

## CROISEMENT DE DEUX BASES DE DONNÉES MÉDICO-ADMINISTRATIVES : MÉTHODOLOGIE ET ÉTUDE DESCRIPTIVE POUR UNE APPLICATION À LA SURVEILLANCE ÉPIDÉMIOLOGIQUE DES CANCERS EN FRANCE

// CROSSING TWO HEALTH ADMINISTRATIVE DATABASES: METHODOLOGY AND DESCRIPTIVE STUDY  
FOR APPLICATION TO EPIDEMIOLOGICAL SURVEILLANCE OF CANCER IN FRANCE

Yao Kudjawu<sup>1</sup> (y.kudjawu@invs.sante.fr), Florence de Maria<sup>1</sup>, Elsa Decool<sup>1</sup>, Michelle Altana<sup>2</sup>, Jean-Marc Harlin<sup>3</sup>,  
Alain Weill<sup>4</sup>, Javier Nicolau<sup>1</sup>, Christian Leroy<sup>1</sup>, Isabelle Grémy<sup>1</sup>

<sup>1</sup> Institut de veille sanitaire, Saint-Maurice, France

<sup>2</sup> Régime social des indépendants, Saint-Denis, France

<sup>3</sup> Mutualité sociale agricole, Paris, France

<sup>4</sup> Caisse nationale d'assurance maladie des travailleurs salariés, Paris, France

Soumis le 26.07.2013 // Date of submission: 07.26.2013

### Résumé // Abstract

**Introduction** – L'utilisation des données croisées du programme de médicalisation des systèmes d'information (PMSI) et d'affections longue durée (ALD) constitue, en France, l'une des alternatives pour estimer localement l'incidence des cancers dans les zones dépourvues de registre. L'objectif de cette étude est de décrire la méthodologie de sélection des séjours PMSI et des bénéficiaires d'ALD et leur croisement, puis de comparer les effectifs obtenus aux données d'incidence de la zone registre de cancer.

**Matériel et méthode** – La période d'étude couvrait 2006-2008. Les données de 21 localisations cancéreuses étaient extraites du PMSI national, des bases nationales de l'assurance maladie et des registres de cancers. La sélection des informations dans le PMSI et les ALD, croisées entre elles grâce à un identifiant anonyme individuel et commun, reposait sur des algorithmes basés sur les codes CIM-10 de diagnostic de cancer.

**Résultats** – Le pourcentage de patients PMSI appariés aux bénéficiaires d'ALD variait de 19% à 60% selon la localisation cancéreuse. L'effectif de patients hospitalisés ou admis en ALD, considérés comme atteints pour la première fois d'un cancer invasif et dénombrés dans la base croisée PMSI-ALD, était plus élevé que le nombre de cas incidents dans la zone registre.

**Conclusion** – Cette description est une étape nécessaire à l'utilisation de la base croisée PMSI-ALD pour l'estimation infranationale de l'incidence des cancers. L'estimation reposera sur l'application, aux zones géographiques considérées, du rapport « incidence cancer/indicateur issu du PMSI-ALD croisé » calculé sur la zone registre.

**Introduction** – The use of crossed data from the French Hospital Discharge Data system (PMSI) and long-term illness (LTI) databases is one of alternatives to estimate subnational incidence of cancer in areas without registries in France. The objective of this study is to describe the methodology of hospital stays and LTI selection and their crossing, and to compare figures obtained from the crossed data to cancer incidence in registry area.

**Material and method** – The study period covered 2006-2008. Data for 21 cancers sites were extracted from national PMSI, health insurance databases and common database of cancer registries. Hospital stays and LTI beneficiaries' selection was based on algorithms related to ICD10 diagnosis codes and were crossed through an anonymous individual and common identifier.

**Results** – The percentage of PMSI patients matched with LTI varied from 19% to 60%, according to cancer site. The number of hospitalized or admitted to LTI patients, considered as suffering from invasive cancer for the first time in the crossed database, was higher than the number of incident cancers in registry area.

**Conclusions** – This description is a necessary step to use PMSI-LTI crossed data for subnational incidence estimation of cancer. The estimation will be based on "cancer incidence/number of patients from PMSI-LTI crossed database" ratio in registry area applied to considered geographical areas.

---

**Mots-clés** : Cancer, PMSI, ALD, Épidémiologie, France

// **Keywords** : Cancer, Hospital discharge data, Long-term illness data, Epidemiology, France

---

## Introduction

En France, la surveillance épidémiologique des cancers repose principalement sur les données des registres des cancers. Ces structures, créées le plus souvent sur une base départementale, constituent les seules sources de données exhaustives et continues d'enregistrement des cas de cancer (selon la localisation, le type histologique, le stade au diagnostic...) sur leur zone géographique. La recherche active des cas et leur validation par les registres nécessitent des moyens humains et financiers importants représentant un obstacle majeur à la mise en place de registres dans la totalité des 101 départements français. Les registres français couvrent environ 20% de la population générale<sup>1,2</sup>. À partir de leurs données, des estimations nationales de l'incidence des cancers sont produites depuis plusieurs années<sup>3</sup>. Elles reposent sur l'extrapolation à la France du rapport « Incidence/Mortalité » observé dans la zone registre (*i.e.* ensemble des départements couverts par un registre). Si ce rapport peut être raisonnablement utilisé pour les estimations nationales, il ne peut l'être au niveau infranational, en raison de sa variabilité interdépartementale<sup>4</sup> empêchant toute estimation d'incidence dans les zones dépourvues de registres.

