2 090	1 359	1 122	968	237	0,54	0,83	30 651
Hommes	1 611	1 005	841	770	164	0,52	0,84	23 937
Femmes	479	354	281	198	73	0,59	0,79	6 714
<50 ans	204	180	137	67	43	0,67	0,76	2 919
50-65 ans	776	548	469	307	79	0,60	0,86	11 610
>65 ans	1 110	631	516	594	115	0,46	0,82	16 129
Loire-Atlantique	514	330	281	233	49	0,55	0,85	-
Bas-Rhin	577	345	283	294	62	0,49	0,82	-
Tarn	200	150	134	66	16	0,67	0,89	-
Doubs	262	168	145	117	23	0,55	0,86	-
Hérault	537	367	279	258	88	0,52	0,76	-

^a Cas présents dans les registres et dans les ALD ; ^b Cas présents dans les registres mais pas dans les ALD ; ^c Cas présents dans les ALD mais pas dans les registres ; * [2].

Tableau 2 Sensibilité (Se) et valeur prédictive positive (VPP) des ALD pour le cancer de la vessie (C67), critères d'appariements stricts / *Table 2 Sensitivity and positive predictive value (PPV) of the health insurance data for bladder cancer (linkage with strict criteria)*

	Base registres 2004 (nombre de cas)	Base ALD 2004 (nombre de cas)	Vrais positifs ^a	Faux négatifs ^b	Faux positifs ^c	Sensibilité (Se)	Valeur prédictive positive (VPP)	Incidence nationale 2005* (nombre de cas)
Total	620	670	311	309	359	0,50	0,46	9 679
Hommes	509	546	260	249	286	0,51	0,48	7 959
Femmes	111	124	51	60	73	0,46	0,41	1 720
<50 ans	16	33	11	5	22	0,69	0,33	304
50-65 ans	115	178	66	49	112	0,57	0,37	2 123
>65 ans	489	459	234	255	225	0,48	0,51	7 252
Loire-Atlantique	121	157	72	49	85	0,60	0,46	-
Bas-Rhin	133	143	50	83	93	0,38	0,35	-
Tarn	86	95	47	39	48	0,55	0,49	-
Doubs	84	82	41	43	41	0,49	0,50	-
Hérault	196	191	101	95	90	0,52	0,53	-

^a Cas présents dans les registres et dans les ALD ; ^b Cas présents dans les registres mais pas dans les ALD ; ^c Cas présents dans les ALD mais pas dans les registres ; * [2].

Tableau 3 Sensibilité (Se) et valeur prédictive positive (VPP) des ALD pour les leucémies (C91 à C95), critères d'appariements stricts / *Table 3 Sensitivity and positive predictive value (PPV) of the health insurance data for leukemia (linkage with strict criteria)*

	Base registres 2004 (nombre de cas)	Base ALD 2004 (nombre de cas)	Vrais positifs ^a	Faux négatifs ^b	Faux positifs ^c	Sensibilité (Se)	Valeur prédictive positive (VPP)	Incidence nationale 2005* (nombre de cas)
Total	613	461	292	321	169	0,48	0,63	6 306
Hommes	350	264	167	183	97	0,48	0,63	3 513
Femmes	263	197	125	138	72	0,48	0,63	2 793
<50 ans	103	100	76	27	24	0,74	0,76	1 193
50-65 ans	131	129	75	56	54	0,57	0,58	1 371
>65 ans	379	232	141	238	91	0,37	0,61	3 742
Loire-Atlantique	205	143	95	110	48	0,46	0,66	-
Bas-Rhin	120	103	57	63	46	0,48	0,55	-
Tarn	59	41	24	35	17	0,41	0,59	-
Doubs	95	64	53	42	11	0,56	0,83	-
Hérault	134	107	63	71	44	0,47	0,59	-

^a Cas présents dans les registres et dans les ALD ; ^b Cas présents dans les registres mais pas dans les ALD ; ^c Cas présents dans les ALD mais pas dans les registres ; * [2].

Pour le cancer de la vessie, le Bas-Rhin présentait une Se de 38%, et la Loire-Atlantique une Se de 60%. Une différence selon les départements était également observée pour la VPP pour ces 3 localisations, plus marquée pour la vessie (de l'ordre de 20%).

En élargissant les règles d'appariement (ALD 2003-2004-2005), il a été possible de calculer une sensibilité élargie. Celle-ci augmentait de 3% pour le poumon et les leucémies, et de 5% pour la vessie (tableau 4).

Les causes de « faux positifs » variaient selon les localisations. Les causes les plus fréquemment rencontrées étaient une différence de l'année de diagnostic, de localisation de cancer, une tumeur bénigne ou une cause inconnue (le dossier n'ayant pu être retrouvé) (tableau 5). Pour le poumon et les leucémies, la cause principale était une différence d'année de diagnostic. Pour la vessie, il s'agissait des tumeurs bénignes.

Discussion

La méthode et les résultats globaux ne sont pas discutés ici (voir [9]). La discussion porte sur la possibilité d'utiliser les ALD pour produire des taux d'incidence de cancers par profession et secteurs d'activité en France. Avant d'étudier la faisabilité de recueillir des informations professionnelles sur les sujets bénéficiant d'une ALD, il était important de s'assurer que l'on retrouve bien les mêmes sujets dans les deux sources. Le résultat idéal serait une concordance parfaite entre les cas « registres » et les cas « ALD » (Se et VPP égales à 100%). Cependant, pour répondre à l'objectif attendu, une sensibilité plus basse peut être acceptable si, par ailleurs, les cas « ALD » sont

bien des cas incidents « registres » (VPP élevée) et si les cas « ALD » sont comparables en termes de facteurs socioprofessionnels aux patients ne bénéficiant pas d'une ALD.

Concernant la VPP, ce travail montre qu'elle est relativement élevée pour le poumon (83%) mais faible pour la vessie (46%). L'étude des « faux positifs » apporte un éclairage sur ces différences. Pour le cancer de la vessie, on retrouve une proportion importante de tumeurs *in situ* en ALD qui n'étaient pas prises en compte dans les registres. Pour les cancers du poumon et les leucémies, environ la moitié des « faux positifs » est due à une différence de date (année de diagnostic dans les registres, année de prise d'effet dans les ALD). Si l'on ajoute aux cas incidents les « faux positifs » pour différence de date (en faisant l'hypothèse que ces cas prévalents s'équilibrent chaque année) et les tumeurs *in situ* pour la vessie (les nouvelles règles de codage des registres vont prendre en compte les tumeurs *in situ*), la VPP serait de 90% pour les cancers du poumon, 87% pour la vessie, et 80% pour les leucémies. Il est peu probable que les autres causes de « faux positifs » (autre localisation de cancer, métastase...) soient liées à des facteurs socioprofessionnels. Ils ne représenteraient donc pas un biais pour la comparaison des taux d'incidence de cancer par métiers. Néanmoins, l'analyse de la VPP ne peut être dissociée de celle de la sensibilité. On sait que tous les sujets atteints de cancer ne font pas de demande de mise en ALD et, d'autre part, qu'il existe des facteurs de surestimation des cas présents dans les ALD [10]. Lorsque l'appariement est réalisé sur des critères stricts, la sensibilité est d'environ 50% quelle que soit la localisation, soit environ 1 cas « registre » sur 2 absent de la base ALD. En élargissant les critères d'appariement, le

gain sur la sensibilité est limité (3 à 5%). L'exhaustivité des cas de cancers incidents à partir des ALD n'est donc pas atteinte, ce qui était pressenti du fait même de la nature des ALD.

Il pouvait être attendu que les résultats de concordance ne soient pas très bons pour les leucémies. En effet, cette catégorie regroupe des pathologies très différentes, pouvant être identifiées au sein des registres mais pas au sein des ALD. Cependant, toutes les leucémies peuvent être associées à des expositions professionnelles, de nature différente selon leur type, et ce regroupement est fréquemment utilisé en épidémiologie des risques professionnels, ce qui justifiait le choix de cette localisation.

Le dernier élément à prendre en compte dans le cadre de ce travail, est la question de l'extrapolation possible de la population « ALD » en terme de répartition des facteurs étudiés. L'enquête « Santé et protection sociale 2008 » montrait que le pourcentage d'exonération du ticket modérateur pour l'ensemble des ALD était inversement proportionnel à l'échelle sociale, puisque 17,2% des ouvriers non qualifiés bénéficiaient d'une ALD contre 9,1% des cadres [11]. Des différences éventuelles de demande de mise en ALD en fonction de facteurs socioprofessionnels sont difficilement mesurables et entraîneraient un biais dans l'utilisation des ALD pour produire des taux d'incidence de cancers par métier. Les résultats par sexe, âge et départements sont parfois meilleurs mais restent insuffisants pour un tel système de surveillance.

Conclusion

Le travail réalisé montre qu'il n'est pas possible d'utiliser les ALD seules de manière satisfaisante pour calculer des taux d'incidence de cancers par métier. En effet, la VPP pourrait être correcte mais ce n'est pas le cas pour la sensibilité. De plus, il est possible que la situation socioprofessionnelle des sujets atteints de cancers intervienne dans le fait qu'ils bénéficient ou non d'une ALD. Même si le nombre des localisations étudiées est limité et leur choix discutable, les résultats généraux pour les autres localisations de cancers ne sont pas non plus concluants [9]. De fait, dans l'objectif d'un système de surveillance basé uniquement sur les ALD, il n'est pas pertinent de mettre en place un recueil actif des informations professionnelles auprès des patients bénéficiant d'une ALD.

Par contre, il est prévu que l'Institut de veille sanitaire teste la possibilité de croiser les données de la Caisse nationale d'assurance vieillesse (Cnav), contenant des informations professionnelles, avec celles des registres. En cas de faisabilité démontrée, il serait intéressant de tester également la faisabilité de croiser les données ALD avec celles de la Cnav. Des comparaisons pourront alors être effectuées entre les deux systèmes dans les départements disposant d'un registre. Sous réserve d'effectifs suffisants, cela pourrait permettre d'éclairer la question du rôle des facteurs socioprofessionnels dans les demandes de mise en ALD.

Remerciements

Aux registres du Bas-Rhin, du Doubs, de l'Hérault, de Loire-Atlantique et du Tarn.

Aux caisses d'assurance maladie participantes du Bas-Rhin, du Doubs, de l'Hérault, de Loire-Atlantique et du

Tableau 4 Nouvelle sensibilité par localisation avec les critères d'appariement élargis sur les bases ALD 2004 / Table 4 New sensitivity with expanded criteria for linking on 2004 ALD bases

	Base registres 2004 (nombre de cas)	Base ALD 2004 (nombre de cas)	Nouveaux vrais positifs (gain en effectif)	Nouvelle sensibilité (gain en %)
Poumon	2 090	1 359	1 186 (+64)	0,57 (+3%)
Vessie	620	670	344 (+33)	0,55 (+5%)
Leucémies	613	461	315 (+23)	0,51 (+3%)

Tableau 5 Causes des faux positifs (cas de cancers présents dans les ALD qui ne sont pas présents dans les registres) / Table 5 Causes of false positives (cancer cases present in the ALD, who are not present in registries)

Causes des « faux positifs »	Poumon		Vessie		Leucémies	
	N	%	N	%	N	%
Autre année de diagnostic	104	43,8	60	16,7	75	44,3
Autre localisation de cancer	48	20,3	15	4,2	6	3,6
Dossier manquant ^a	32	13,5	44	12,3	30	17,8
Autre département	16	6,8	9	2,5	10	5,9
Tumeur bénigne ou non invasive	15	6,3	214	59,6	24	14,2
Métastase ou extension du primitif	10	4,2	0	0,0	0	0,0
Tumeur maligne non confirmée	6	2,5	2	0,6	1	0,6
Autre cause	2	0,8	1	0,3	3	1,8
Hors Circ ^b	2	0,8	0	0,0	0	0,0
Aucune tumeur	1	0,4	1	0,3	5	3,0
Ayant droit	1	0,4	0	0,0	0	0,0
Base certificat décès	0	0,0	0	0,0	1	0,6
Double ALD	0	0,0	1	0,3	0	0,0
Même sur-groupe de cancer ^c	0	0,0	12	3,3	14	8,3
Total	237	100	359	100	169	100

^a Le dossier du patient n'a pas été retrouvé, il n'est donc pas possible de déterminer la raison de l'absence des cas des registres.

^b D'après les règles du Circ (Centre international de recherche sur le cancer) relatives au codage pour les cancers multiples, préconisées par les registres français.

^c Ensemble des cancers thoraciques (y compris mésothéliomes) pour les poumons, ensemble des cancers urologiques pour la vessie, ensemble des hémopathies pour les leucémies.

Tarn, ainsi qu'aux DIM des structures de soins prenant en charge la population de ces départements.

Références

- [1] Tordjman I. Cancers professionnels. État des connaissances en date du 3 décembre 2008. Coll. Fiches Repère. Boulogne-Billancourt: Institut national du cancer. 4 p. Disponible à : <http://www.cancer-environnement.fr/LinkClick.aspx?fileticket=T4q-Bmd2V9w%3D&tabid=165&mid=771>
- [2] Belot A, Velten M, Grosclaude P, Bossard N, Launoy G, Benhamou E, et al. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2005. Saint-Maurice: Institut de veille sanitaire; 2008. 132 p.
- [3] Imbernon E. Estimation du nombre de cas de certains cancers attribuables à des facteurs professionnels en France. Saint-Maurice: Institut de veille sanitaire; 2003. 28 p.
- [4] Geoffroy-Perez B. Analyse de la mortalité et des causes de décès par secteur d'activité de 1968 à 1999 à partir de l'Échantillon démographique permanent. Saint-Maurice: Institut de veille sanitaire; 2006. 158 p.
- [5] Goldberg M, Imbernon E, Rolland P, Gilg Soit Ilg A, Savès M, de Quillacq A, et al. The French National Mesothelioma Surveillance Program. *Occup Environ Med.* 2006;63(6):390-5.
- [6] Coset : cohortes pour la surveillance épidémiologique en lien avec le travail. [Internet]. Disponible à : <http://www.coset.fr>
- [7] Lauzeille D, Marchand JL, Trétarre B, Rébillard X, Molinié F, Leux C, et al. Recueil d'informations professionnelles auprès de patients atteints de cancer. Résultats de deux études pilotes menées avec les registres dans les départements de l'Hérault et de Vendée. Saint-Maurice: Institut de veille sanitaire; 2011. 26 p.
- [8] Goldberg M. Les bases de données d'origine administrative peuvent-elles être utiles pour l'épidémiologie ? *Rev Epidemiol Sante Publique.* 2006;54(4):297-303.
- [9] Grosclaude P, Dentan C, Trétarre B, Velten M, Fournier E, Molinié F. Utilité des bases de données médico-administratives pour le suivi épidémiologique des cancers. Comparaison avec les données des registres au niveau individuel. *Bull Epidemiol Hebd.* 2012 ;(5-6) :63-7.
- [10] Uhry Z, Remontet L, Grosclaude P, Belot A, Colonna M, Boussac-Zarebska M, et al. Tendances récentes des données d'affections de longue durée. Intérêt pour la surveillance nationale de l'incidence des cancers, période 1997-2009. Saint-Maurice: Institut de veille sanitaire; 2011. 40 p.
- [11] Allonier C, Dourgnon P, Rochereau T. Enquête sur la santé et la protection sociale 2008. Paris: Irdes; 2010. 258 p.

