

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

Références

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

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

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

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

3/ Réseau français des registres de cancer Francim, France

5/ Registre des cancers du Tarn, Toulouse, France

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

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

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

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

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

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

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

Résumé / Abstract

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

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

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

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

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

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

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

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

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

Mots clés / Key words

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

Introduction

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

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

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

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

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

Matériel et méthode

Données

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

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

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

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

Méthode

Principe

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

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

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

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

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

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

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

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

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

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

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

Sélection des localisations cancéreuses pour une estimation départementale

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

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

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

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

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

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

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

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

Résultats

Analyses statistiques

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

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

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

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

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

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

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

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

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

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

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

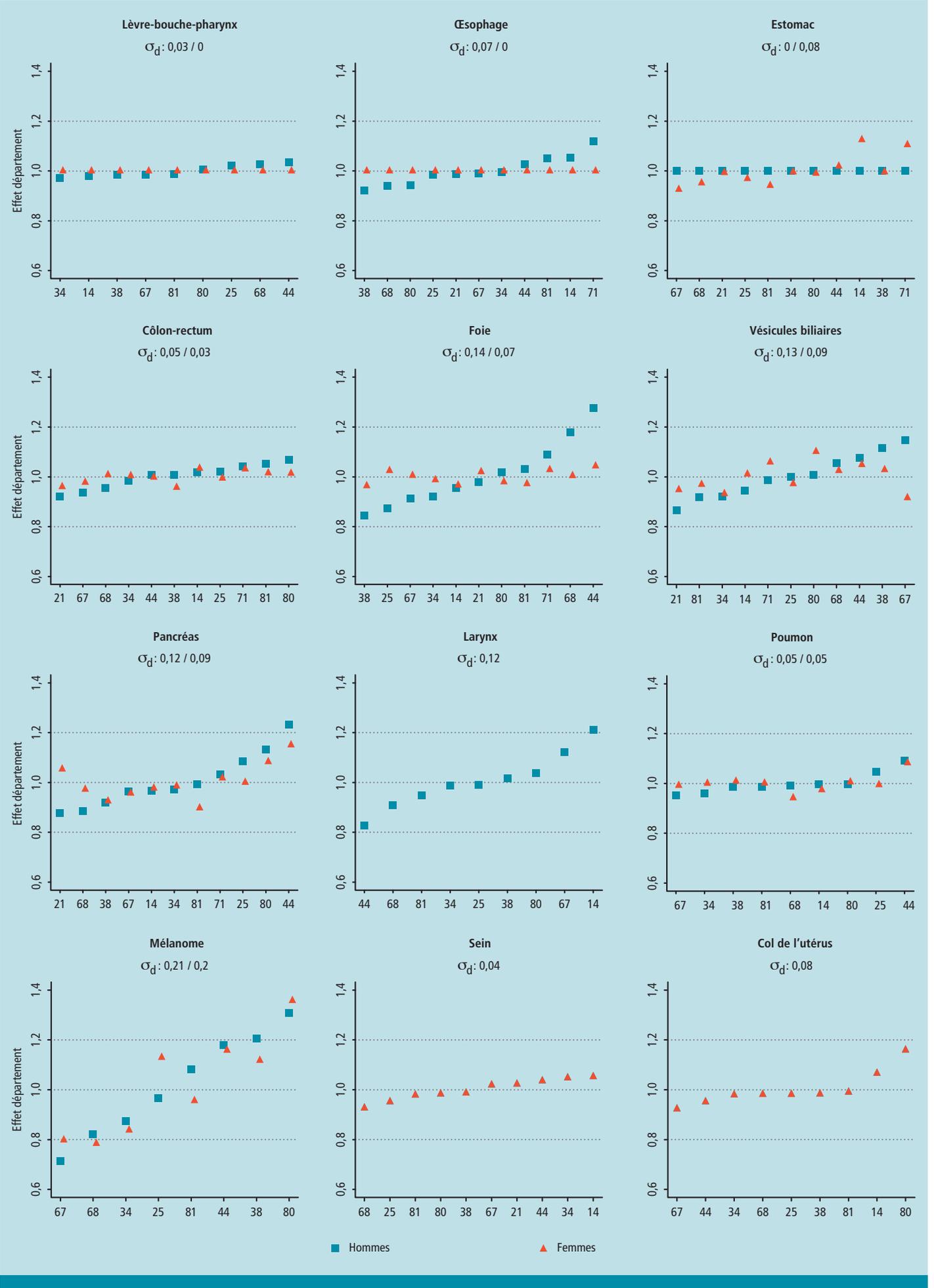
^a Validation croisée : nombre de départements avec une mesure d'écart K significativement trop élevé, avec $K=(O-P)^2/P$.