Les bases nationales de données médico-administratives que sont le Programme de médicalisation des systèmes d'information (PMSI) et les bases contenant les Affections longue durée (ALD) n'ont pas été créées pour mesurer l'incidence des cancers<sup>5,6</sup>. Cependant, depuis quelques années, les ALD sont utilisées pour suivre les tendances récentes de l'incidence nationale des cancers<sup>7</sup>. Des développements méthodologiques ont également été réalisés avec ces bases pour estimer l'incidence des cancers au niveau infranational. Ils reposent sur l'utilisation du rapport « Incidence/PMSI (ou ALD) de la zone registre » appliqué au PMSI (ou ALD) des départements (ou régions) sans registre<sup>1,8</sup>. Des estimations départementales (ou régionales) ont ainsi pu être fournies pour certains cancers mais pas pour tous, car l'hypothèse d'égalité du rapport « Incidence/PMSI (ou ALD) » entre les départements (ou régions) n'a pu être validée pour plus de la moitié des localisations cancéreuses<sup>8,9</sup>. Pour ces localisations, l'une des orientations repose sur la construction d'un nouvel indicateur, proxy de l'incidence, obtenu en croisant les bases PMSI et ALD, qui serait soumis à la même méthodologie pour produire des estimations infranationales. Cette démarche doit être précédée d'une étude exploratoire de l'indicateur construit à partir du croisement PMSI-ALD, afin de comprendre les phénomènes induits par le croisement d'informations médico-administratives de différente nature.

L'objectif de l'étude exploratoire présentée dans cet article est de :

1) décrire la méthodologie de :

- sélection des patients atteints pour la première fois d'un cancer invasif dans chacune des bases ;
- croisement des bases PMSI et ALD ;
- construction de l'indicateur proxy de l'incidence issu du croisement PMSI-ALD ;

2) comparer cet indicateur avec les cas incidents de cancers invasifs de la zone registre qui est la référence.

## Matériel et méthode

La période d'étude couvrait 2006-2008 et concernait 21 localisations cancéreuses (tableau) identifiées en utilisant la Classification internationale des maladies, 10<sup>e</sup> révision (CIM-10).

### Trois informations différentes et une information de référence

#### Les séjours hospitaliers issus du PMSI (période 2004-2008)

Le PMSI enregistre exhaustivement les hospitalisations des établissements (privés et publics) de court séjour. Chacune produit un résumé de sortie anonyme comportant des informations administratives (âge, sexe, code postal de résidence) et médicales : code CIM-10 de diagnostics principal (DP), relié (DR) et associé significatif (DAS) enregistré sur quatre caractères, actes réalisés. Pour mémoire, le DP représente (jusqu'en 2009) le « motif de prise en charge qui a mobilisé l'essentiel de l'effort médical et soignant au cours de l'hospitalisation ». Le DR « a pour rôle, en association avec le DP et lorsque celui-ci n'y suffit pas, de rendre compte de la prise en charge du malade en termes médico-économiques » : il s'agit de tout diagnostic permettant d'éclairer le contexte pathologique, essentiellement lorsque le DP n'est pas lui-même une affection. Pour le cancer, il est fréquemment complété avec un code de cancer invasif lors de la réalisation de séances de chimiothérapie ou de radiothérapie, elles-mêmes étant codées en DP. Les DAS sont notamment « les diagnostics, symptômes et autres motifs de recours significatifs d'une majoration de l'effort de soins et de l'utilisation des moyens, par rapport aux DP et DR » : il peut s'agir d'une complication du DP ou du couple DP+DR, d'une complication du traitement ou d'une affection distincte supplémentaire (source : <http://www.atih.sante.fr>).

Chaque année l'Institut de veille sanitaire (InVS), dans le cadre de ses missions de surveillance, reçoit de l'Agence technique de l'information sur l'hospitalisation (Atih) une extraction du PMSI national. Pour cette étude, cinq bases annuelles de séjours PMSI pour cancers de 2004 à 2008 ont été utilisées. La base PMSI de chaque année N regroupe tous les séjours hospitaliers de l'année N, en lien avec un cancer.

#### Les consommations de soins pour cancer reconnu comme ALD, issues du Sniiram (période 2006-2008)

Le Sniiram (Système national d'information inter-régimes de l'assurance maladie) est un ensemble de bases de données informationnelles des régimes d'assurance maladie. Il est constitué des données d'activité et de dépenses détaillées, individualisées et chaînées, couvrant les soins de ville, les soins hospitaliers sanitaires et sociaux, publics et privés.

Dans ces bases figure l'information d'ALD, un dispositif financier de l'assurance maladie de prise en charge totale des soins et traitements en lien avec une maladie « dont la gravité et/ou le caractère chronique nécessitent un traitement prolongé et une thérapeutique particulièrement coûteuse »<sup>10</sup> comme les cancers. Ces données d'ALD contiennent des informations médicales (code CIM-10 de la pathologie ayant entraîné la mise en ALD, enregistré sur trois caractères, dates de début et de fin d'ALD). Ces informations peuvent être chaînées avec des informations administratives (année et mois de naissance, sexe, code postal de résidence, régime d'affiliation) présentes dans le Sniiram. Trois extractions du Sniiram 2006-2008 relatives aux cancers, transmises par l'assurance maladie, ont été utilisées. La base Sniiram de l'année N regroupe toutes les consommations de soins de l'année N associées à une ALD cancer, que celle-ci ait été accordée l'année N ou les années précédentes (exemple : la base Sniiram 2006 contient des informations sur les patients mis en ALD en 2006 ou avant 2006 qui ont consommé des soins en lien avec le cancer en 2006).

#### Les assurés sociaux bénéficiant pour la première fois du dispositif ALD de la Mutualité sociale agricole (MSA) en raison d'un cancer (période 2006-2008)

En raison de leur absence du Sniiram, les données des assurés sociaux de la MSA ont spécifiquement été extraites des bases des bénéficiaires d'ALD pour cancer de la MSA (informations administratives et médicales). Tout comme le Sniiram, trois extractions de bases MSA, correspondant chacune à une année, de 2006 à 2008, ont été utilisées. La base MSA de l'année N regroupe tous les patients atteints d'un cancer reconnu comme ALD par la MSA pour la première fois l'année N.