Évaluation de l'utilisation des données d'affection de longue durée (ALD) pour estimer l'incidence départementale des cancers pour 24 localisations cancéreuses, France, 2000-2005

Zoé Uhry (z.uhry@invs.sante.fr)¹, Laurent Remontet², Marc Colonna^{3,4}, Aurélien Belot^{1,2}, Pascale Grosclaude^{3,5}, Nicolas Mitton^{3,4}, Solène Delacour-Billon^{3,6}, Julie Gentil^{3,7}, Marjorie Boussac-Zarebska¹, Nadine Bossard², Arlette Danzon¹, Michelle Altana⁸, François Frete⁹, Alain Weill¹⁰, Agnès Rogel¹

1/ Institut de veille sanitaire, Saint-Maurice, France

3/ Réseau français des registres de cancer Francim, France

5/ Registre des cancers du Tarn, Toulouse, France

7/ Registre des cancers gynécologiques de Côte-d'Or, Dijon, France

8/ Direction de la gestion des risques et de l'action sociale, Régime social des indépendants, La Plaine Saint-Denis, France

9/ Échelon national du contrôle médical, Mutualité sociale agricole, Bagnolet, France

10/ Direction de la stratégie, des études et des statistiques, Caisse nationale d'assurance maladie des travailleurs salariés, Paris, France

2/ Service de biostatistique, Hospices civils de Lyon, Lyon, France

4/ Registre du cancer de l'Isère, Grenoble, France

6/ Registre des cancers de Loire-Atlantique et de Vendée, Nantes, France

Résumé / Abstract

Introduction – Cette étude a pour objectif d'évaluer pour 24 localisations cancéreuses la validité d'estimations départementales de l'incidence des cancers à partir des données d'affections de longue durée (ALD). Ces estimations utilisent le rapport entre nombre d'admissions en ALD et nombre de cas incidents observé dans les départements couverts par un registre (rapport ALD/incidence). Ces estimations départementales sont valides à condition que le rapport varie peu d'un département français à un autre, ou que les variations de ce rapport soient modérées, comparées à l'ampleur des variations géographiques d'incidence.

Matériel et méthode – Les données d'admissions en ALD proviennent du régime général d'assurance maladie, de la Mutualité sociale agricole (MSA) et du Régime social des indépendants. La variabilité géographique du rapport ALD/incidence a été étudiée, ajustée sur l'âge, dans les départements couverts par un registre pour la période 2000-2005. La variabilité géographique de l'incidence a également été étudiée.

Résultats – Une estimation départementale de l'incidence est possible pour huit localisations : lèvres-bouche-pharynx, œsophage/hommes, estomac, colon-rectum, poumon, sein, ovaire et testicule. Les autres localisations ont été écartées en raison d'une variabilité géographique majeure du rapport ALD/incidence (foie, vésicule, pancréas, larynx, mélanome, thyroïde, myélome multiple, leucémie) ou trop importante comparée aux variations géographiques d'incidence (col de l'utérus, corps de l'utérus, prostate, rein, système nerveux central). Enfin, trois localisations ont été écartées en raison de difficultés de codage spécifiques (vessie) ou d'effectifs trop faibles (œsophage/femmes, maladie de Hodgkin).

Discussion – Cette étude présente, pour la première fois, une évaluation systématique de la validité d'estimations départementales de l'incidence des cancers à partir des données ALD. Ce travail est une première étape, l'objectif à terme étant d'utiliser l'ensemble des données médico-administratives pertinentes pour estimer l'incidence départementale et régionale des cancers.

Evaluation of the use of long-term disease health insurance data to estimate cancer incidence for 24 cancer sites at district level, 2000-2005

Introduction – This study aims at evaluating for 24 cancer sites the validity of cancer incidence estimates at district level based on long-term illness data (LTI) issued from the health funds. These estimations are based on the ratio between the number of LTI and the number of incident cases observed in the districts covered by a cancer registry (LTI/incidence ratio). These district estimations are valid only if the ratio varies little throughout the districts or if its variations remain moderate compared to the size of the geographical variations in incidence.

Material and method – LTI data are issued from the three main French national health funds (general, agricultural and self-employed). The geographical variability of the LTI/incidence ratio was studied, adjusted on age, in the districts covered by a registry for the period 2000-2005. The geographical variability of incidence was also studied.

Results – District-level estimation is possible for eight cancer sites: lip-oral cavity-pharynx, oesophagus/men, stomach, colon-rectum, lung, breast, ovary and testis. The other cancers sites were excluded because the geographical variability of the LTI/incidence ratio was high (liver, gall bladder, pancreas, larynx, melanoma, thyroid, multiple myeloma, leukaemia) or because it was too important compared to geographical variations in incidence (cervix uteri, corpus uteri, prostate, kidney, central nervous system). Three cancer sites were excluded because of coding issues (bladder) or small numbers (oesophagus/women, Hodgkin's disease).

Discussion – This study presents for the first time a systematic evaluation of the validity of estimating cancer incidence at district level using LTI data. This study is a first step, the objective being to use all pertinent medico-administrative databases to estimate cancer incidence at district and regional levels.

Mots clés / Key words

Cancer, incidence, affection de longue durée, variations géographiques, départements, France / Cancer, incidence, long-term illness, geographical variations, district, France

Introduction

En France, l'incidence des cancers est mesurée par les registres départementaux des cancers qui couvrent actuellement environ 20% de la population. L'incidence nationale est estimée à partir des données des registres et de mortalité, en extrapolant à la France le rapport entre incidence et mortalité observé dans la zone couverte par les registres [1;2]. Cette méthode est adaptée au niveau national, mais elle présente des limites au niveau régional [3] et ne peut être appliquée au niveau départemental. À ce jour, il n'existe pas d'estimations départementales systématiques de l'incidence des cancers en France.

Les données d'affections de longue durée des caisses d'assurance maladie (ALD) offrent une possibilité à évaluer pour l'estimation de l'incidence des cancers dans les départements sans registre. Le rapport entre nombre de demandes d'ALD et nombre de cas incidents de cancers (rapport ALD/incidence), observable dans les départements avec registre, pourrait en effet être utilisé pour redresser les données d'ALD et estimer ainsi l'incidence. L'hypothèse fondamentale de telles estimations est que le rapport ALD/incidence est identique dans tous les départements français ou, autrement dit, que la « différence » entre ALD et incidence est la même dans tous les départements.

La demande de mise en ALD a pour objectif une prise en charge financière sans ticket modérateur de l'assuré pour les soins liés à une pathologie lourde. Ces données couvrent l'ensemble du territoire, y compris les départements d'outre-mer, et sont codées selon des règles standardisées. La possibilité d'utilisation de ces données à des fins de surveillance épidémiologique est ainsi une question récurrente. Les demandes de mise en ALD et les cas incidents ne coïncident que partiellement [4;5], mais cela ne constitue pas obligatoirement un obstacle pour notre objectif, dès lors que la « différence » entre ALD et incidence est la même dans tous les départements.

Cette étude avait pour objectif d'évaluer, pour 24 localisations cancéreuses, la validité d'utiliser le rapport ALD/incidence observé dans les départements couverts par un registre pour estimer l'incidence des cancers dans les départements non couverts par un registre. Cette évaluation visait à sélectionner les localisations pour lesquelles l'utilisation de ce rapport est valide. La qualité des estimations a été évaluée en étudiant la variabilité géographique du rapport ALD/incidence. Par ailleurs, la capacité des estimations à décrire les disparités géographiques d'incidence dépendant à la fois de leur qualité mais aussi de l'ampleur de ces disparités, cet aspect a également été étudié.

Cet article résume la méthodologie adoptée et les principaux résultats de cette évaluation. L'analyse complète et les estimations départementales pour les localisations retenues ont été publiées de façon détaillée [4].

Matériel et méthode

Données

L'analyse a porté sur les nombres de cas incidents et de premières admissions en ALD pour cancer sur la période 2000-2005, dans les départements couverts par un registre général ou spécialisé. La

période d'étude a été déterminée par la disponibilité commune des deux sources au moment de l'analyse (les données de registres sont disponibles avec un délai de 3 à 4 ans).

Vingt-quatre localisations cancéreuses ont été étudiées [4]. Les données d'incidence ont été extraites de la base commune des registres. Par précaution, compte tenu de l'objectif de l'étude, les données des registres ayant signalé des difficultés de recueil sur la période d'étude n'ont pas été incluses dans l'analyse. Pour cette période d'étude, l'analyse portait sur 8 à 11 départements selon la localisation cancéreuse étudiée, représentant 11 à 15% de la population française. Les données ALD sont issues des trois principales caisses d'assurance maladie : régime général, Mutualité sociale agricole (MSA) et Régime social des indépendants (RSI), qui couvrent 96% de la population résidant en France. Le département correspondait au département d'affiliation et non de résidence pour le régime général et la MSA. La concordance entre ces deux informations est toutefois supérieure à 98% pour tous les départements analysés, à l'exception d'un département (Doubs, 94%), pour lequel une correction a été utilisée [4].

Les données ont été agrégées par localisation cancéreuse, département, sexe, et classe d'âge quinquennale.

Méthode

Principe

La méthode d'estimation de l'incidence départementale est détaillée dans une publication récente [4]. Le rapport ALD/incidence est estimé dans la zone couverte par les registres, où ces deux informations sont disponibles. Dans un département sans registre, le nombre d'admissions en ALD est ensuite redressé, en le divisant par ce rapport, afin d'estimer le nombre de cas incidents pour le département. L'hypothèse fondamentale de cette estimation est que le rapport ALD/incidence, pour un âge donné, est identique dans tous les départements français. La qualité des ces estimations est donc évaluée selon deux critères, qui ne peuvent être estimés que dans les départements avec registre : 1/ la variabilité géographique du rapport ALD/incidence ; 2/ la comparaison de l'incidence réelle aux estimations obtenues en utilisant le rapport ALD/incidence (validation croisée, voir ci-après).

Par ailleurs, la capacité des estimations d'incidence à décrire les disparités géographiques dépend de la fois de leur qualité et de l'ampleur des variations géographiques d'incidence. Si l'incidence est très contrastée sur le territoire, des estimations entachées d'erreurs peuvent malgré tout apporter une information sur les gradients géographiques.

Ainsi, la validité des estimations départementales est évaluée sur la qualité des estimations en premier lieu, mais également en mettant en regard qualité et ampleur des disparités géographiques de l'incidence.

Variabilité géographique du rapport ALD/incidence dans les départements avec registre

Le rapport entre les nombres d'admissions en ALD et de cas incidents est modélisé en fonction de l'âge et du département (effet aléatoire

d'écart-type σ_d) dans le cadre d'un modèle de Poisson à effets mixtes. σ_d mesure la variabilité géographique du rapport ALD/incidence. À titre d'illustration, une valeur de σ_d de 0,05 correspond à un rapport variant de $\pm 10\%$ environ selon le département, une valeur de 0,10 à un rapport variant de $\pm 20\%$ environ. La qualité des estimations d'incidence obtenues en utilisant le rapport ALD/incidence, observé dans la zone couverte par des registres, est liée à σ_d : plus σ_d est grand, plus les erreurs d'estimations seront importantes. Cette variabilité est par ailleurs prise en compte dans les intervalles de prédiction des estimations.

Comparaison de l'incidence observée et prédite dans les départements avec registre (validation croisée)

La qualité des estimations est de plus évaluée en confrontant ces estimations à l'incidence observée. Cette confrontation n'est possible que dans les départements avec registre. Pour cela, pour chaque département avec registre, la procédure suivante de « validation croisée » est effectuée : le nombre de cas incidents est prédit à partir d'un rapport ALD/incidence obtenu en excluant le département, et cette prédiction est comparée au nombre de cas observé. L'écart entre le nombre total de cas incidents prédit (P) et observé (O) est évalué par une mesure de type Chi 2 (K). L'erreur relative entre le nombre de cas prédit et le nombre observé (ER) est également présentée. Les résultats détaillés de la validation croisée sont disponibles par ailleurs [4] et incluent aussi une mesure des écarts par âge.

Qualité des estimations comparée à l'ampleur des disparités géographiques de l'incidence

La capacité des estimations à décrire les disparités géographiques d'incidence dépend de leur qualité mais également de l'ampleur de ces disparités géographiques. Cette dernière est également étudiée dans les départements avec registre. Pour en rendre compte de façon comparable à la mesure de la variabilité du rapport ALD/incidence, un modèle de Poisson du même type est réalisé, dans lequel le taux d'incidence est modélisé en fonction de l'âge et d'un effet départemental aléatoire (d'écart-type σ_k) [4]. σ_k mesure la variabilité géographique de l'incidence ou, autrement dit, l'ampleur des variations géographiques d'incidence.

La comparaison de σ_d (qualité des estimations) à σ_k (ampleur des disparités d'incidence) donne une indication sur la possibilité de décrire les disparités géographiques d'incidence à partir des estimations obtenues : σ_d doit être largement inférieur à σ_k .

Sélection des localisations cancéreuses pour une estimation départementale

La possibilité d'utiliser le rapport ALD/incidence pour estimer l'incidence départementale est donc évaluée en fonction de la qualité des estimations (variabilité géographique du rapport σ_d et validation croisée) et de l'ampleur des variations géographiques d'incidence (σ_k), selon le principe suivant :

- variabilité du rapport ALD/incidence majeure ($\sigma_d \geq 0,09$) : localisation non sélectionnée ;
- variabilité du rapport ALD/incidence mineure ($\sigma_d \leq 0,03$) : localisation sélectionnée ;
- variabilité du rapport ALD/incidence intermédiaire (σ_d de 0,04 à 0,08) : localisation sélectionnée

ou non selon l'ampleur des variations géographiques de l'incidence (comparaison de σ_d et σ_k) ;

- faibles effectifs : localisation non exploitable, quelle que soit la variabilité du rapport ALD/incidence.