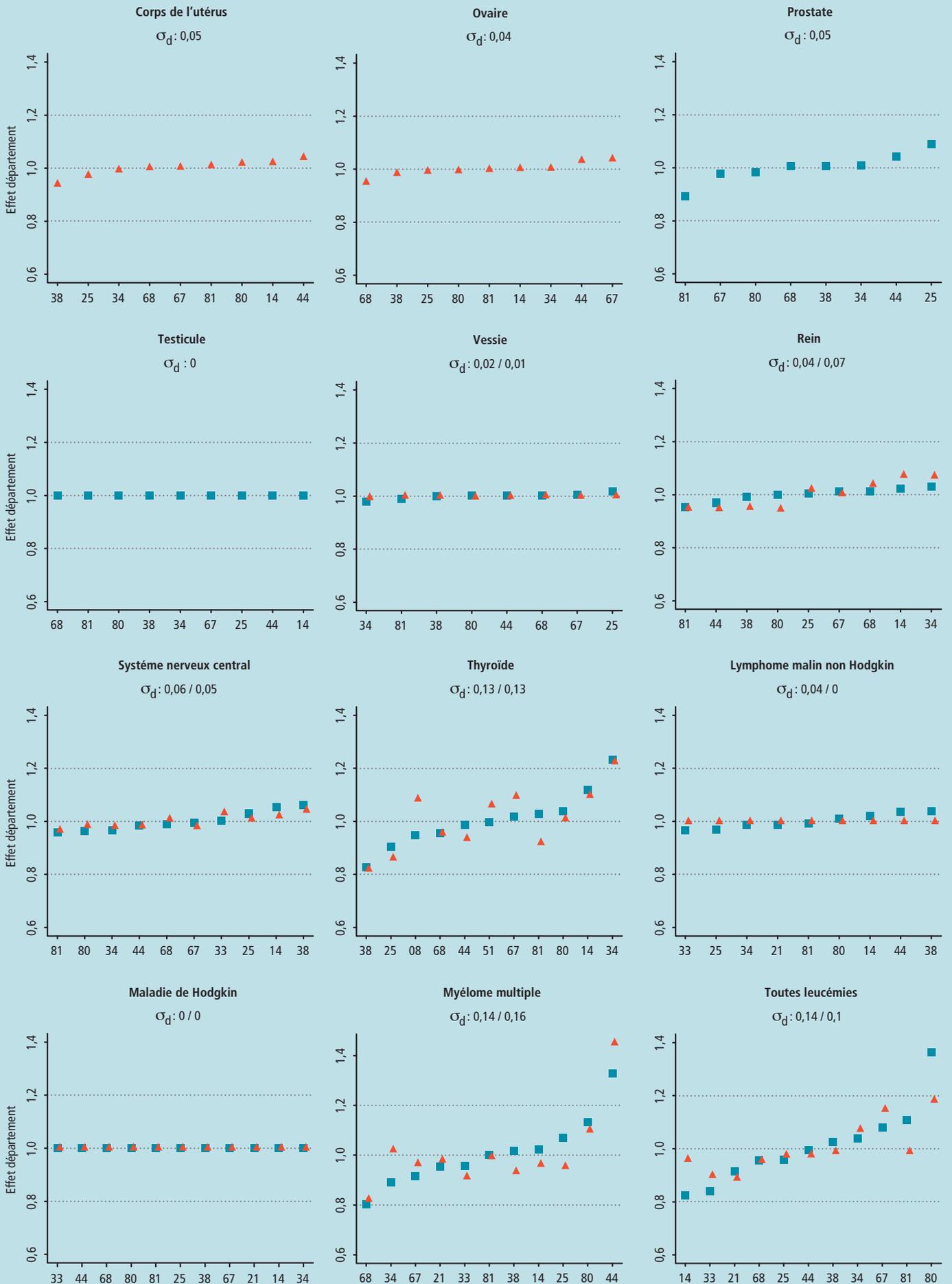
^b Validation croisée : erreur relative ER maximale (en valeur absolue) observée, avec $ER=(P-O)/O$.

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

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

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





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

Variabilité géographique majeure du rapport ALD/incidence

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

Variabilité géographique intermédiaire du rapport ALD/incidence

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

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

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

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

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

Variabilité géographique mineure du rapport ALD/incidence

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

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

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

Discussion

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

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

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

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

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

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

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

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

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

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

Références

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

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

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

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

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

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

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

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

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

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

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

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

1/ Hospices civils de Lyon, Service de Biostatistique, Lyon, France ; Université de Lyon ; Université Lyon I, Villeurbanne, France ; CNRS ; UMR 5558, Laboratoire Biostatistique Santé, Pierre-Bénite, France

2/ Centre Léon Bérard, Département d'information médicale, Lyon, France

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