#### La référence : les cas de cancer invasif de la zone registre (période 2006-2008)

L'indicateur proxy de l'incidence issu du croisement PMSI-ALD a été comparé à l'incidence de la zone registre. Les données des registres ont été extraites de la base commune des registres des cancers.

#### Étape préliminaire : un identifiant anonyme commun

Les données des bases utilisées ont bénéficié d'une anonymisation qui reposait sur le numéro d'inscription au répertoire, le sexe et la date de naissance complète, grâce à la fonction d'occultation d'informations nominatives de l'assurance maladie. Il en est résulté un identifiant anonyme unique par personne, commun aux bases médico-administratives.

#### Sélection des patients hospitalisés ou admis en ALD atteints pour la première fois d'un cancer invasif

La méthode de sélection des patients diffère entre les bases PMSI, Sniiram et MSA en raison de leur mode de constitution non identique (figure 1).

#### Sélection dans les bases PMSI

Dans la base PMSI de l'année N, les séjours avec un code CIM-10 de cancer invasif en DP ont été sélectionnés et chaînés entre eux (même code CIM-10 en DP) puis avec ceux des deux années précédentes (mêmes codes CIM-10 en DP ou DR ou DAS) afin d'exclure les cancers invasifs prévalents. Les patients avec un code CIM-10 de cancer invasif l'année N sans séjour hospitalier pour ce même cancer en N-1 ou N-2 ont été retenus, permettant de constituer trois bases d'étude : PMSI 2006, PMSI 2007 et PMSI 2008.

#### Sélection dans les bases Sniiram et MSA

Dans la base Sniiram de l'année N, les consommations de soins avec un code CIM-10 de cancer invasif au niveau de la pathologie associée à l'ALD, et avec une date de mise en ALD l'année N, ont été chaînées entre elles pour obtenir les consommateurs de soins avec un code CIM-10 de cancer invasif dont la date de mise en ALD correspondait à l'année du fichier Sniiram. Au final, trois bases d'étude ALD-Sniiram ont été constituées, de 2006 à 2008.

Dans la base MSA de l'année N, la sélection reposait sur les assurés avec un code CIM-10 de cancer invasif au niveau de la pathologie associée à l'ALD. Au final, trois bases d'étude ALD-MSA ont été constituées, de 2006 à 2008.

Pour chacune des trois années, les consommateurs ALD-Sniiram ont été fusionnés avec les assurés ALD-MSA. Les doublons ont été éliminés afin de tenir compte des éventuels transferts d'assurés entre régimes.

#### Croisement des bases PMSI et ALD-Sniiram/MSA et traitement des non appariés

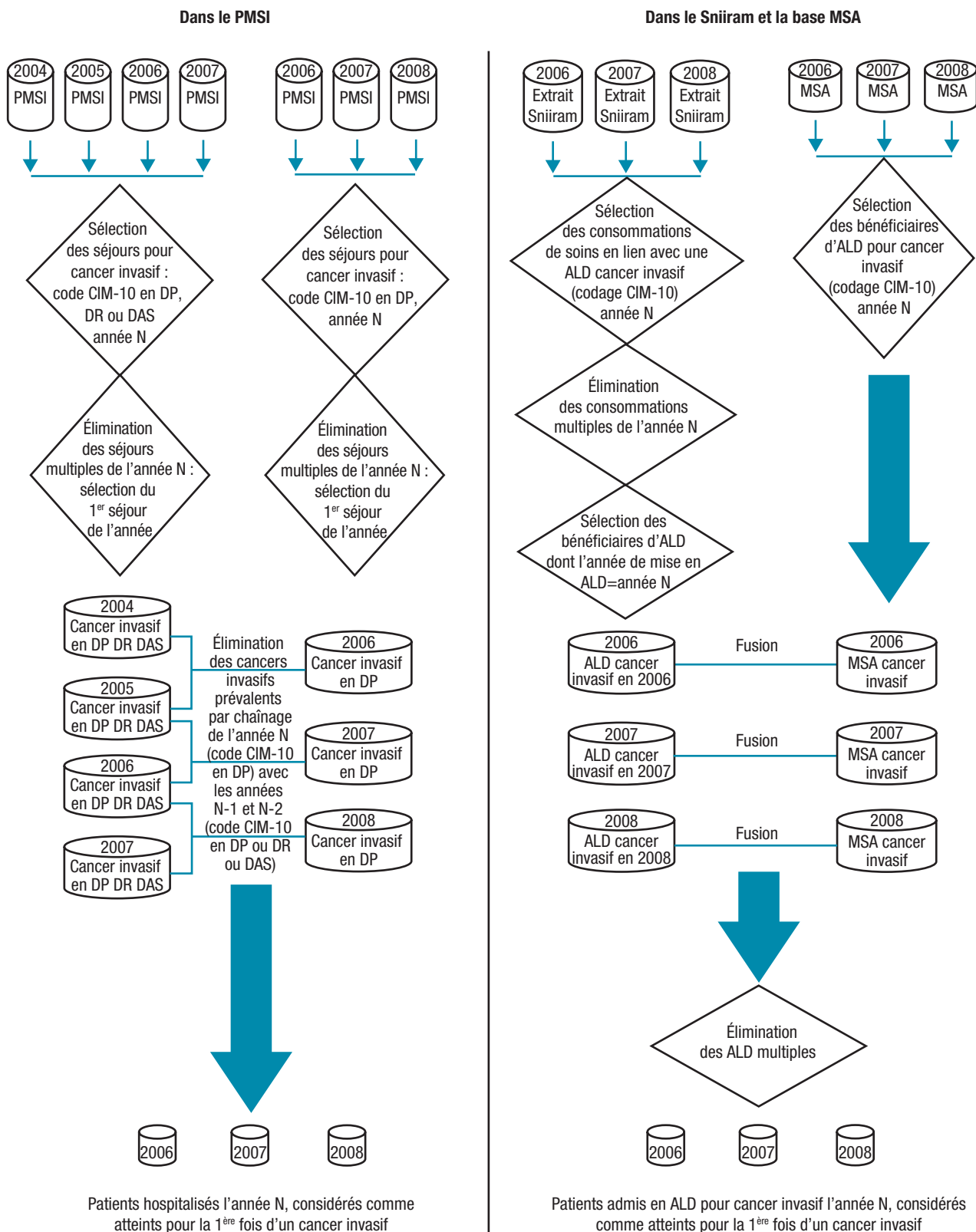
Les patients issus des bases d'étude PMSI et ALD-Sniiram/MSA (appelées dans la suite de l'article « bases ALD ») ont été croisés grâce à l'identifiant individuel anonyme commun et aux codes CIM-10 de cancer invasif (figure 2). À l'issue du croisement, trois groupes de patients ont été identifiés : patients PMSI-ALD appariés, patients PMSI non appariés aux bénéficiaires ALD et bénéficiaires ALD non appariés aux patients PMSI.