Des éléments épidémiologiques (difficultés de codage, inclusion de tumeurs non invasives dans les ALD, prise en charge des cancers, etc.) interviennent également dans cette évaluation.

Par ailleurs, les effectifs variant de 600 à 35 000 selon la localisation étudiée (tableau), la puissance d'analyse est extrêmement variable. Pour les localisations avec des effectifs très importants, une erreur de prédiction modérée peut être statistiquement significative, et on considérera alors l'ordre de grandeur des erreurs plutôt que leur significativité.

Analyse de sensibilité en lien avec le régime spécial d'Alsace-Moselle

L'Alsace-Moselle bénéficie d'un régime local spécifique d'assurance maladie plus favorable, qui pourrait conduire à des admissions en ALD moins fréquentes, à pathologie égale, dans les départements concernés. Une analyse de sensibilité en excluant les départements alsaciens (Bas-Rhin et Haut-Rhin) a donc été réalisée.

Résultats

Analyses statistiques

La figure illustre la variabilité géographique du rapport ALD/incidence. La qualité des estimations d'incidence obtenues en utilisant ce rapport est directement liée à cette variabilité. La figure présente les effets départementaux aléatoires

estimés dans l'analyse du rapport ALD/incidence, par localisation cancéreuse, accompagné du σ_d estimé. Un effet département de 1,1 signifie que le département a un rapport ALD/incidence supérieur de 10%, à âge égal, à la moyenne de l'ensemble des départements. La variabilité géographique du rapport ALD/incidence est majeure pour certaines localisations (par exemple, pour le mélanome ou les leucémies) et, à l'inverse, mineure pour d'autres localisations (par exemple pour les cancers des lèvres-bouche-pharynx). Pour les localisations à faibles effectifs (tableau), un manque de puissance peut conduire à une estimation de la variabilité nulle.

Le tableau résume les principaux résultats de la modélisation du rapport ALD/incidence et de la validation croisée. Pour les lymphomes malins non-hodgkiniens (LMNH), les départements alsaciens ont été exclus pour l'analyse, car ils avaient les deux rapports ALD/Incidence les plus faibles pour ce cancer, possiblement en lien avec la spécificité du régime d'assurance maladie de ces départements.

La variabilité géographique du rapport ALD/incidence est majeure ($\sigma_d \geq 0,09$) pour les localisations foie, vésicule, pancréas, larynx, mélanome, thyroïde, myélome et leucémies. Elle est importante ($\sigma_d \geq 0,07$) pour les localisations œsophage (hommes), estomac (femmes), col de l'utérus, rein (femmes). Elle est moins importante ($0,04 \leq \sigma_d \leq 0,06$) pour les localisations : côlon-rectum (hommes), poumon, sein, corps de l'utérus, ovaires, prostate, rein (hommes), système nerveux central (SNC), LMNH (hommes). Elle est mineure ou n'est pas mise en évidence ($\sigma_d \leq 0,03$) pour les localisations : lèvres-bouche-pharynx (LBP), œsophage (femmes), estomac (hommes), côlon-rectum (femmes), LMNH (femmes), testicules, vessie et maladie de Hodgkin.

Les erreurs maximales observées en validation croisée dépendent à la fois de σ_d et des effectifs, les localisations avec de faibles effectifs étant plus susceptibles de présenter des erreurs relatives importantes. Comme attendu, les erreurs sont particulièrement élevées et significatives dans un nombre élevé de départements pour les localisations avec une variabilité géographique du rapport ALD/incidence majeure (foie, vésicule, pancréas, larynx, mélanome, thyroïde, myélome, leucémies).

Soulignons qu'une variabilité géographique importante ($\sigma_d \geq 0,07$) du rapport ALD/incidence est parfois observée malgré de faibles effectifs (<100 cas/département), pour les cancers du foie chez la femme, de la thyroïde chez l'homme et de la vésicule biliaire.

Au sein des départements avec registre, les disparités géographiques d'incidence sont majeures ($\sigma_k \geq 0,17$) pour les localisations LBP, œsophage, foie, vésicule, larynx, pancréas, mélanome, testicule, thyroïde. À l'inverse, elles sont mineures pour les localisations sein, prostate et maladie de Hodgkin ($\sigma_k \leq 0,05$), et modérées pour côlon-rectum et SNC.

Sélection des localisations cancéreuses pour une estimation départementale de l'incidence

La sélection des localisations pour une estimation départementale de l'incidence, à partir des

Tableau Qualité des estimations : (1) variabilité géographique du rapport ALD/incidence (σ_d) et (2) comparaison de l'incidence observée et prédite (validation croisée), et ampleur des disparités géographiques d'incidence (σ_k). Départements avec registre, 2000-2005, France | **Table** Quality of estimations: (1) geographical variability of ALD/incidence ratio (σ_d) and (2) comparison of observed and predicted incidence (cross-validation), and extent of the geographical variations in incidence (σ_k). Districts covered by a cancer registry, 2000-2005, France

Localisations	Sexe	Nombre de départements	Cas	ALD	ALD /cas	σ_d	N ^a	ER max ^b (%)	σ_k ^c
Lèvres-bouche-pharynx	H	9	7 350	6 064	0,83	0,03	2	8	0,22
Lèvres-bouche-pharynx	F	9	1 599	1 282	0,80	0,00	0	14	0,08
Œsophage	H	11	3 297	2 705	0,82	0,07	2	23	0,27
Œsophage	F	11	632	497	0,79	0,00	0	27	0,22
Estomac	H	11	3 567	2 305	0,65	0,00	0	9	0,09
Estomac	F	11	2 025	1 373	0,68	0,08	3	26	0,11
Côlon-rectum	H	11	16 326	12 298	0,75	0,05	6	12	0,08
Côlon-rectum	F	11	13 247	10 527	0,79	0,03	4	8	0,06
Foie	H	11	4 477	2 332	0,52	0,14	7	35	0,28
Foie	F	11	1 035	672	0,65	0,07	5	37	0,19
Vésicule	H	11	773	406	0,53	0,13	5	37	0,21
Vésicule	F	11	1 015	596	0,59	0,09	4	33	0,21
Pancréas	H	11	2 888	2 050	0,71	0,12	5	38	0,15
Pancréas	F	11	2 587	1 895	0,73	0,09	3	31	0,19
Larynx	H	9	2 262	2 112	0,93	0,12	3	32	0,19
Poumon	H	9	16 435	11 486	0,70	0,05	4	14	0,10
Poumon	F	9	4 026	3 077	0,76	0,05	2	15	0,11
Mélanome	H	8	2 343	1 203	0,51	0,21	6	46	0,27
Mélanome	F	8	2 788	1 508	0,54	0,20	7	50	0,19
Sein	F	10	34 785	39 365	1,13	0,04	5	9	0,04
Col de l'utérus	F	9	2 168	2 062	0,95	0,08	2	34	0,16
Corps de l'utérus	F	9	4 223	3 186	0,75	0,05	2	17	0,14
Ovaire	F	9	3 243	2 981	0,92	0,04	2	12	0,11
Prostate	H	8	34 550	28 868	0,84	0,05	3	13	0,05
Testicule	H	9	1 376	1 437	1,04	0,00	1	15	0,17
Vessie	H	8	4 760	5 772	1,21	0,02	2	12	0,15
Vessie	F	8	1 037	1 260	1,22	0,01	0	14	0,04
Rein	H	9	4 084	3 283	0,80	0,04	1	19	0,10
Rein	F	9	2 114	1 853	0,88	0,07	2	16	0,14
Système nerveux central	H	10	1 860	1 733	0,93	0,06	3	23	0,08
Système nerveux central	F	10	1 532	1 500	0,98	0,05	2	23	0,07
Thyroïde	H	11	1 034	783	0,76	0,13	2	43	0,26
Thyroïde	F	11	3 359	2 694	0,80	0,13	5	30	0,33
Lymphome malin non-hodgkien ^d	H	9	3 613	2 579	0,71	0,04	4	16	0,10
Lymphome malin non-hodgkien ^d	F	9	3 245	2 331	0,72	0,00	0	7	0,10
Maladie de Hodgkin	H	11	792	781	0,99	0,00	0	13	0,06
Maladie de Hodgkin	F	11	669	715	1,07	0,00	0	15	0,06
Myélome multiple	H	11	1 962	1 574	0,80	0,14	4	45	0,13
Myélome multiple	F	11	1 681	1 470	0,87	0,16	2	65	0,11
Leucémies	H	11	3 923	3 444	0,88	0,14	5	51	0,14
Leucémies	F	11	2 907	2 706	0,93	0,10	5	31	0,10

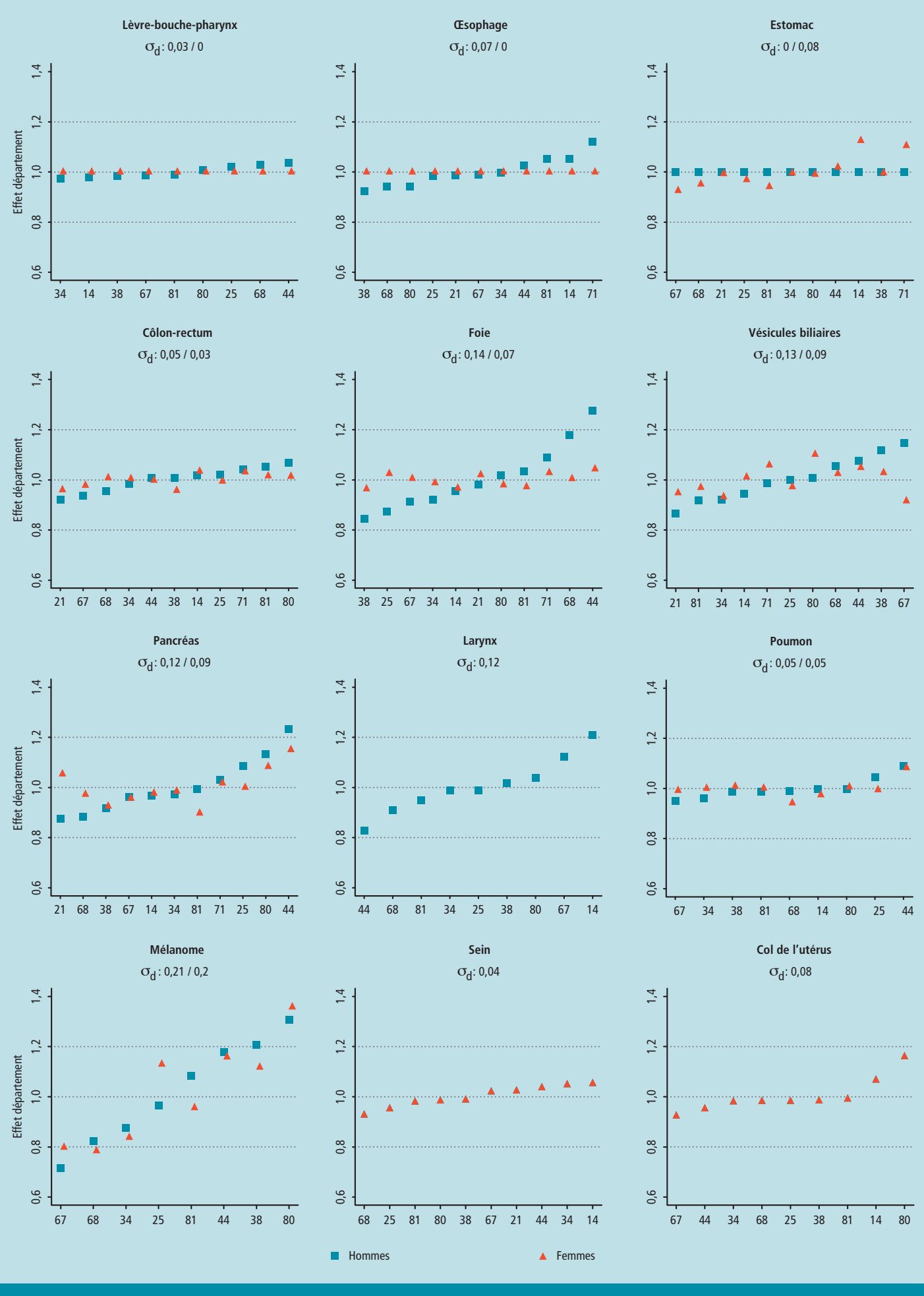
^a Validation croisée : nombre de départements avec une mesure d'écart K significativement trop élevé, avec $K=(O-P)^2/P$.

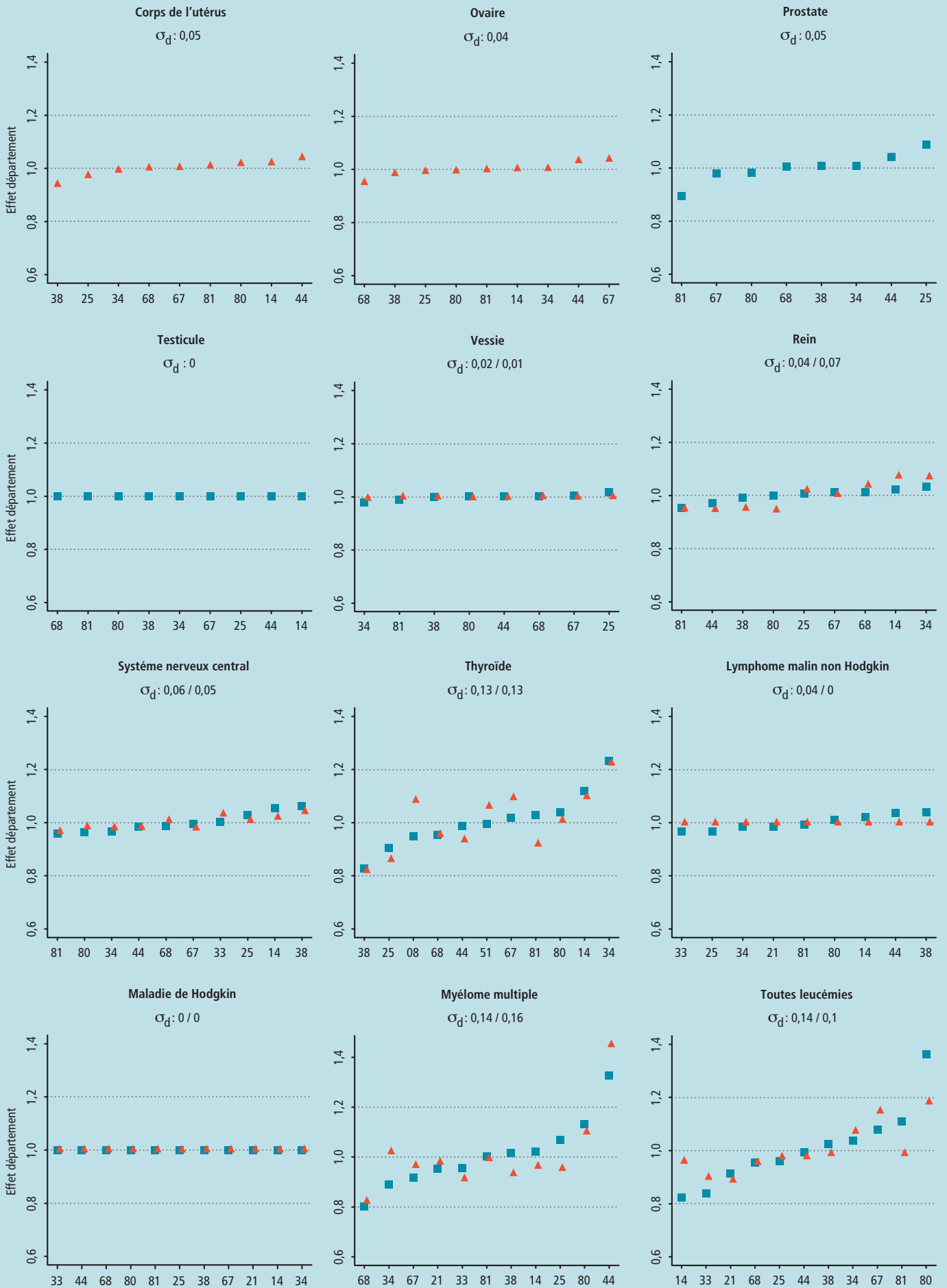
^b Validation croisée : erreur relative ER maximale (en valeur absolue) observée, avec $ER=(P-O)/O$.

