

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.anccr.nu>

[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.

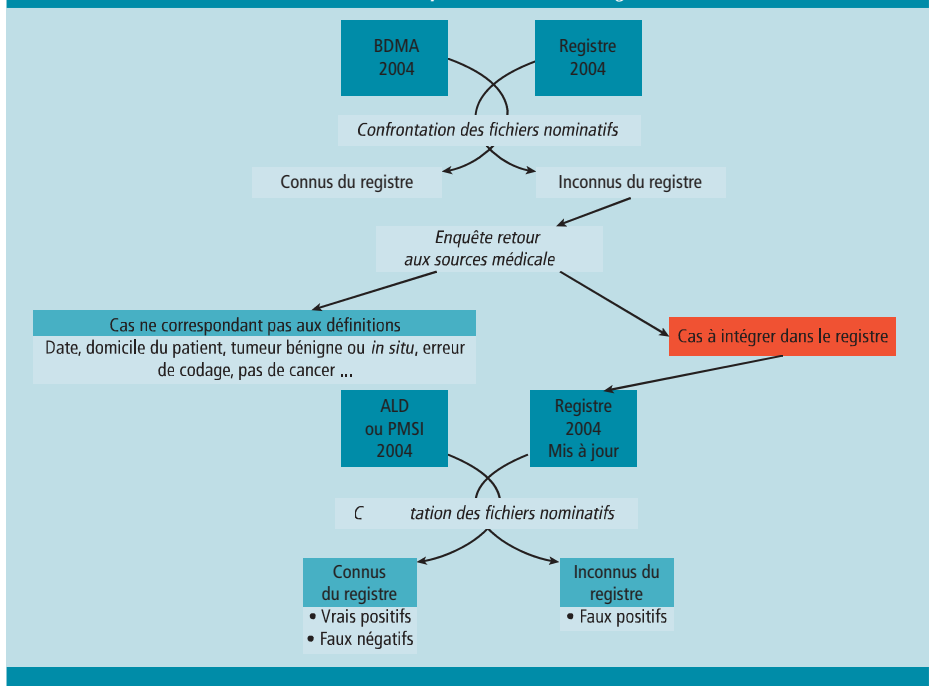
Pour le PMSI, le registre a demandé aux Départements d'information médicale (DIM) des établissements de soins les RSS de tous les séjours ayant un diagnostic principal et/ou un diagnostic relié avec un code CIM10 de cancer durant l'année 2004, ainsi que pour les années 2003 et 2005 (même raison que ci-dessus) 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 de 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
Mésotéliome (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'avions 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.