Pour chaque patient non apparié au sein d'une source, une vérification du caractère initial ou invasif du cancer a été réalisée à partir des informations contenues dans l'autre source. Elle consistait :

- pour le caractère initial du cancer d'un patient non apparié mis en ALD pour cancer, à rechercher parmi les bénéficiaires d'ALD non appariés de l'année N ceux qui avaient effectué un séjour hospitalier dans les deux années précédentes pour le même cancer, à l'aide des bases PMSI N-1 et PMSI N-2 ;
- pour le caractère initial du cancer d'un patient non apparié enregistré dans le PMSI, à rechercher parmi les patients PMSI non appariés de l'année N, ceux dont le cancer avait été reconnu comme ALD antérieurement à l'année N,

Figure 1

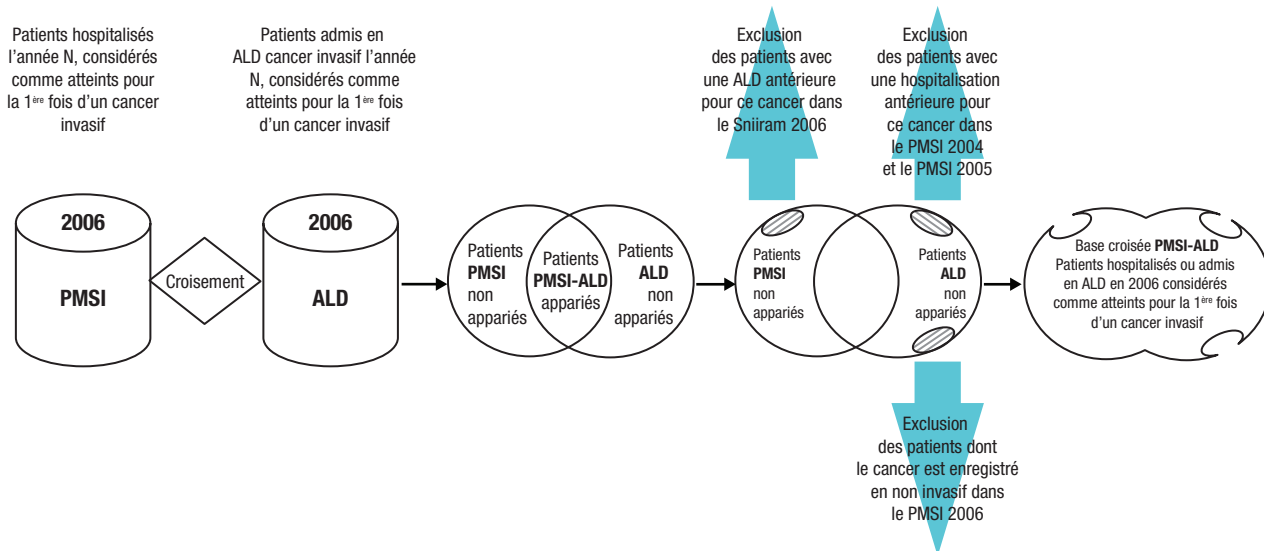
**Sélection des patients hospitalisés ou admis en ALD, atteints pour la première fois d'un type de cancer invasif**



ALD : Affection longue durée ; PMSI : Programme de médicalisation des systèmes d'information ; Sniiram : Système national d'information inter-régimes de l'assurance maladie ; MSA : Mutualité sociale agricole ; DP : diagnostic principal ; DR : diagnostic relié ; DAS : diagnostic associé, significatif.

Figure 2

**Croisement des bases PMSI et ALD (Sniiram/MSA) puis traitement des non appariés : exemple pour l'année 2006**



ALD : Affection longue durée ; PMSI : Programme de médicalisation des systèmes d'information ; Sniiram : Système national d'information inter-régimes de l'assurance maladie ; MSA : Mutualité sociale agricole.

enregistrés dans la base Sniiram de l'année N. Cette recherche ne pouvait pas être réalisée au sein des bases ALD-MSA en raison de leur indisponibilité pour les années antérieures à 2006 ;

- pour le caractère invasif du cancer d'un patient non apparié mis en ALD pour cancer, à rechercher parmi les bénéficiaires d'ALD non appariés de l'année N, ceux dont le diagnostic de cancer était identifié comme *in situ* (ou tumeur bénigne pour le système nerveux central) dans la base PMSI de l'année N (code CIM-10 en DP) ;
- pour le caractère invasif du cancer d'un patient non apparié enregistré dans le PMSI, la recherche n'a pas été réalisée en raison de la qualité du diagnostic final enregistré dans le PMSI (basé sur des informations histologiques ou hématologiques issues d'examen complémentaires), considérée *a priori* comme supérieure à celle accompagnant les mises en ALD.