^c Mesure des disparités géographiques d'incidence (σ_k) dans les départements avec registre.

^d Sans les départements du Haut-Rhin et du Bas-Rhin (sauf pour σ_k , qui est mesuré sur l'ensemble des registres).

Figure Variabilité géographique du rapport ALD/incidence dans les départements avec registre : effets départementaux et σ_d associé (hommes/femmes), 2000-2005, France / Figure Geographical variability of ALD/incidence ratio in the districts covered by a registry: district effect and σ_d (men/women), 2000-2005





analyses statistiques et de considérations épidémiologiques, est détaillée ci-après.

Variabilité géographique majeure du rapport ALD/incidence

Les localisations présentant une variabilité géographique majeure du rapport ALD/incidence ont été écartées d'emblée : foie, vésicule, pancréas, larynx, mélanome, thyroïde, myélome et leucémies.

Variabilité géographique intermédiaire du rapport ALD/incidence

Pour les localisations cancéreuses présentant une variabilité géographique intermédiaire du rapport ALD/incidence, le choix dépend de la validation croisée, de l'ampleur des variations géographiques d'incidence et du contexte épidémiologique.

Les localisations cancéreuses présentant une variabilité géographique importante du rapport ALD/incidence ont été écartées (col de l'utérus, rein/femmes), sauf deux exceptions (œsophage/hommes et estomac/femmes). Le cancer de l'œsophage a été sélectionné car la variabilité du rapport ALD/incidence est très inférieure à l'ampleur des variations géographiques de l'incidence (σ_d 4 fois plus petit que σ_k). Le cancer de l'estomac chez la femme a été sélectionné car un classement différent selon le sexe ne semblait pas justifié. De plus, les cancers de l'œsophage et de l'estomac étant fortement létaux, une confrontation aux données de mortalité, en termes de disparités géographiques, permettra de consolider les résultats.

Quatre localisations ont été sélectionnées : côlon-rectum/hommes, poumon, sein, ovaire. Le cancer du sein a été sélectionné du fait d'erreurs de prédiction modérées en validation croisée. Le cancer du côlon-rectum chez l'homme a été sélectionné car un classement différent selon le sexe ne semblait pas justifié et les erreurs de prédiction restaient acceptables. Le cancer du poumon et le cancer de l'ovaire ont été sélectionnés, avec des erreurs de prédiction atteignant jusqu'à 15% pour le poumon, car la variabilité géographique du rapport ALD/incidence est inférieure à l'ampleur des variations géographiques de l'incidence (σ_d 2 à 3 fois plus petit que σ_k). Comme pour les cancers de l'œsophage et de l'estomac, une confrontation aux données de mortalité permettra de consolider les résultats pour les cancers du poumon.

Six localisations ont été écartées (corps de l'utérus, prostate, rein/hommes, SNC, LMNH/hommes). Les cancers du corps de l'utérus, du rein et le LMNH chez l'homme n'ont pas été sélectionnés en raison d'erreurs de prédiction importantes en validation croisée. Pour les cancers du corps de l'utérus, les ALD codées « utérus, sans précision » représentent de plus une source de variabilité du rapport ALD/incidence. Le cancer de la prostate n'a pas été sélectionné en raison d'erreurs de prédiction importantes en validation croisée et de faibles variations géographiques de l'incidence. De plus, les cancers de la prostate diagnostiqués à un stade très précoce sont de plus en plus surveillés plutôt que traités. Cette évolution de la prise en charge est probablement source de variabilité du rapport ALD/incidence. Le cancer du système nerveux central n'a pas été sélectionné en raison d'erreurs de prédiction importantes en validation croisée et de

difficultés de codage (erreur de codage possible entre tumeurs *in situ* et bénignes et tumeurs invasives). La sensibilité des ALD pour identifier les cas incidents n'est que de 39% pour le SNC et la valeur prédictive positive de 59% [5].

Variabilité géographique mineure du rapport ALD/incidence

Les localisations cancéreuses présentant une variabilité géographique mineure du rapport ALD/incidence ont été d'emblée sélectionnées (LBP, estomac/hommes, côlon-rectum/femmes, testicules) sauf deux exceptions (vessie, LMNH/femmes). Le LMNH n'a pas été pris en compte chez la femme car un classement différent selon le sexe ne semblait pas justifié. De plus, la complexité et l'évolutivité des classifications des tumeurs en hématologie est un contexte défavorable pour l'utilisation de données médico-administratives.

Le cancer de la vessie n'a pas été sélectionné en raison de l'épidémiologie très particulière de ce cancer (grande diversité des formes évolutives, *continuum* des tumeurs bénignes vers les tumeurs malignes, récurrences fréquentes...) et des difficultés de codage qui en découlent. Un tiers environ des ALD pour cancer de la vessie correspondraient à des tumeurs *in situ* ou bénignes [5].

Enfin, 2 localisations avec des faibles effectifs (œsophage/femmes et Hodgkin) ont été considérées comme non exploitables et n'ont pas été sélectionnées.

Discussion

Cette étude présente pour la première fois une évaluation systématique, pour 24 localisations cancéreuses, de l'utilisation du rapport ALD/incidence pour estimer l'incidence départementale des cancers, sur la période 2000-2005. L'utilisation de ce rapport à l'échelle départementale repose sur l'hypothèse qu'il est identique dans tous les départements. Cette évaluation a été réalisée dans les départements couverts par un registre et s'est appuyée principalement sur l'évaluation de la qualité des estimations (variabilité géographique du rapport ALD/incidence et validation croisée), mais également, sur l'ampleur des variations géographiques d'incidence. Des considérations épidémiologiques ont pu également intervenir.

Pour la période 2000-2005, 8 localisations ont finalement été sélectionnées pour une estimation départementale de l'incidence : LBP, œsophage/hommes, estomac, côlon-rectum, poumon, sein, ovaire et testicule. Les estimations départementales pour ces localisations sont présentées dans une autre publication [4]. Pour le cancer de la prostate, le contexte défavorable a été un élément important dans le choix d'écarter cette localisation. En effet, les cancers diagnostiqués précocement et mis sous surveillance ne font vraisemblablement pas l'objet d'une demande d'ALD, et cette évolution de la prise en charge médicale ne s'est probablement pas développée de façon homogène sur tout le territoire. Une autre localisation (vessie) a été écartée malgré des résultats probants dans les départements avec registre, en raison des difficultés de codage spécifiques à ces tumeurs.

La principale limite de cette évaluation est le faible nombre de départements avec registre inclus dans l'analyse (une dizaine). La qualité des estimations,

appréciée dans ces seuls départements, peut parfois être mal évaluée. L'extrapolation du rapport ALD/incidence dans les départements non couverts par un registre est toujours un pari. Des éléments épidémiologiques mettant en cause l'hypothèse que ce rapport est identique sur le territoire ont ainsi contribué à écarter certaines localisations (vessie, prostate, SNC). La sélection des localisations, au vu des analyses statistiques et du contexte épidémiologique, a été établie et discutée au sein par l'ensemble des auteurs de l'étude.

L'ancienneté de la période d'étude (2000-2005) est également une limite de ce travail. Un nouveau parcours de soins a notamment été introduit en 2006, la demande d'ALD étant dorénavant établie par le médecin traitant, accompagné d'un protocole précisant les soins relevant de l'ALD. Cette évolution du dispositif des ALD a pu augmenter la fréquence des demandes d'ALD parmi les cas incidents, et peut-être diminuer la variabilité départementale du rapport ALD/incidence. Toutefois, les tendances nationales des taux d'ALD pour cancer ne présentent pas de brusque augmentation suite à cette évolution [6]. Il est peu probable que cette évolution conduise à un changement substantiel de la qualité des estimations départementales d'incidence obtenues en redressant les données ALD. Ce type d'évaluation doit néanmoins être mis à jour régulièrement, les déterminants de la variabilité géographique du rapport ALD/incidence pouvant évoluer avec le temps.

Par ailleurs, les demandes d'ALD parmi les cas incidents de cancers sont moins fréquentes chez les personnes âgées [4]. La prise en charge des personnes âgées est complexe et la variabilité géographique du rapport ALD/incidence pourrait ainsi être plus importante pour ce groupe. Une analyse restreinte aux personnes de moins de 75 ans a donc été réalisée : les résultats étaient peu modifiés dans l'ensemble. La variabilité géographique observée dans l'analyse principale ne semble pas provenir des personnes âgées en particulier.

Enfin, on remarquera que le rapport ALD/incidence est très variable selon la localisation cancéreuse étudiée, témoignant des multiples déterminants de la demande d'ALD et des contextes spécifiques à chaque localisation. Le nombre d'ALD est en général inférieur au nombre de cas incidents (rapport ALD/incidence <1), sauf pour les cancers du sein, de la vessie, et du SNC. Les données ALD incluent en effet, pour ces cancers, des tumeurs *in situ* ou bénignes nécessitant un traitement similaire aux tumeurs invasives et codées de ce fait en tumeur invasive dans les ALD. Toutefois, comme évoqué en introduction, le niveau du rapport ALD/incidence ne détermine pas directement la qualité des estimations départementales, sauf lorsqu'il est particulièrement faible (mélanome, foie, vésicule biliaire), cette configuration étant en général associée à une variabilité géographique majeure du rapport ALD/incidence.

Cette étude s'inscrit dans le cadre plus général de l'utilisation des données médico-administratives et de mortalité pour estimer l'incidence à un niveau infranational. Une étude similaire a également été menée sur les données hospitalières (PMSI) [7-9]. Le croisement individuel et anonyme des données PMSI et ALD est maintenant disponible et sera

étudié selon la même méthodologie. Ces études sont complémentaires des travaux qui ont effectué un croisement individuel entre les données des registres et les données ALD et PMSI [5].

L'objectif à terme est d'établir une stratégie de production d'estimations départementales et régionales de l'incidence des cancers à partir des différentes sources de données disponibles. Ceci implique en particulier de déterminer quelles sources utiliser, en fonction de la localisation cancéreuse étudiée, pour obtenir les meilleures estimations possibles, selon une méthodologie à définir.

Références

- [1] Belot A, Grosclaude P, Bossard N, Jouglu E, Benhamou E, Delafosse P, *et al.* Cancer incidence and mortality in France over the period 1980-2005. *Rev Epidemiol Santé Publique.* 2008;56:159-75.
[2] Belot A, Grosclaude P, Bossard N, Jouglu E, Benhamou E, Delafosse P, *et al.* Cancer incidence and mortality in France over the period 1980-2005.

Disponible à : <http://www.invs.sante.fr/surveillance/cancers>

- [3] Colonna M, Bossard N, Mitton N, Remontet L, Belot A, Delafosse P, *et al.* Éléments d'interprétation des estimations régionales de l'incidence du cancer en France sur la période 1980-2005. *Rev Epidemiol Santé Publique.* 2008;56(6): 434-40.

- [4] Uhry Z, Remontet L, Colonna M, Belot A, Grosclaude P, Mitton N, *et al.* Estimation départementale de l'incidence des cancers à partir des données d'affection de longue durée (ALD) et des registres. Évaluation par localisation cancéreuse, 2000-2005. Saint-Maurice: Institut de veille sanitaire; 2012. 101 p. (sous presse)

- [5] Grosclaude P, Dentan C, Trétarre B, Velten M, Fournier E, Molinié F. Utilité des bases de données médico-administratives pour le suivi épidémiologique des cancers Comparaison avec les données des registres au niveau individuel. *Bull Epidemiol Hebd.* 2012;(5-6):63-7.

- [6] Uhry Z, Remontet L, Grosclaude P, Belot A, Colonna M, Boussac-Zarebska M, *et al.* Tendances

récentes des données d'affections de longue durée : intérêt pour la surveillance nationale de l'incidence des cancers. Période 1997-2009. Saint-Maurice: Institut de veille sanitaire; 2011. 40 p. Disponible à : <http://www.invs.sante.fr/Publications-et-outils/Rapports-et-syntheses/Maladies-chroniques-et-traumatismes/2011/Tendances-recentes-des-donnees-d-affections-de-longue-duree-interet-pour-la-surveillance-nationale-de-l-incidence-des-cancers>

- [7] Mitton N, Colonna M, Trombert B, Olive F, Gomez F, Iwaz J, *et al.* A suitable approach to estimate cancer incidence in area without cancer registry. *J Cancer Epidemiol.* 2011;418968:1-11.

- [8] Remontet L, Mitton B, Couris CM, Iwaz J, Gomez F, Olive F, *et al.* Is it possible to estimate the incidence of breast cancer from medico-administrative databases? *Eur J Epidemiol.* 2008;23(10):681-8.

- [9] Bossard N, Gomez F, Remontet L, Uhry Z, Olive F, Mitton N, *et al.* Utilisation des données du PMSI pour estimer l'incidence des cancers en France à l'échelon infranational : la démarche du groupe Oncepi. *Bull Epidemiol Hebd.* 2012;(5-6):77-80.

