

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

4/ Département d'information médicale, Centre hospitalier universitaire de 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

* 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	France métropolitaine		38,0	24,8	98,6
Lozère	48	39,8	20,9	98,1					

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 *i* 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 Epidémiol 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 Epidémiol 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 Epidémiol 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 Epidémiol 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 Epidémiol 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 Epidémiol Santé Publique. 2011;59(1):53-8. Erratum in: Rev Epidémiol 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 Epidémiol 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 Epidémiol 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.