Les patients PMSI ou ALD non appariés pour lesquels ces recherches étaient positives ont été exclus.

**Obtention de la population d'étude : les patients hospitalisés ou admis en ALD considérés comme atteints pour la première fois d'un cancer invasif**

Pour chaque cancer, le nouvel indicateur de surveillance était constitué de l'union :

- des patients PMSI appariés aux bénéficiaires ALD ;
- des patients PMSI non appariés aux bénéficiaires ALD, desquels ont été exclus ceux dont le cancer a été identifié comme prévalent grâce aux informations ALD ;
- des bénéficiaires ALD non appariés aux patients PMSI, desquels ont été exclus ceux dont le

cancer a été identifié comme prévalent ou non invasif grâce aux informations PMSI.

**Étude du nouvel indicateur : description et comparaison avec la référence**

Une recherche de cohérence entre les patients appariés, concernant le sexe, l'âge et le lieu de résidence, a été réalisée. Le taux d'appariement (nombre de patients appariés divisé par le nombre de patients issus de l'union PMSI-ALD) a été également calculé. Enfin, l'indicateur a été comparé aux cas incidents de cancers de la zone registre à travers le calcul d'un ratio (figure 3).

**Résultats**

Sur la période d'étude, une stabilité relative dans le temps des effectifs issus du croisement était observée pour chaque cancer. Le nombre de patients hospitalisés ou admis en ALD considérés comme atteints pour la première fois d'un cancer invasif variait selon le cancer (tableau) de 6 800 (lymphome de Hodgkin) à 187 500 (cancer de la prostate).

**Cohérence des informations entre les patients appariés**

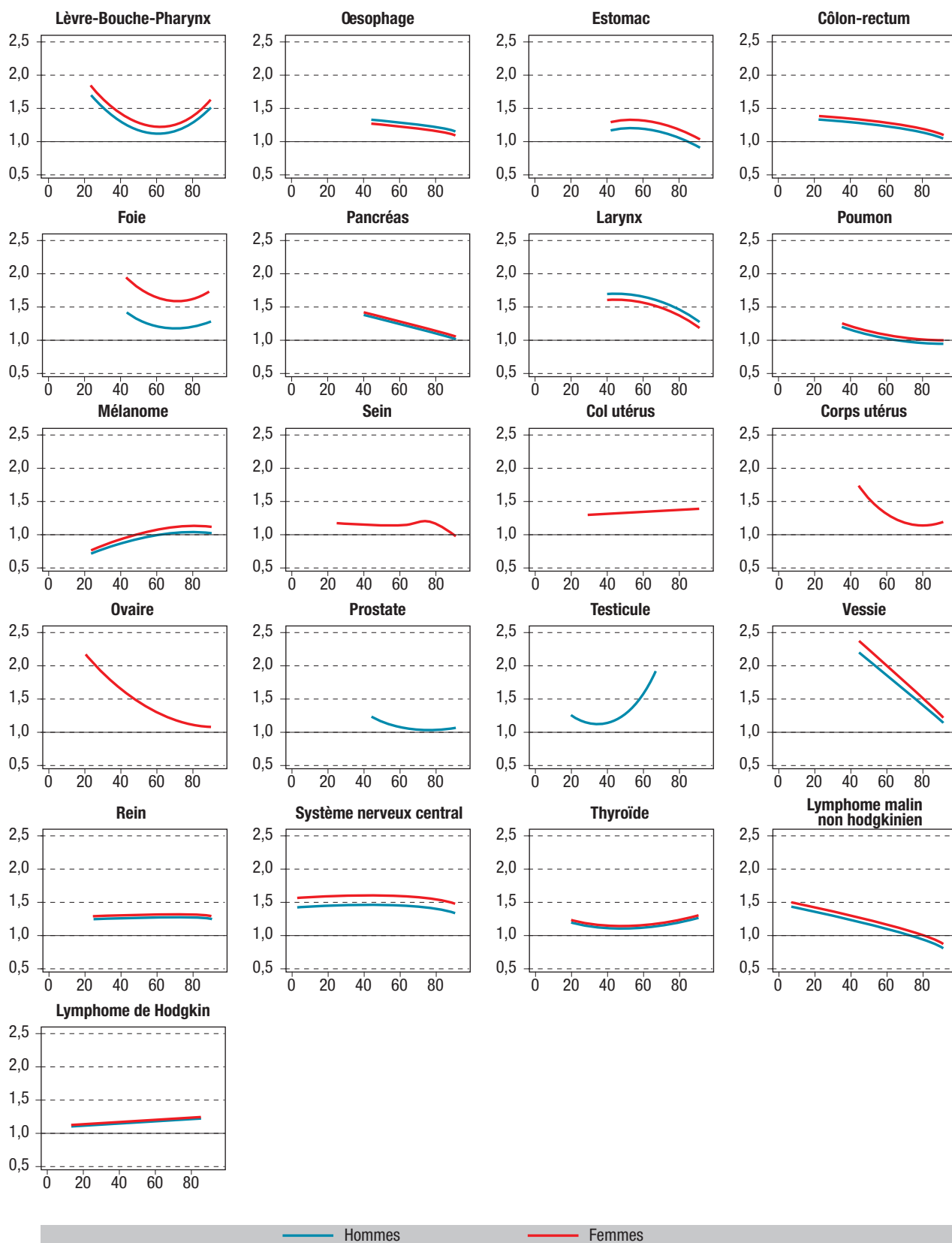
La comparaison individuelle des patients PMSI appariés aux bénéficiaires ALD montrait une cohérence parfaite sur le sexe et sur l'âge. La cohérence de l'information relative au département de résidence variait de 94% à 98% selon le cancer.