Utilisation des données du PMSI pour estimer l'incidence des cancers en France à l'échelon infranational : la démarche du groupe Oncepi*

Nadine Bossard (nadine.bossard@chu-lyon.fr)¹, Frédéric Gomez², Laurent Remontet¹, Zoé Uhry³, Frédéric Olive⁴, Nicolas Mitton⁵, Stéphanie Polazzi⁶, Anne-Marie Schott⁶, Béatrice Trombert⁷, Marc Colonna^{5,8}

- 1/ Hospices civils de Lyon, Service de Biostatistique, Lyon, France ; Université de Lyon ; Université Lyon I, Villeurbanne, France ; CNRS ; UMR 5558, Laboratoire Biostatistique Santé, Pierre-Bénite, France
2/ Centre Léon Bérard, Département d'information médicale, Lyon, France
3/ Institut de veille sanitaire, Saint-Maurice, France
4/ Département d'information médicale, Centre hospitalier universitaire de Grenoble, France
5/ Registre du cancer de l'Isère, Grenoble, France
6/ Hospices civils de Lyon, Pôle Information médicale évaluation recherche, Lyon, France ; Université de Lyon, EA 4129 Laboratoire Santé individu société, Bron, France
7/ Service de santé publique et d'information médicale, CHU de Saint-Etienne, France
8/ Réseau des Registres de cancers Francim, Toulouse, France

* Observatoire épidémiologique Rhône-Alpes en oncologie

Résumé / Abstract

À ce jour, hormis les zones couvertes par des registres, aucune estimation de l'incidence du cancer n'est disponible au niveau départemental. L'objectif de l'Observatoire épidémiologique Rhône-Alpes en oncologie (groupe Oncepi) est de développer une méthodologie appropriée pour estimer l'incidence du cancer dans l'ensemble des départements français en utilisant, outre les données départementales des registres, les données du PMSI (Programme de médicalisation des systèmes d'information) qui sont exhaustives sur l'ensemble du territoire.

Afin d'atteindre cet objectif, une réponse aux deux questions suivantes devait être donnée : 1) Les données du PMSI peuvent-elles être utilisées en leur appliquant un facteur de correction basé sur leur sensibilité et leur spécificité ? 2) Si elles ne le peuvent pas, est-il possible d'utiliser les données du PMSI comme corrélat de l'incidence ?

Les différentes études menées par le groupe Oncepi ont abouti à la production d'estimations infranationales grâce à l'utilisation du rapport « Incidence PMSI/Incidence observée dans les registres ». Ce rapport est estimé par modélisation dans les départements avec registres, puis appliqué aux données du PMSI dans les départements sans registre. Des estimations départementales ont été produites pour les cancers du côlon-rectum et du sein. Plusieurs difficultés méthodologiques liées à la qualité des données du PMSI et à la modélisation ont été identifiées, mais certaines d'entre elles sont en cours de résolution : ceci rendra bientôt possible les estimations départementales pour d'autres cancers.

Les perspectives du groupe Oncepi sont de poursuivre les développements méthodologiques nécessaires, d'étudier (en lien avec l'Institut de veille sanitaire) l'intérêt des autres données médico-administratives disponibles au niveau national dans l'estimation de l'incidence des cancers à l'échelon infranational, et enfin d'étendre ses objectifs épidémiologiques à l'estimation de la prévalence hospitalière.

Using data from the French hospital database (PMSI) to estimate district-specific cancer incidence: the ONCEPI* group approach

To date, the estimation of district-specific cancer incidence is not available in France, except in fifteen districts (French Départements) covered by cancer registries. The objective of the Observatoire Epidémiologique Rhône-Alpes en Oncologie (ONCEPI group) is to develop an appropriate methodology to estimate cancer incidence in each French district, using jointly cancer registry data and the nationwide hospital data.

To reach this objective, two questions had to be addressed: 1) Can hospital data be used in combination with factors that amend their lack of sensitivity and specificity? 2) If not, is it possible to use hospital data as a correlate of incidence?

Several studies were carried out by the ONCEPI group. In particular, district-specific cancer incidence could be provided for some cancer sites (colon-rectum, breast) using the modelling of the relationship between cancer registry data and hospital data in districts with cancer registries, and inferring the incidence from district-specific hospital data in districts without registries. Several methodological difficulties have been encountered, regarding data quality and modelling, but some of them have been already solved: this will soon yield results about other cancer sites.

The perspectives of the ONCEPI group are now to deal with the remaining difficulties, explore the possibility of using other medico-administrative data with nationwide coverage to estimate district specific cancer incidence (together with the Institut de veille sanitaire), and extend the epidemiological objectives to estimate hospital prevalence.

Mots-clés / Key words

Cancer, incidence, données médico-administratives, PMSI, registres des cancers / Cancer, incidence, medical and administrative data, PMSI, cancer registries

Introduction

En France métropolitaine, l'enregistrement des cas de cancer est réalisé par une vingtaine de registres départementaux qui couvrent environ 20% du territoire. À partir des données de ces registres, des estimations nationales de l'incidence des cancers (tendances et projections) sont régulièrement produites depuis plusieurs années [1-4]. Ces estimations nationales sont calculées à partir des données de mortalité (utilisée comme corrélat de l'incidence) en extrapolant à la France entière le rapport entre incidence et mortalité observées dans l'ensemble des zones couvertes par les registres. Si ce rapport peut être raisonnablement utilisé pour des estimations France entière, il n'en est pas de même pour des estimations départementales car il est très variable d'un département à un autre et n'est donc pas pertinent. Les données du PMSI sont disponibles pour l'ensemble du territoire français, mais il est établi qu'elles ne peuvent pas, à l'heure actuelle, être utilisées pour mesurer directement l'incidence [5]. Cela est confirmé par P. Grosclaude et coll. [6], qui montrent leur faible valeur prédictive positive. À ce jour, hormis les zones couvertes par des registres, aucune estimation de l'incidence du cancer n'est disponible au niveau départemental. L'objectif de l'Observatoire épidémiologique Rhône-Alpes en oncologie (groupe Oncepi) est de développer une méthodologie appropriée pour estimer l'incidence du cancer dans l'ensemble des départements français en utilisant, de façon conjointe, les données départementales des registres et les données du PMSI. La réalisation de cet objectif permettra également de produire des estimations régionales.

Pour atteindre cet objectif, le groupe Oncepi a considéré plusieurs étapes visant à répondre aux questions suivantes : 1) les données du PMSI peuvent-elles être utilisées après application d'un facteur de correction basé sur la sensibilité et la spécificité, elles-mêmes estimées sur des données d'un registre ? – l'incidence réelle étant considérée à ce jour comme celle mesurée par les registres ; 2) Si non, est-il possible d'utiliser les données du PMSI comme corrélat de l'incidence ? – le rapport cas PMSI / Incidence registres étant estimé à partir de la zone registre et extrapolé aux différents départements, sous l'hypothèse d'un rapport stable d'un département à l'autre (contrairement au rapport Incidence / Mortalité).

Un tel travail nécessite la mise en commun de compétences dans divers domaines. C'est pourquoi le groupe Oncepi associe plusieurs équipes : le Registre général du cancer de l'Isère, les Départements d'information médicale du Centre Léon Bérard et des CHU de Grenoble, de Lyon et de Saint-Etienne, le Service de biostatistique des Hospices civils de Lyon et, plus récemment, l'Institut de veille sanitaire.

Nous rapportons ici les différentes étapes du travail mené par le groupe Oncepi et les conclusions de chacune d'entre elles, puis les perspectives du groupe.

Les données du PMSI peuvent-elles être utilisées après application d'un facteur de correction basé sur la sensibilité et la spécificité ?

Puisqu'il est bien établi que les données du PMSI ne peuvent constituer une mesure "directe" de

l'incidence [5;6], des méthodes de correction ont été développées. L'une d'entre elles a été explorée par le groupe Oncepi. Elle repose sur un facteur de correction qui utilise la sensibilité et la spécificité des données du PMSI [5]. Les performances des données du PMSI, en termes de spécificité et de sensibilité, ont été estimées pour le cancer du sein, dans le département de l'Isère, à partir des données des registres et des données PMSI de 2002. Le croisement des données a été effectué par appariement grâce à un identifiant personnel anonyme (logiciel Anonymat®). Deux algorithmes d'extraction des cas de cancer ont été testés : un algorithme large (le cancer du sein figure en diagnostic principal dans le PMSI) et un algorithme plus précis ("cancer du sein" en diagnostic principal + mention d'acte chirurgical en lien possible avec le cancer du sein). Le nombre de cas de cancers a été ensuite calculé pour chacun des autres départements avec registre à partir du nombre de cas PMSI corrigé par la sensibilité et la spécificité estimées à l'étape précédente. Ce nombre corrigé a été enfin comparé, pour chaque département avec registre, au nombre de cas fournis par le registre. Cette comparaison a montré que l'utilisation de la sensibilité et de la spécificité mesurées au niveau d'un département (précisément l'Isère) ne conduisait pas à des estimations correctes de l'incidence dans les autres départements avec registre. Ce résultat pourrait s'expliquer par une hétérogénéité des pratiques médicales et/ou des pratiques de codage des actes médicaux dans le PMSI. Par ailleurs, au sein du PMSI, la distinction entre cas prévalents et cas incidents est difficile, et des problèmes d'identification des cas sont survenus lors de l'estimation de la sensibilité et de la spécificité. Ces difficultés, qui ont contribué à mettre en échec la méthode utilisée dans cette étude, se sont probablement atténuées depuis 2002. Toutefois, cette approche reste lourde à appliquer parce que le croisement des données est très fastidieux. Dans la mesure où la question centrale est de diminuer l'effet de l'hétérogénéité inhérente aux données du PMSI, une alternative était d'utiliser les données du PMSI comme corrélat de l'incidence et d'explorer une méthode reposant sur la modélisation du rapport données PMSI / Incidence registres sur l'ensemble de la zone des registres. L'homogénéité interdépartementale de ce rapport est l'hypothèse majeure de cette approche. Cette hypothèse est forte, comme le suggère cette toute première étude, mais la possibilité de modéliser statistiquement le rapport au niveau de l'ensemble des départements avec registres offrait des perspectives intéressantes pour son exploration.

La première étape n'ayant pas positivement abouti, est-il possible d'utiliser les données du PMSI comme corrélat de l'incidence ?

Le rapport données PMSI / Incidence registres dans les départements de la zone registre a donc été ensuite estimé par une approche de type calibration qui suppose que les données du PMSI mesurent l'incidence avec une certaine erreur. Un modèle mixte (avec effet aléatoire sur le département) incluant l'âge comme variable continue a permis d'estimer le rapport données PMSI / Incidence registres sur la zone registre.

Cette modélisation présentait l'avantage de gommer les fluctuations aléatoires de l'effet de l'âge, améliorant ainsi les qualités prédictives de l'approche. L'application de ce rapport aux données du PMSI permettait ainsi, théoriquement, une estimation de l'incidence dans l'ensemble des départements (avec ou sans registre). Dans les 12 départements avec registre qui ont participé à l'étude (pour lesquels on disposait donc à la fois de l'incidence observée au sein du registre et de l'incidence estimée par le modèle), il a été possible de calculer l'erreur de prédiction du modèle en utilisant la méthode de validation croisée.

Cette méthode a été utilisée dans un premier temps pour estimer le nombre de cas de cancer du sein à partir des données PMSI de 2002. Le chaînage des séjours n'étant pas encore utilisable pour cette année-là, seul le nombre de séjours était exploitable et non le nombre de cas : c'est donc le rapport Séjours PMSI / Incidence registres qui a été estimé [7]. Ici aussi, deux algorithmes d'extraction des données PMSI ont été explorés (Algorithme 1 : "cancer du sein" comme diagnostic principal et Algorithme 2 : "cancer du sein" comme diagnostic principal + mention d'acte chirurgical en lien possible avec le cancer du sein). Les principaux résultats ont montré que les prédictions du modèle étaient satisfaisantes sur la zone registre quand l'Algorithme 1 (le plus large) était utilisé. Cela signifiait une certaine homogénéité du rapport Séjours PMSI / Incidence registres au niveau de la zone registre. Ce résultat était attendu parce que l'algorithme le plus large est, par nature, moins sujet à l'hétérogénéité de codage. Cependant, dans les autres départements, on pouvait parfois noter une grande variabilité du nombre de séjours par patient (variabilité explorée seulement pour les séjours pour lesquels l'identification du patient était possible), ce qui remettait complètement en cause l'utilisation du nombre de séjours.

Le modèle a donc été ensuite appliqué aux données de séjours chaînés de 2004, seule année pour laquelle les données du PMSI et des registres étaient disponibles de façon concomitante, autorisant ainsi l'estimation du rapport cas PMSI / Incidence registres, où le numérateur était le nombre de patients repérés par le PMSI et non plus le nombre de séjours. Ce travail a été effectué pour les cancers du côlon-rectum, du sein, du rein, et de l'ovaire [8]. Les prédictions étaient satisfaisantes sur la zone registre avec l'Algorithme 1 pour le cancer du sein et du côlon-rectum. Des estimations départementales ont donc été produites pour l'année 2007 en appliquant le rapport cas PMSI / Incidence registres de 2004 aux données PMSI de 2007 (le rapport a donc été supposé constant de 2004 à 2007, ce qui apparaissait raisonnable). Il n'en était pas de même pour les cancers à plus faible incidence, comme celui du rein ou de l'ovaire, pour lesquels l'étape de validation croisée, essentielle pour la mise en évidence d'un effet "département", n'a pu être appliquée. Dans ces cas, seule une exploration graphique était possible, ce qui n'était pas satisfaisant.

Ainsi, il a été possible de fournir des estimations départementales pour 2 cancers sur les 4 étudiés

(tableau 1), mais l'hypothèse d'une homogénéité du rapport cas PMSI / Incidence registres reste cruciale et peut s'avérer difficile à mettre en évidence.

Les causes de cette hétérogénéité ont été finement analysées par le groupe Oncepi [9]. Les principales conclusions ont été que, même en présence d'un chaînage des séjours, le codage des cancers est parfois difficile par nature (cancers de la vessie, de l'ovaire, cancers ORL notamment) et les règles de codage du diagnostic principal peuvent être plus ou moins maîtrisées selon les structures de soin, ces mêmes règles pouvant de plus évoluer dans le temps. Il n'est pas exclu non plus que les règles de production de l'information du PMSI pour la facturation affectent l'homogénéité du codage. La difficulté de distinguer les cas incidents des cas prévalents au sein des données du PMSI reste importante. Cependant, le chaînage des hospitalisations

étant devenu tout à fait performant, une approche longitudinale retraçant la trajectoire du patient est devenue possible et permettra, à l'avenir, d'atténuer cette difficulté [8;9]. En ce qui concerne l'évaluation de la qualité du chaînage des séjours hospitaliers, nous invitons le lecteur à prendre connaissance du récent article de A. Bocquier et coll. [10].