**Exclusion des cancers prévalents ou non invasifs, et taux d'appariement**

La majorité des patients exclus grâce au croisement étaient des patients hospitalisés mis en ALD avant leur hospitalisation. Le pourcentage

Figure 3

**Rapport entre le nombre de patients hospitalisés ou admis en ALD considérés comme atteints pour la première fois d'un cancer invasif dénombrés dans la base croisée PMSI-ALD, et le nombre de cas incidents, selon l'âge, par sexe, zone registre, 2006-2008**



L'axe des abscisses correspond à l'âge et l'axe des ordonnées au rapport entre le nombre de patients hospitalisés ou admis en ALD, considérés comme atteints pour la première fois d'un cancer invasif dénombrés dans la base croisée PMSI-ALD, et le nombre de cas incidents de cancer de la zone registre.

**Répartition des patients hospitalisés ou admis en ALD sélectionnés, considérés comme atteints pour la première fois d'un cancer invasif dans la base PMSI-ALD croisée, par localisation cancéreuse, effectifs cumulés 2006-2008**

Localisation cancéreuse	Code CIM-10	Apport du croisement : exclusion supplémentaire de faux-positifs				Patients hospitalisés ou admis en ALD considérés comme atteints pour la première fois d'un cancer invasif			
		Effectif et % de patients PMSI non appariés avec une ALD antérieure à l'hospitalisation (2)	Effectif et % de patients ALD non appariés avec une hospitalisation antérieure à l'ALD (3)	Effectif et % de patients appariés retrouvés parmi les patients PMSI atteints de tumeurs in situ ou bénignes (4)	Effectif « net », après exclusion des cancers non initiaux ou non invasifs (5) <sup>a</sup>	% de patients PMSI appariés aux patients ALD (taux d'appariement)			
Lèvre-Bouche-Pharynx	C00 à C14	5 241	9,7%	247	0,5%	346	0,6%	48 105	41%
Œsophage	C15	1 071	5,4%	102	0,5%	74	0,4%	18 452	41%
Estomac	C16	1 286	4,8%	131	0,5%	34	0,1%	25 419	35%
Côlon-Rectum-Anus	C18 à C21	8 186	5,4%	876	0,6%	516	0,3%	140 965	46%
Foie	C22	1 306	4,4%	262	0,9%	4	0,0%	27 877	19%
Pancréas	C25	1 346	4,0%	134	0,4%	7	0,0%	32 003	33%
Larynx	C32	1 200	6,6%	113	0,6%	182	0,0%	16 687	31%
Poumon	C33-C34	7 144	6,3%	488	0,4%	127	0,1%	105 170	39%
Mélanome de la peau	C43	1 585	5,0%	121	0,4%	398	1,3%	29 646	24%
Sein	C50	208 841	13,6%	323	0,2%	7 402	3,5%	172 735	59%
Col de l'utérus	C53	14 794	6,1%	30	0,2%	527	3,6%	13 334	32%
Corps de l'utérus et utérus sans autre indication	C54-C55	1 339	5,2%	136	0,5%	119	0,5%	24 294	42%
Ovaire	C56	1 070	5,5%	53	0,3%			18 195	38%
Prostate	C61	19 891	9,5%	1 624	0,8%	43	0,0%	187 547	44%
Testicule	C62	564	7,2%	8	0,1%	9	0,1%	7 222	60%
Vessie	C67	6 769	8,4%	2 107	2,6%	689	0,9%	71 016	31%
Rein	C64 à C66, C68	2 532	6,0%	194	0,5%	23	0,1%	39 229	36%
Système nerveux central	C70 à C72	1 324	5,7%	53	0,2%	1 151 <sup>b</sup>	5,0%	20 599	30%
Thyroïde	C73	2 697	9,6%	228	0,8%	59	0,2%	25 120	44%
Lymphome malin non hodgkinien	C82 à C85	3 346	7,9%	322	0,8%			38 771	33%
Lymphome de Hodgkin	C81	564	7,6%	16	0,2%			6 842	44%

<sup>a</sup> (5) = (1) - [(2) + (3) + (4)]

<sup>b</sup> Correspond à des tumeurs bénignes ou des kystes cérébraux du système nerveux central.

Les zones grisées correspondent à des localisations cancéreuses pour lesquelles il n'existe pas de code CIM-10 de cancer *in situ*.

ALD : Affection longue durée ; PMSI : Programme de médicalisation des systèmes d'information ; CIM-10 : Classification internationale des maladies, 10<sup>e</sup> révision.

de patients PMSI appariés aux bénéficiaires d'ALD variait selon la localisation cancéreuse. Le cancer du foie et le mélanome cutané présentaient un faible taux d'appariement (19% et 24%) et les cancers du sein et du testicule un taux élevé (59% et 60%). Ce taux variait de 30% à 46% pour les autres cancers (tableau).

### Comparaison sur la zone registre, des effectifs issus du croisement PMSI-ALD avec l'incidence

Pour tous les cancers, le nombre de patients hospitalisés ou admis en ALD considérés comme atteints pour la première fois d'un cancer invasif dénombrés dans la base croisée PMSI-ALD était plus élevé que le nombre de cas incidents de cancer invasif de la zone registre (figure 3). Le rapport PMSI-ALD/Incidence variait avec l'âge.

## Discussion

Les informations des bases médico-administratives ne sont pas recueillies à des fins épidémiologiques. Cependant, malgré leurs limites<sup>11</sup>, elles présentent des avantages qui leur permettent d'alimenter certains systèmes de surveillance comme celui des cancers en France. L'interprétation des résultats doit toutefois rester prudente.

La description de l'indicateur issu du croisement montre un taux d'appariement variable de 19% à 60% selon le cancer et un indicateur supérieur à l'incidence de la zone registre, suggérant la persistance de cancers prévalents ou non invasifs parmi les patients non appariés, malgré les exclusions déjà réalisées. En effet, si un appariement faible était attendu pour quelques cancers comme ceux nécessitant un geste chirurgical sans traitement long et coûteux ou ceux essentiellement traités en ambulatoire (mélanome de la peau) ou encore ceux à pronostic sombre à court terme ou siège de tumeur secondaire (foie et système nerveux central), il aurait dû être élevé pour les autres, ce qui n'est globalement pas observé.

Compte tenu de cette limite, la discussion sur l'appariement obtenu ne peut porter que sur ses valeurs extrêmes - élevées pour le testicule et le sein, faibles pour le foie et le mélanome cutané - et vise à confirmer des hypothèses sur le parcours et la prise en charge des patients.

Le cancer du testicule est un cancer rare, qui touche principalement les adultes jeunes entre 20 et 35 ans<sup>12</sup>. Les jeunes atteints sont non seulement des primo-accédants aux ALD, mais également des patients nécessairement pris en charge en secteur hospitalier, compte tenu du mode radical de traitement de ce cancer.

Pour le cancer du sein, la nécessité d'une demande de mise en ALD liée à une prise en charge coûteuse et de longue durée, après un diagnostic de plus en plus précoce grâce au programme de dépistage organisé, expliquerait le taux d'appariement élevé.

Le pronostic du cancer du foie est très mauvais. La survie brute à un an est de 36% et de 16% à 3 ans (période 1989-2007)<sup>13</sup>. Ce mauvais pronostic pourrait être à l'origine de peu de demandes spécifiques d'ALD pour ce cancer au profit de l'utilisation de l'ALD « maladies chroniques actives du foie et cirrhose » le plus souvent à l'origine de ce cancer.

Quant au mélanome cutané, son traitement chirurgical peut être réalisé en ambulatoire ou dans un cabinet privé (où il n'existe pas de PMSI). Par ailleurs, le traitement post-chirurgical peut ne pas s'avérer coûteux ou ne pas faire l'objet d'un long suivi.

Peu d'études françaises se sont intéressées à la qualité du PMSI ou des ALD comme indicateur de suivi épidémiologique des cancers à l'aide de données nominatives<sup>6,14</sup>. Outre la persistance de cancers prévalents difficilement repérables, ces études évoquent aussi, pour expliquer la persistance de faux positifs, des erreurs de codage de certaines lésions tumorales en raison :

- de leur localisation à la frontière d'organes (œsophage-estomac, cancers ORL...);
- d'un cancer métastatique considéré comme primitif (foie, poumon, système nerveux central); par exemple, dans notre étude, le rapport élevé PMSI-ALD/Incidence pour le cancer du foie chez la femme reflète probablement le codage de métastases hépatiques de cancers gynécologiques primitifs;
- de tumeurs bénignes du système nerveux central ou de tumeurs à stade inclassable de la vessie, codées en tumeurs malignes en raison d'une prise en charge hospitalière identique à celle des cancers invasifs, ce qui, dans notre étude, pourrait expliquer en partie le faible appariement observé (environ 30%) pour ces localisations.

Par ailleurs, le passage de la situation d'ayant-droit à celle d'assuré social serait également à l'origine de faux positifs. Le statut d'ayant-droit permet à une personne non reconnue comme assuré social par l'assurance maladie (enfant, conjoint qui n'a jamais travaillé...) de bénéficier d'une protection sociale. Dans les bases de données médico-administratives, les ayants-droits n'ont pas de numéro spécifique et sont enregistrés sous le numéro de l'assuré social dont ils dépendent. D'après une étude de l'assurance maladie, cette situation est très fréquente avant 25 ans et marginale ensuite (de l'ordre de 0,5% à 2%)<sup>15</sup>. Son impact serait limité sur les tumeurs solides, qui surviennent principalement chez des personnes âgées.

Une des possibilités pour exclure davantage de faux positifs consisterait à rendre plus spécifique les algorithmes de sélection. Dans la littérature, l'utilisation épidémiologique du PMSI repose sur la sélection des séjours à partir des codes diagnostiques de la pathologie étudiée, associés ou non aux actes thérapeutiques. Ainsi, pour les tumeurs



du système nerveux central, la sélection en « DP ou DR » permettrait d'observer une diminution des faux positifs avec une hausse de la valeur prédictive positive, mais au prix d'une diminution de la sensibilité<sup>14</sup>. Dans notre étude, la sélection PMSI reposait sur le DP afin d'exclure d'emblée les cancers prévalents (patient en cours de suivi médical, patient dont le cancer est connu et qui est hospitalisé pour une autre pathologie, récurrence) et de diminuer, pour certains cancers, le risque de variabilité géographique du rapport Incidence/PMSI par rapport à une sélection incluant les actes chirurgicaux<sup>16</sup>. La sélection PMSI associant « DP ou DR » a été testée : la proportion de faux positifs exclus avec une sélection en « DP » variait, selon le cancer, de 4,4% à 17,3% contre 5,4% à 20,4% avec une sélection en « DP ou DR » sans gain sur le taux d'appariement, sauf pour le mélanome cutané (+1 point) et le cancer du poumon (+3 points).

Autre option envisageable au niveau de la sélection PMSI : l'allongement de la durée de chaînage, fixée à deux ans dans l'étude, afin d'exclure également d'emblée les cancers prévalents. Les travaux relatifs aux estimations d'incidence à partir du PMSI retrouvés dans la littérature<sup>8,9</sup> ne comportant pas de chaînage entre les années, il était difficile de se prononcer sur une durée de chaînage pertinente. Toutefois, l'extension du chaînage à quatre années antérieures, réalisée pour l'année 2008, permettait de montrer qu'au moins 90% des cancers prévalents étaient repérés dans les deux premières années de chaînage, sauf pour les cancers du sein, de la prostate et de la vessie, pour lesquels une année supplémentaire de chaînage était nécessaire pour atteindre ce résultat.