Perspectives

L'expérience du groupe Oncepi est encourageante, puisque des estimations départementales d'incidence du cancer du sein et du côlon-rectum pour une année donnée ont pu être produites à partir de données du PMSI. Par ailleurs, certaines contraintes qui existaient au début des activités du groupe se sont assouplies. D'une part, la qualité du chaînage s'est grandement améliorée dans le temps et, d'autre part, l'utilisation concomitante des données du PMSI

et des registres est devenue possible sur des périodes plus longues, permettant ainsi d'augmenter les effectifs et de diminuer les fluctuations aléatoires pour certaines localisations de faible incidence. C'est pourquoi, des estimations départementales pour d'autres localisations sont en cours de réalisation (non publiées). Les questions et les perspectives restent cependant nombreuses.

Une méthode de validation des prédictions reste à développer pour les cancers de faible incidence. Rappelons ici que cette étape de validation est essentielle pour apprécier l'hypothèse d'homogénéité interdépartementale du rapport cas PMSI / Incidence registres. Cette hypothèse reste l'élément le plus fragile car, même avec des données correctement chaînées, un codage hétérogène est toujours possible (difficultés de codage inhérentes au cancer étudié ou application hétérogène des règles de codage

Tableau 1 Taux d'incidence estimés standardisés monde (pour 100 000) pour les cancers du côlon-rectum et du sein dans les départements français en 2007 / **Table 1** Estimated world-standardized incidence rates (for 100,000) for colon-rectum and breast cancers in French districts in 2007

Département	Code	Côlon-rectum		Sein	Département	Code	Côlon-rectum		Sein
		Homme	Femme				Homme	Femme	
Ain	01	35,1	25,2	82,0	Maine-et-Loire	49	42,8	25,7	127,1
Aisne	02	44,1	26,1	103,2	Manche	50	32,6	18,6	86,8
Allier	03	38,9	27,4	81,8	Marne	51	38,9	27,2	106,6
Alpes-de-Haute-Provence	04	43,5	24,3	98,1	Haute-Marne	52	41,2	18,4	99,4
Hauts-Alpes	05	35,9	17,4	94,3	Mayenne	53	42,2	20,4	91,4
Alpes-Maritimes	06	37,0	26,1	98,6	Meurthe-et-Moselle	54	41,7	26,5	109,9
Ardèche	07	44,3	26,0	84,9	Meuse	55	34,8	31,3	101,7
Ardennes	08	43,6	19,5	95,7	Morbihan	56	38,1	23,3	92,0
Ariège	09	40,5	25,8	80,7	Moselle	57	41,6	24,4	101,8
Aube	10	28,8	23,3	84,1	Nièvre	58	48,6	22,8	80,6
Aude	11	35,2	24,3	95,6	Nord	59	41,6	26,2	102,0
Aveyron	12	30,3	23,0	80,3	Oise	60	38,8	26,3	93,7
Bouches-du-Rhône	13	37,7	26,5	111,6	Orne	61	41,5	27,4	112,6
Calvados	14	36,6	20,3	92,5	Pas-de-Calais	62	40,5	27,7	97,6
Cantal	15	27,3	20,5	83,3	Puy-de-Dôme	63	42,2	24,7	87,7
Charente	16	38,2	24,8	93,5	Pyrénées-Atlantiques	64	37,4	22,9	91,8
Charente-Maritime	17	43,8	27,0	92,5	Hautes-Pyrénées	65	39,6	23,8	83,8
Cher	18	40,2	27,7	94,9	Pyrénées-Orientales	66	40,9	24,0	96,3
Corrèze	19	32,9	22,9	84,7	Bas-Rhin	67	38,1	23,5	97,7
Côte-d'Or	21	35,7	22,7	84,9	Haut-Rhin	68	35,3	25,3	86,8
Côtes-d'Armor	22	40,3	28,7	95,1	Rhône	69	39,9	24,6	99,3
Creuse	23	41,0	27,0	83,2	Haute-Saône	70	42,0	28,5	100,5
Dordogne	24	40,8	22,8	91,6	Saône-et-Loire	71	43,4	25,3	97,9
Doubs	25	34,5	24,7	89,6	Sarthe	72	31,2	19,8	103,9
Drôme	26	37,3	28,3	97,9	Savoie	73	35,3	23,5	88,8
Eure	27	43,4	25,2	97,2	Haute-Savoie	74	29,0	17,8	96,3
Eure-et-Loir	28	38,2	25,9	101,1	Paris	75	35,8	25,5	117,6
Finistère	29	39,3	24,0	94,4	Seine-Maritime	76	34,9	24,3	103,5
Corse-du-Sud	2A	48,9	29,2	79,5	Seine-et-Marne	77	37,5	30,2	101,6
Haute-Corse	2B	23,1	22,5	107,1	Yvelines	78	34,9	26,1	107,1
Gard	30	40,8	23,6	109,0	Deux-Sèvres	79	44,1	23,4	91,3
Haute-Garonne	31	34,4	22,5	96,8	Somme	80	44,6	23,6	104,7
Gers	32	34,6	26,0	79,7	Tarn	81	28,7	21,2	87,5
Gironde	33	37,9	28,0	93,6	Tarn-et-Garonne	82	34,1	23,6	108,8
Hérault	34	38,6	22,7	94,0	Var	83	36,3	24,6	109,1
Ille-et-Vilaine	35	30,8	19,6	90,3	Vaucluse	84	45,6	25,7	101,6
Indre	36	34,6	27,6	86,8	Vendée	85	40,2	25,3	103,6
Indre-et-Loire	37	35,0	27,5	108,8	Vienne	86	39,7	20,1	93,7
Isère	38	35,9	23,4	96,3	Haute-Vienne	87	39,6	24,8	89,3
Jura	39	36,3	19,5	81,8	Vosges	88	37,7	25,3	89,6
Landes	40	37,5	24,4	92,7	Yonne	89	43,2	25,4	99,5
Loir-et-Cher	41	37,9	27,6	104,9	Territoire-de-Belfort	90	43,6	26,6	88,0
Loire	42	40,7	27,1	97,5	Essonne	91	35,1	26,3	101,3
Haute-Loire	43	43,0	25,3	77,9	Hauts-de-Seine	92	35,6	25,5	100,8
Loire-Atlantique	44	39,8	23,7	108,4	Seine-Saint-Denis	93	37,7	26,3	100,0
Loiret	45	40,0	22,7	96,7	Val-de-Marne	94	35,6	24,6	104,2
Lot	46	35,3	22,3	83,3	Val-d'Oise	95	34,1	23,2	103,9
Lot-et-Garonne	47	32,3	24,2	71,4					
Lozère	48	39,8	20,9	98,1	France métropolitaine		38,0	24,8	98,6

éventuellement renforcée par des évolutions réglementaires non uniformément et rapidement appliquées). Par ailleurs, les données du PMSI ne sont évidemment pas utilisables pour des cancers dont la prise en charge est ambulatoire (le mélanome, par exemple). Ainsi, des critères d'éligibilité devront donc être formalisés *in fine* afin de rendre compte de la faisabilité ou non de produire des estimations départementales pour chaque localisation.

Ensuite, les données du PMSI, même de bonne qualité, ont une antériorité limitée par rapport à celles des registres, ce qui rend incomplète leur exploitation à visée épidémiologique, notamment pour l'étude des tendances sur de longues périodes.

L'utilisation d'autres sources médico-administratives doit être bien sûr envisagée. Cela a été réalisé par Uhry et coll. [10] avec les données ALD (Affections longue durée) dans une étude qui a utilisé la même approche méthodologique que celle retenue par le groupe Oncepi pour les données du PMSI (reposant donc sur l'utilisation du rapport Incidence ALD / Incidence registres). Des estimations départementales ont pu être produites pour huit localisations cancéreuses : lèvre-bouche-pharynx, œsophage (hommes), estomac, côlon-rectum, poumon, sein, ovaire, testicule. Il semble raisonnable de considérer qu'aucune donnée médico-administrative ne pourra, seule, fournir des estimations infranationales des incidences de toutes les localisations cancéreuses. En effet, si la question de l'hétérogénéité du rapport se pose de la même façon pour les données ALD et pour celles du PMSI, les raisons de cette hétérogénéité peuvent être différentes et donc ne pas affecter les mêmes localisations. Il sera donc nécessaire de choisir, pour chaque localisation, la source de données à utiliser : données du PMSI et/ou données ALD. La comparaison des estimations départementales obtenues d'après les données médico-administratives constitue de fait une étape essentielle pour la mise à disposition d'estimations départementales. En parallèle, quelques développements méthodologiques et statistiques doivent être envisagés pour améliorer la valeur prédictive des modèles utilisés. L'Institut de veille sanitaire, qui est à l'origine des travaux d'estimations départementales à partir des ALD, a récemment rejoint le groupe Oncepi. Une réflexion concernant l'utilisation conjointe des données du PMSI et des ALD devrait être menée à terme.

Dans le but de compléter l'utilisation épidémiologique des données du PMSI, l'indicateur que constitue la prévalence est en cours d'étude au sein du groupe Oncepi. Une première étude de la prévalence dite «hospitalière» à partir des données PMSI seules a été menée et publiée pour le cancer du sein [12]. La prévalence hospitalière dénombre les personnes ayant bénéficié de soins pour une pathologie cancéreuse au cours d'une période donnée. Ces estimations de prévalence hospitalière ont été ensuite confrontées aux estimations de prévalence totale et de prévalence partielle. On rappelle que la prévalence totale correspond au nombre (ou à la proportion) de personnes atteintes ou ayant été atteintes d'une pathologie cancéreuse et vivantes à une date donnée, alors que la prévalence partielle limite le groupe constituant la prévalence totale aux personnes dont le diagnostic a été réalisé depuis une période de temps limitée (par exemple depuis 1, 3, 5 ou 10 ans au plus). Dans les deux cas, le nombre de personnes ayant eu recours au système de soins hospitalier pour leur cancer n'est pas quantifiable de manière précise. De ce fait, ces estimations de prévalence hospitalière illustrent l'intérêt des données du PMSI dans l'évaluation des besoins médicaux des personnes atteintes d'un cancer du sein. L'estimation de la prévalence hospitalière pour d'autres localisations cancéreuses constitue également un axe de travail du groupe Oncepi.

Conclusion

Le groupe Oncepi a su proposer une approche méthodologique pour des estimations infranationales de l'incidence des cancers en utilisant les données du PMSI, avec une analyse détaillée des avantages et des limites de ces données. Dans le même esprit, et dans le but de compléter l'utilisation épidémiologique des données du PMSI, l'analyse de la prévalence est en cours.

Remerciements

Le groupe Oncepi a bénéficié d'une aide de l'INCa dans le cadre de l'Appel à projets 2007 (« Pour une observation des cancers au niveau régional ») et de l'Appel à projets 2010 (« Recherche en sciences humaines et sociales, en santé publique et épidémiologie »).

Le groupe Oncepi remercie Jean Iwaz pour sa relecture attentive de cet article.

Références

[1] Danzon A, Le Moal J, Chérié-Challine L, Viso AC. La surveillance épidémiologique des cancers en France :

outils actuels et perspectives. Bull Epidemiol Hebd. 2012;(5-6):54-8.

[2] Remontet L, Estève J, Bouvier AM, Grosclaude P, Launoy G, Menegoz F, *et al.* Cancer incidence and mortality in France over the period 1978-2000. Rev Epidemiol Santé Publique. 2003;51(1 Pt 1):3-30.

[3] Belot A, Grosclaude P, Bossard N, Jouglu E, Benhamou E, Delafosse P, *et al.* Cancer incidence and mortality in France over the period 1980-2005. Rev Epidemiol Santé Publique. 2008;56(3):159-75.

[4] Remontet L, Belot A, Bossard N. Tendances de l'incidence et de la mortalité par cancer en France et projections pour l'année en cours : méthodes d'estimation et rythme de production. Bull Epidemiol Hebd. 2009;(38):405-8.

[5] Couris CM, Polazzi S, Olive F, Remontet L, Bossard N, Gomez F, *et al.* Breast cancer incidence using administrative data: correction with sensitivity and specificity. J Clin Epidemiol. 2009;62(6):660-6.

[6] Grosclaude P, Dentan C, Trétarre B, Velten M, Fournier E, Molinié F. Utilité des bases de données médico-administratives pour le suivi épidémiologique des cancers. Comparaison avec les données des registres au niveau individuel. Bull Epidemiol Hebd. 2012;(5-6):63-7.

[7] Remontet L, Mitton N, Couris CM, Iwaz J, Gomez F, Olive F, *et al.* Is it possible to estimate the incidence of breast cancer from medico-administrative databases? Eur J Epidemiol. 2008;23(10):681-8.

[8] Mitton N, Colonna M, Trombert B, Olive F, Gomez F, Iwaz J, *et al.* A suitable approach to estimate cancer incidence in area without cancer registry. J Cancer Epidemiol 2011; 418968:1-11.

[9] Olive F, Gomez F, Schott AM, Remontet L, Bossard N, Mitton N, *et al.* Critical analysis of French DRG based information system (PMSI) databases for the epidemiology of cancer: a longitudinal approach becomes possible. Rev Epidemiol Santé Publique. 2011;59(1):53-8. Erratum in: Rev Epidemiol Santé Publique. 2011;59(3):199.

[10] Bocquier A, Thomas N, Zitouni J, Lewandowski E, Cortaredona S, Jardin M, *et al.* Évaluation de la qualité du chaînage des séjours hospitaliers pour l'étude des variations spatiales de santé à partir des données du PMSI. Étude de faisabilité dans trois régions françaises. Rev Epidemiol Santé Publique 2011;59(4):243-9.

[11] Uhry Z, Remontet L, Colonna M, Belot A, Grosclaude P, Mitton N, *et al.* Évaluation de l'utilisation des données d'affection de longue durée (ALD) pour estimer l'incidence départementale des cancers pour 24 localisations cancéreuses, France, 2000-2005. Bull Epidemiol Hebd. 2012;(5-6):71-7.

[12] Trombert Paviot B, Gomez F, Olive F, Polazzi S, Remontet L, Bossard N, *et al.* Identifying prevalent cases of breast cancer in the French case-mix databases. Methods Inf Med. 2011;50(2):124-30.

Système multi-sources de surveillance des cancers (SMSC) : bilan de l'étude pilote

Yao Kudjawa (y.kudjawa@invs.sante.fr), Céline Caserio-Schönemann, Agnès Guillet, Javier Nicolau, Arnaud Musset, Jérémie Rudant, Arlette Danzon, Laurence Chérié-Challine

Institut de veille sanitaire, Saint-Maurice, France

Résumé / Abstract

Le développement d'un système de surveillance nationale, le système multi-sources cancer (SMSC), en complément du système reposant sur les registres a été recommandé par la commission d'évaluation des conséquences sanitaires de l'accident de Tchernobyl et affirmé par les Plans cancer gouvernementaux successifs.

Le principe repose sur le croisement de données anonymisées de trois sources : Programme de médicalisation des systèmes d'information (PMSI), affections de longue durée (ALD) et comptes-rendus d'anatomie et cytologie pathologiques (ACP). L'objectif de l'étude pilote était d'étudier la faisabilité de la mise en place du SMSC. Elle a porté sur le cancer de la thyroïde dans deux régions : Île-de-France et Nord-Pas-de-Calais. Le recueil des données du compte-rendu ACP a été réalisé rétrospectivement (2004) grâce à un formulaire informatisé.

Au total, 423 fiches de données ACP de cancer de la thyroïde ont été transmises par les anatomo-cytopathologistes volontaires des deux régions. Elles ont été croisées avec celles des ALD (1 159) et du PMSI (2 947) de l'année 2004. Le taux d'appariement (nombre de cas ACP retrouvés dans les autres sources) observé était respectivement de 84% (les 3 sources), de 81% (ACP, PMSI) et de 48% (ACP, ALD).

La mise en place du système a mis en évidence le fait que la solution informatique retenue était trop complexe pour un développement au niveau national. Cette complexité étant liée aux difficultés techniques rencontrées au cours du projet.

À l'issue du pilote, deux axes de développement du SMSC au niveau national ont été identifiés :

- dans l'immédiat, croisement des 2 sources PMSI et ALD, à des fins descriptives et de production d'indicateurs de surveillance ;
- à moyen terme et en parallèle avec le premier axe, étudier les conditions de mise en place de la base nationale de données ACP.

Mots clés / Key words

Cancer, surveillance, SMSC, données médico-administratives, ACP / Cancer, surveillance, MSSC, medical and administrative data, ACP

Introduction

En France, le système de surveillance des cancers repose principalement sur les registres de cancer et l'enregistrement des causes médicales de décès par le Centre d'épidémiologie sur les causes médicales de décès (CépiDc) de l'Institut national de la santé et de la recherche médicale (Inserm). Les registres de cancer assurent l'enregistrement actif et exhaustif des cas incidents de cancer dans une zone géographique délimitée (en général le département) et couvrent actuellement environ 20% de la population [1]. Des estimations nationales et régionales de l'incidence sont produites à partir des données de mortalité du territoire considéré, en extrapolant à la France le rapport entre incidence et mortalité observées dans les départements couverts par un registre de cancer [2;3].

En complément à ce système de surveillance, des travaux basés sur l'utilisation des données des bases médico-administratives de l'assurance maladie ou du Programme de médicalisation des systèmes d'information (PMSI) se développent [4-6]. Les données de ces bases médico-administratives présentent des intérêts de centralisation au niveau national et de rapide

disponibilité ainsi qu'une couverture nationale, mais également des limites relatives à l'exhaustivité de leurs données.

L'objectif du système multi-sources cancer (SMSC) est de permettre une surveillance épidémiologique des cancers sur l'ensemble du territoire et avec réactivité. Cet objectif a été décrit dans le premier Plan cancer gouvernemental 2003-2007 [7] et réaffirmé dans le second [8]. Il vise à renforcer le dispositif actuel de surveillance des cancers en France suite aux travaux de la commission multidisciplinaire d'évaluation des conséquences sanitaires de l'accident de Tchernobyl en France, mise en place et pilotée par l'Institut de veille sanitaire (InVS) [9;10].

Le principe du SMSC est de croiser de façon automatisée et passive et à un niveau individuel les trois sources de données suivantes : 1) les données issues du PMSI, 2) les données de l'assurance maladie relatives aux exonérations du ticket modérateur pour cancer, correspondant à l'affection longue durée n°30 ou ALD cancer, 3) les données issues des comptes-rendus d'anatomie et cytologie pathologique (ACP) qui permettent d'attester du diagnostic de cancer.

Multi source surveillance system of cancers (MSSC): assessment of the pilot study

The development of a national multi source surveillance system of cancers (MSSC) in addition to the registry-based system has been recommended by the Assessment Commission on Health Effects of the Chernobyl Accident, and asserted by the successive governmental cancer plans.

The principle is to cross anonymous data from three sources: hospital discharge database (PMSI), long-term illness (LTI) database and pathologist's report called ACPs. The objective of the pilot study was to study the feasibility of the MSSC implementation. The study concerned thyroid cancer in two French regions: Ile-de-France and Nord-Pas-de-Calais.

ACP data were retrospectively (2004) collected by electronic form.

Four hundred and twenty-three (423) forms of ACP data on thyroid cancer were transmitted by pathologists of both regions on a voluntary basis. They were crossed with data on LTI thyroid cancer (1,159) and PMSI thyroid cancer (2,947) for the year 2004. The rate of matched data (number of ACP cases retrieved in other sources) observed was 84% (3 sources), 81% (PMSI, ACP), and 48% (ACP, LTI). The pilot showed that the implementation of the system at national level would be too complex due to technical problems related to technical computer systems.

At the end of the pilot, two lines of MSSC development at national level were identified:

- *In the short term, crossing both PMSI and LTI cancer sources, for descriptive purposes, and in order to produce surveillance indicators.*
- *In the mid term, and in parallel with the first line, study the conditions of ACP data implementation at national level.*

L'objectif de l'étude pilote était d'étudier la faisabilité du croisement des trois sources et la pertinence de la mise en place du SMSC [11]. Cette étude pilote conduite par l'InVS a porté sur le cancer de la thyroïde et s'est déroulée de 2004 à 2009. Ses objectifs spécifiques étaient d'étudier la faisabilité de :

- la mise en place d'une solution informatique pour le recueil des données ACP, leur anonymisation, leur transfert depuis les structures publiques et privées jusqu'à l'InVS ;
- l'appariement des trois sources de données ACP, PMSI et ALD à l'aide d'un numéro d'anonymat commun ;
- la concordance entre les trois sources.

Nous rapportons dans cet article le déroulement et les résultats de cette étude pilote et les perspectives envisagées pour le dispositif.

Matériel et méthode

Sites pilotes

L'étude pilote ciblait deux régions, l'Île-de-France (IdF) et le Nord-Pas-de-Calais (NPdC) définies par le Plan cancer au regard :

- des priorités de santé publique dans ces deux régions ;

- du développement, en parallèle, de deux études de faisabilité pour la mise en place de registres de cancer en zones urbaines (Val-de-Marne en région IdF et région lilloise dans le NPdC), le SMSC devant constituer une alternative aux registres en cas d'échec.

Afin de réaliser l'étude pilote dans les conditions les plus proches possibles de la situation des ACP au niveau national, les structures ACP volontaires des deux régions ayant une activité en cancérologie de la thyroïde ont été sélectionnées sur les caractéristiques suivantes de représentativité : secteur d'activité (privé et public), taille des structures (grosse, petite) et variété des équipements informatiques « métier ». Sur cette base, en IdF, les structures ACP ont été sélectionnées sur proposition du Centre de regroupement informatique et statistique en ACP (Crisap)¹. Dans le NPdC, en l'absence de Crisap, le choix des ACP a été fait lors d'une réunion de présentation du projet par l'InVS aux pathologistes en présence de la direction régionale des affaires sanitaires et sociales (Drass).

Huit structures ACP, cinq en IdF (Institut Gustave Roussy, Groupe Hospitalier Pitié-Salpêtrière, Hôpital Ambroise Paré, cabinet de pathologie Tolbiac et Institut de pathologie de Paris) et trois dans le NPdC (CHRU de Lille, Groupe hospitalier de l'Institut catholique de Lille et le laboratoire de Valenciennes) ont été incluses sur la base du volontariat pour participer à l'étude pilote.

Justification de la localisation cancéreuse étudiée

La localisation cancéreuse étudiée était le cancer de la thyroïde. Le choix de cette localisation était cohérent avec les travaux de la commission pluridisciplinaire qui faisait le constat d'un dispositif de surveillance épidémiologique insuffisant pour juger de l'impact de l'accident de Tchernobyl en France sur l'incidence des cancers thyroïdiens. Bien que non considérée comme une priorité de santé publique, cette localisation présentait des qualités pour tester le SMSC : faible incidence, donc faible enregistrement et faible charge de travail pour les ACP, diagnostic histologique systématique (donc passage par les ACP), prise en charge chirurgicale (donc séjour présent dans le PMSI).

Sources de données, recueil d'informations et anonymisation des données

Les données ACP concernaient les cas de cancer de la thyroïde survenus chez des patients résidant en IdF et dans le NPdC diagnostiqués en 2004 au sein de l'une des huit structures.

Les données ALD concernaient les patients domiciliés en IdF ou dans le NPdC et mis en ALD en 2004 pour un cancer thyroïdien (code CIM10 : C73). Elles ont été extraites du Système national d'information interrégimes de l'assurance maladie (Sniiram).

Les données PMSI ont été extraites de la base 2004 du PMSI national et concernaient les

séjours de patients domiciliés en IdF ou dans le NPdC avec un résumé de sortie anonyme mentionnant en diagnostic principal, associé ou relié, un cancer de la thyroïde (CIM10 : C73).

Les données ACP recueillies ont été anonymisées grâce à la fonction Foin² déjà utilisée pour les données PMSI et ALD. La transmission sécurisée de ces données était assurée par le logiciel Camel³. Le croisement des trois sources de données a été réalisé grâce à l'identifiant anonyme commun.

Réseau de partenaires impliqués et remontée des données à l'InVS

Les partenaires impliqués dans l'étude pilote se situaient à trois niveaux géographiques :

- au niveau local, les huit structures ACP recueillaient rétrospectivement les comptes-rendus ACP de cancer de la thyroïde diagnostiqués en 2004 pour les transmettre au niveau régional via le circuit informatique sécurisé mis en place ;
- au niveau régional, le Crisap d'IdF (Crisapif) et les Drass d'IdF et du NPdC anonymisaient les données ACP (premier niveau) grâce à une clé propre, et les transféraient au niveau national ;
- au niveau national, la Caisse nationale d'assurance maladie des travailleurs salariés (CnamTS) anonymisait une deuxième fois, grâce à une clé propre, les données identifiantes transmises par le niveau régional, puis les transmettait à l'InVS. L'Institut fédératif de recherche 69 de l'Inserm (IFR69) jouait le rôle de tiers de confiance (mobilisable dans le cadre d'une alerte pour le retour au cas).

Un troisième niveau d'anonymisation irréversible était réalisé par l'InVS sur les données des trois sources grâce à une clé propre à l'InVS.

Les trois niveaux d'anonymisation avaient pour intérêt d'empêcher la levée d'anonymat des données lors de leur réception et pendant leur traitement, mais aussi d'assurer la protection de la base constituée. Ce projet avait reçu l'avis favorable de la commission nationale de l'informatique et des libertés (Cnil).

Test de retour au cas en cas d'alerte

Afin de tester la capacité du dispositif à retourner au cas en présence d'alerte de cas groupés de cancer, un tirage aléatoire simple sans remise de 100 fiches ACP de six des huit structures volontaires a été réalisé. Les deux structures qui n'ont pas participé au test mettaient en avant une contrainte de temps.

Les identifiants anonymes des fiches tirées ont été transmis dans un premier temps au tiers de confiance (IFR69) dans le but de retrouver, par les numéros de fiche, la structure ACP puis le patient dans la structure ACP.

Analyse des données

Une description des données de chaque source a été réalisée. Une recherche de cohérence entre

les trois sources concernant le sexe, l'âge et le lieu de résidence a été réalisée. Le taux d'appariement des données croisées des trois sources (nombre de cas communs aux bases croisées sur le nombre total de cas ACP) a également été calculé.

Résultats

Description des données

Un total de 423 fiches individuelles de cancer de la thyroïde transmises par les ACP des régions pilotes d'IdF (250) et du NPdC (173) a été validé par l'InVS. Le pourcentage des informations renseignées et exploitables variait de 53% à 100%. La variable la moins renseignée était le code postal de résidence pour lequel on ne disposait que de 23% d'informations en IdF. L'âge moyen des cas au diagnostic était de 50,7 ans chez les hommes et de 49,4 ans chez les femmes, qui représentaient plus de 75% des cas. La forme histologique la plus fréquente était la forme papillaire (82%).

En ce qui concerne les données PMSI, 2 947 cas de cancer de la thyroïde avec identifiants valides ont été repérés, soit 2 240 cas en IdF (76%) et 694 cas dans le NPdC (24%). La base ALD30 du Sniiram contenait 1 159 cas de cancer de la thyroïde dont 921 (79%) domiciliés en IdF et 238 (21%) dans le NPdC. Les ALD provenaient pour 93% du régime général de l'assurance maladie et pour 4% du Régime social des indépendants (RSI). Les données de la Mutualité sociale agricole (MSA) n'alimentaient pas la base Sniiram.

Données croisées

La figure 1 montre la répartition des effectifs provenant des différentes sources, ainsi que les effectifs de cas appariés. Sur 423 cas ACP, 343 ont été retrouvés dans le PMSI (81%), et 202 dans les ALD (48%), avec 191 cas retrouvés dans le PMSI et les ALD (45%).

Un total de 354 cas ACP ont été retrouvés dans le PMSI ou dans les ALD, soit un taux d'appariement entre la source ACP et les deux autres sources de 84%. Ce taux était de 89% dans le NPdC et de 80% en IdF (tableau 1).

On notait une cohérence parfaite sur le sexe et l'âge pour les données appariées (tableau 1). Cette cohérence était également parfaite sur le département de domiciliation dans le NPdC mais elle est non mesurable en IdF compte tenu du grand nombre de données manquantes pour cette variable dans les données ACP.

À l'issue du test de retour au cas, la totalité des dossiers transmis a été retrouvée.

Discussion

Les résultats de l'étude pilote ont montré que le croisement des données individuelles des trois sources à l'aide d'un numéro anonyme commun était techniquement réalisable. Le taux d'appariement des données entre les trois sources s'élevait à 84% et n'était pas identique dans les deux régions étudiées du fait probablement de différences de pratiques médicales. Le pourcentage de cohérence sur le sexe et le département de résidence entre les trois sources de données était parfait.