Dans le Sniiram, la sélection des bénéficiaires d'ALD reposait sur les patients mis en ALD une année donnée et ayant réalisé leur première consommation de soins en lien avec l'ALD la même année. Ce processus ne permettait pas de prendre en compte les 4% à 7% de bénéficiaires ayant débuté leur consommation de soins postérieurement à l'année de leur mise en ALD. Toutefois, la sélection élargie ne permettait ni d'exclure davantage de faux positifs ni d'augmenter le taux d'appariement.

## Conclusion et perspectives

La méconnaissance de l'ampleur et du devenir des faux positifs issus des deux sources croisées semble poser davantage de questions qu'en résoudre, en termes d'impact sur les faux positifs : communs aux deux sources ou propres à une seule source ? En l'absence de littérature relative au croisement de bases médico-administratives, l'évaluation de la qualité de l'indicateur PMSI-ALD croisé pour le suivi épidémiologique des cancers devrait être envisagée à l'aide de données nominatives issues des registres. Ces travaux complèteraient la perspective prochaine de production d'estimations infranationales de l'incidence à partir de l'indicateur PMSI-ALD, sous réserve de sa validation. ■

## Remerciements

Les auteurs remercient vivement Zoé Uhry, qui a réalisé la comparaison de l'indicateur avec la zone registre, Laurent Duchet, David Dias, Sophie Gosselin, Mathilde Risse-Fleury et Pascale Grosclaude pour leur aide précieuse, ainsi que les relecteurs pour leurs remarques pertinentes.

## Références

- [1] 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. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=10441](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10441)
- [2] Grémy I. Portrait des registres des cancers en France. *Environ Risques Santé.* 2012;11(2):162-5. [http://www.jle.com/fr/revues/sante\\_pub/ers/e-docs/00/04/76/3E/article.phtml](http://www.jle.com/fr/revues/sante_pub/ers/e-docs/00/04/76/3E/article.phtml)
- [3] Binder-Foucard F, Belot A, Delafosse P, Remontet L, Woronoff A, Bossard N. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2012. Étude à partir des registres des cancers du réseau Francim - Partie 1 : tumeurs solides. Saint-Maurice: Institut de veille sanitaire; 2013. 122 p. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=11619](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=11619)
- [4] Colonna M, 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.
- [5] Couris C, 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. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=10438](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10438)
- [7] Uhry Z, Remontet L, Grosclaude P, Belot A, Colonna M, 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. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=10437](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10437)
- [8] Uhry Z, Remontet L, Colonna M, Belot A, Grosclaude P, Mitton N, *et al.* Cancer incidence estimation at a district level without a national registry: a validation study for 24 cancers sites using French health insurance and registry data. *Cancer Epidemiol.* 2013;37(2):99-114.
- [9] Estimations régionales de l'incidence par cancer, 2008-2010. In: Surveillance épidémiologique des cancers. Dossier thématique. Saint-Maurice: Institut de veille sanitaire (site Internet). <http://www.invs.sante.fr/Dossiers-thematiques/Maladies-chroniques-et-traumatismes/Cancers/Surveillance-epidemiologique-des-cancers/Estimations-de-l-incidence-et-de-la-mortalite/Estimations-regionales-de-l-incidence-par-cancer-2008-2010>.
- [10] Définition d'une affection de longue durée exonérante et d'une affection de longue durée non exonérante. Assurance maladie. <http://www.ameli.fr/assures/soins-et-remboursements/combien-serez-vous-rembourse/en-cas-d-affection-de-longue-duree/qu-est-ce-qu-une-affection-de-longue-duree-ald.php>.
- [11] Kudjawa Y, Rudant J, Decool E, Danzon A, Grémy I, Musset A, *et al.* Système multi sources de surveillance des cancers. Croisement des données ALD et PMSI. Description des données 2006-2008 et construction d'indicateurs. Saint-Maurice: Institut de veille sanitaire; 2012. 70 p. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=10637](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10637).

[12] Kudjawu Y, Danzon A, Bloch J. Évolution nationale et variations régionales du taux de patients opérés pour cancer du testicule en France, 1998-2008. *Bull Epidémiol Hebd.* 2012;(7-8-9):106-10. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=10472](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10472).

[13] Grosclaude P, Remontet L, Belot A, Danzon A, Rasamimanana-Cerf N, Bossard N. Survie des personnes atteintes de cancer en France 1989-2007. Étude à partir des registres des cancers du réseau Francim. Saint-Maurice: Institut de veille sanitaire; 2013. 412 p. [http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=11316](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=11316).

[14] Coureau G, Baldi I, Savès M, Jaffré A, Barat C, Gruber A, *et al.* Évaluation des performances du PMSI pour l'identification des tumeurs incidentes du système nerveux central par rapport à un registre spécialisé en Gironde, France, en 2004. *Rev Epidémiol Santé Publique.* 2012;60(4):295-304.

[15] Chaignot C, Weill A, Blotière PO, de Roquefeuil L, Ricordeau P, Alla F, *et al.* Intérêt de l'identifiant bénéficiaire anonyme unique vie entière dans le Sniiram et de son utilisation après chaînage dans le PMSI. *Rev Epidémiol Santé Publique.* 2013;61S:S16-S17.

[16] 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. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3065037/>.

#### Citer cet article

Kudjawu Y, de Maria F, Decool E, Altana M, Harlin JM, Weill A, *et al.* Croisement de deux bases de données médico-administratives : méthodologie et étude descriptive pour une application à la surveillance épidémiologique des cancers en France. *Bull Epidémiol Hebd.* 2013; (Hors-série):49-58.