¹ Crisap : association de pathologistes ayant pour mission la mise en commun et l'analyse des données ACP au niveau régional.

² Foin (Fonction d'occultation d'informations nominatives) est une librairie d'anonymisation développée par la CnamTS, qui comprend un algorithme informatique associé à un code ou secret d'anonymisation.

³ Camel (Chiffrement assurance maladie élargie) : logiciel qui permet le contrôle, le cryptage et le transport des données.

Figure 1 Répartition des effectifs appariés selon les sources, Île-de-France et Nord-Pas-de-Calais, année 2004 (8 structures ACP, premières mises en ALD issues du régime général et du Régime social des indépendants ; patients avec un séjour PMSI codé en diagnostic principal, associé ou relié) / Figure 1 Distribution of matched data according to sources, Ile-de-France and Nord-Pas-de-Calais regions, 2004 (ACP data from 8 laboratories, first long-term illnesses implemented by the general social security scheme and the social security scheme for self employed workers, patients with a hospital stay coded as main, associated or related diagnosis)

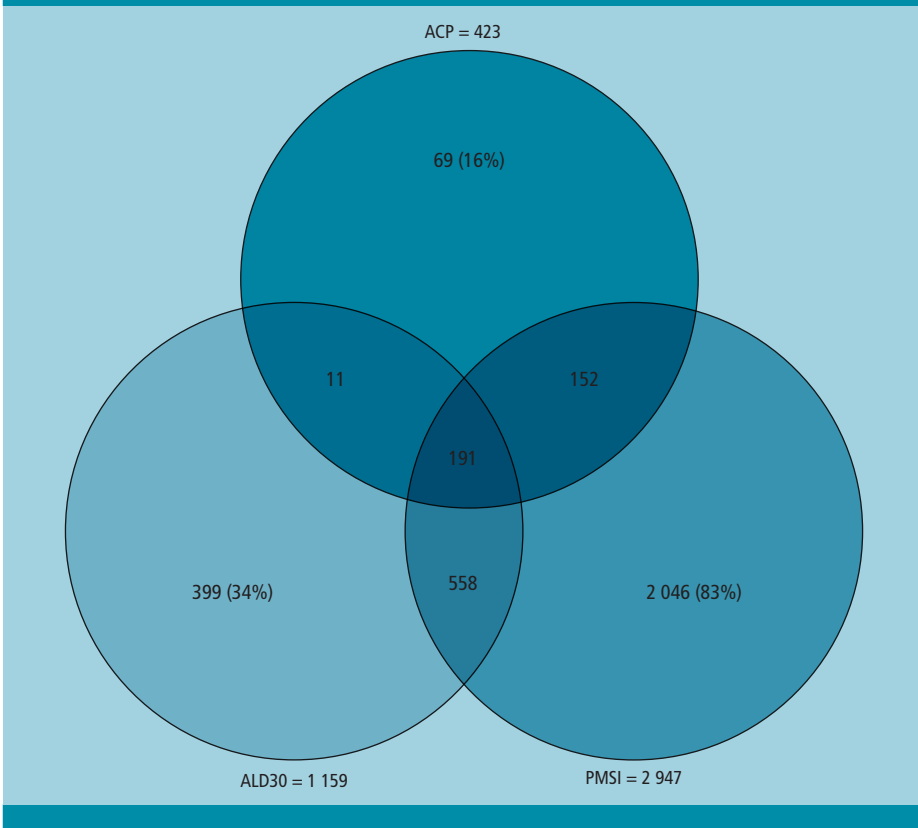


Tableau 1 Taux d'appariement entre la source ACP et les ALD et/ou le PMSI, et niveau de cohérence entre les données appariées / Table 1 Rate of matched data between ACP source and long-term illness data and/or hospital discharge data, and level of consistency between matched data

	ACP et ALD et PMSI (n=191)	ACP et ALD (n=202)	ACP et PMSI (n=343)	ACP et (ALD ou PMSI) (n=354)
Taux d'appariement^a				
Global	45%	48%	81%	84%
Région Nord-Pas-de-Calais	39%	40%	88%	89%
Région Île-de-France	49%	53%	76%	80%
Niveau de cohérence sur				
- Sexe	99%	100%	99%	
- Âge	100%	100%	100%	
- Département de résidence ^b	100%	100%	99%	

^a Dénominateur : nombre de cas ACP (n=423).

^b En raison du faible taux de remplissage d'information sur le département de résidence en Île-de-France, le niveau de cohérence sur le département de résidence n'a porté que sur les deux départements de la région du Nord-Pas-de-Calais.

L'absence de certains cas ACP dans les bases médico-administratives analysées pourrait s'expliquer par plusieurs facteurs :

- la mobilité des patients dont le lieu de résidence est proche d'une autre région administrative, avec le cas échéant une offre de soin spécifique à la maladie considérée comme plus performante. Cependant, à l'échelle nationale, l'effet de cette mobilité disparaîtrait ;
- les identifiants anonymes invalides de la base PMSI, qui représentaient 4% des séjours, n'ont pas été pris en compte dans les analyses ;
- les prises en charge hospitalières situées en fin d'année : par exemple, les cas diagnostiqués en fin de l'année n peuvent être comptabilisés dans le PMSI pour l'année $n+1$ (date de sortie d'hôpital) ;

- les résultats histologiques de la lésion traitée qui peuvent arriver après le codage du séjour hospitalier, empêchant son repérage par le PMSI parmi les données de cancer de la thyroïde ;
- la mise en ALD n'étant pas une obligation, certains assurés peuvent volontairement et pour différents motifs (déjà en ALD pour un autre motif, acte chirurgical unique sans suivi thérapeutique coûteux...) choisir de ne pas en bénéficier. Ils n'apparaissent donc pas dans la base ;
- la base ALD n'est pas complète et ne comprend notamment pas les données de la MSA ;
- la déclaration de la maladie peut se faire à une caisse d'assurance non domiciliée au lieu de résidence du patient. Ces cas échappent également à l'analyse.

La charge de travail des ACP était différente selon les structures. Elle variait de 2 à 155 fiches saisies, validées et transmises au niveau régional, avec parfois la nécessité d'une seconde saisie pour le même patient en cas d'informations complémentaires non disponibles au moment de la validation [11].

Le dispositif tel que testé est réalisable et opérationnel à une échelle limitée (deux régions, huit structures ACP) mais ne peut pas être déployé au niveau national, en raison notamment du niveau de complexité de la solution informatique utilisée, de l'hétérogénéité des systèmes d'information hospitalière et du manque de maîtrise des systèmes d'information des structures impliquées dans le recueil, l'anonymisation et la remontée des informations. Des dysfonctionnements liés au transfert des données étaient source de ruptures récurrentes sur toute la chaîne de transmission des informations.

De plus, les fortes contraintes techniques liées à l'utilisation des outils Foin (anonymisation) et Camel (sécurisation des données), qui n'autorisaient pas la technologie libre basée sur l'utilisation du web pour la transmission des données, ont compliqué l'intégration du système dans l'environnement informatique déjà hétérogène et complexe.

Par ailleurs, la constitution de la base ACP nécessiterait de pouvoir disposer d'outils standardisés et informatisés au sein des structures ACP (compte-rendu ACP), ce qui n'est pas encore le cas la plupart du temps.

Notons enfin que les résultats de cette étude (concordance entre sources) ne sont pas extrapolables aux autres localisations cancéreuses ou aux autres régions du fait des caractéristiques propres à chaque cancer (fréquence, modalités de prise en charge) et des variations de pratiques dans les départements [12].

Perspectives

Au regard de ces difficultés, deux axes de développement du SMSC au niveau national ont été identifiés :

- 1) le premier axe consiste à développer des travaux autour du croisement des deux sources médico-administratives (PMSI et ALD) déjà disponibles au niveau national à des fins descriptives, de production d'indicateurs de surveillance des cancers et de leurs comparaisons avec les données de registres dans les départements couverts ;
- 2) le deuxième axe, à moyen terme et en parallèle avec le premier, sera consacré à la constitution de la base ACP à partir du dossier communicant de cancérologie (DCC), dont l'objectif premier est d'améliorer la qualité de la prise en charge des patients [13].

Conclusion

L'étude pilote a montré que le SMSC était techniquement possible, avec en particulier la constitution *de novo* d'une base de données d'anatomie et de cytologie pathologiques et la création d'un identifiant anonyme commun aux trois sources. Toutefois, la solution informatique pour la collecte et la remontée des données ACP s'est avérée trop complexe pour un déploiement national. Il a donc été décidé d'utiliser les données ACP qui alimenteront le DCC.

Remerciements

À l'ensemble des partenaires du projet, en particulier les différents services concernés de la CnamTS ; les médecins pathologistes des structures ACP impliquées ; les Drass Île-de-France et Nord-Pas-de-Calais ; Sintel ; le Crisapif ; l'IFR69 et l'Inserm ; la Cnil ; les membres du comité de pilotage.

Références

- [1] Surveillance épidémiologique des cancers en France. Dossier thématique, site internet de l'Institut de veille sanitaire. Disponible à : <http://www.invs.sante.fr/Dossiers-thematiques/Maladies-chroniques-et-traumatismes/Cancers/Surveillance-epidemiologique-des-cancers>
- [2] Belot A, Grosclaude P, Bossard N, Jouglu E, Benhamou E, Delafosse P, *et al.* Cancer incidence and mortality in France over the period 1980-2005. *Rev Epidemiol Santé Publique* 2008;56(3):159-75.
- [3] Colonna M, Bossard N, Mitton N, Remontet L, Belot A, Delafosse P, *et al.* Some interpretation of regional estimates of the incidence of cancer in France over the period 1980-2005. *Rev Epidemiol Santé Publique* 2008;56(6):434-40.
- [4] Mitton N, Colonna M, Trombert B, Olive F, Gomez F, Iwaz J, *et al.* A suitable approach to estimate cancer incidence in area without cancer registry. *J Cancer Epidemiol.* 2011;418968:1-11.
- [5] Remontet L, Mitton N, Couris CM, Iwaz J, Gomez F, Olive F, *et al.* Is it possible to estimate the incidence of breast cancer from medico-administrative databases? *Eur J Epidemiol.* 2008;23(10):681-8.
- [6] Uhry Z, Colonna M, Remontet L, Grosclaude P, Carré N, Couris CM, *et al.* Estimating infra-national and national thyroid cancer incidence in France from cancer registries data and national hospital discharge database. *Eur J Epidemiol.* 2007;22(9):607-14.
- [7] Bilan du plan cancer 2003-2007 [Internet]. Institut national du cancer, France. Disponible à : <http://www.e-cancer.fr/institut-national-du-cancer/presentation/1993-plan-cancer-2003-2007>
- [8] Le plan cancer 2009-2013 [Internet]. Institut national du cancer, France. Disponible à : <http://www.e-cancer.fr/plancancer-2009-2013>
- [9] Leenhardt L, Grosclaude P, Chérié-Challine L, *et al.* Mise en place d'un dispositif de surveillance épidémiologique nationale des cancers thyroïdiens. Saint-Maurice: Institut de veille sanitaire; 2001. 146 p. Disponible à : <http://www.invs.sante.fr/Dossiers-thematiques/Maladies-chroniques-et-traumatismes/Cancers/Surveillance-epidemiologique-des-cancers/Publications>
- [10] Verger P, Chérié-Challine L. Évaluation des conséquences sanitaires de l'accident de Tchernobyl en France. Dispositif de surveillance épidémiologique, état des connaissances, évaluation des risques et perspectives. Saint-Maurice: Institut de veille sanitaire; 2001. 24 p. Disponible à : <http://www.invs.sante.fr/Dossiers-thematiques/Maladies-chroniques-et-traumatismes/Cancers/Surveillance-epidemiologique-des-cancers/Publications>
- [11] Caserio-Schönemann C, Kudjawu Y, Chérié-Challine L, Bloch J, Guillet A, Musset A, *et al.* Système multi sources de surveillance des cancers (SMSC) : Étude pilote sur le cancer de la thyroïde en Ile-de-France et dans le Nord-Pas-de-Calais, 2010. Saint-Maurice: Institut de veille sanitaire, France. Disponible à : <http://www.invs.sante.fr/Publications-et-outils/Rapports-et-syntheses/Maladies-chroniques-et-traumatismes/2011/Systeme-multi-sources-de-surveillance-des-cancers-SMSC>
- [12] Le choix des thérapeutiques du cancer du rectum. Conférence de consensus, Société nationale française de colo-proctologie, 1994. Disponible à : <http://www.med.univ-rennes1.fr/uv/snfcp/pratique>
- [13] Dossier communicant de cancérologie. Institut national du cancer, France [Internet]. Disponible à : <http://www.e-cancer.fr/soins/parcours-de-soins/dossier-communicant-de-cancerologie>

La publication d'un article dans le BEH n'empêche pas sa publication ailleurs. Les articles sont publiés sous la seule responsabilité de leur(s) auteur(s) et peuvent être reproduits sans copyright avec citation exacte de la source.

Retrouvez ce numéro ainsi que les archives du Bulletin épidémiologique hebdomadaire sur <http://www.invs.sante.fr/Publications-et-outils/BEH-Bulletin-epidemiologique-hebdomadaire>

Directrice de la publication : Dr Françoise Weber, directrice générale de l'InVS
Rédactrice en chef : Judith Benrekassa, InVS, redactionBEH@invs.sante.fr
Rédactrice en chef adjointe : Laetitia Gouffé-Benadiba
Secrétaires de rédaction : Farida Mihoub, Annie Cholin

Comité de rédaction : Dr Sabine Abitbol, médecin généraliste ; Dr Thierry Ancelle, Faculté de médecine Paris V ; Dr Pierre-Yves Bello, Direction générale de la santé ; Dr Juliette Bloch, CNSA ; Cécile Brouard, InVS ; Dr Sandrine Danet, Drees ; Dr Claire Fuhrman, InVS ; Dr Bertrand Gagnière, Cire Ouest ; Anabelle Gilg Soit Ilg, InVS ; Dorothée Grange, ORS Île-de-France ; Philippe Guilbert, Inpes ; Dr Rachel Haus-Cheymol, Service de santé des Armées ; Éric Jouglu, Inserm CépiDc ; Dr Nathalie Jourdan-Da Silva, InVS ; Agnès Lefranc, InVS ; Dr Bruno Morel, ARS Rhône-Alpes ; Dr Valérie Schwoebel, Cire Midi-Pyrénées ; Hélène Therre, InVS.