

MÉTHODE D'ESTIMATION DE L'INCIDENCE RÉGIONALE DES CANCERS À PARTIR DES DONNÉES D'INCIDENCE DES REGISTRES, DES DONNÉES DE MORTALITÉ PAR CANCER ET DES BASES DE DONNÉES MÉDICO-ADMINISTRATIVES

// A METHOD TO ESTIMATE CANCER INCIDENCE AT THE ADMINISTRATIVE DISTRICT LEVEL IN FRANCE USING INCIDENCE DATA FROM CANCER REGISTRIES, MORTALITY DATA BY CANCER SITE AND HEALTH ADMINISTRATIVE DATABASES

Marc Colonna^{1,2} (mcolonna.registre@wanadoo.fr), Nicolas Mitton¹, Laurent Remontet³, Aurélien Belot³, Nadine Bossard³, Pascale Grosclaude^{2,4}, Zoé Uhry⁵

¹ Registre des cancers de l'Isère, Grenoble, France

² Réseau français des registres de cancer Francim, France

³ Service de biostatistique, Hospices civils de Lyon, France

⁴ Registre des cancers du Tarn, Institut Claudius Regaud (ICR), Toulouse, France

⁵ Institut de veille sanitaire, Saint-Maurice, France

Soumis le 02.07.2013 // Date of submission: 07.02.2013

Résumé // Abstract

Objectifs – En France, les registres des cancers couvrent seulement 20% de la population. Afin d'estimer l'incidence (I) régionale des cancers, différents corrélats peuvent être utilisés : Mortalité (M), Programme de médicalisation des systèmes d'information (PMSI) et Affection de longue durée (ALD). L'objectif de ce travail était d'évaluer la validité de ces corrélats et de les hiérarchiser pour fournir des estimations régionales d'incidence sur la période 2008-2010.

Méthode – Le principe de l'estimation est identique pour les trois méthodes et consiste à appliquer le rapport M/I de la zone registre (ou ALD/I ou PMSI/I, respectivement) aux données régionales de mortalité (ou ALD ou PMSI, respectivement). Du fait de la couverture départementale des registres en France, ces méthodes ne peuvent être évaluées, et donc validées, qu'à cet échelon administratif dans les départements couverts par un registre. Un arbre de décision a été adopté pour classer chaque méthode par cancer, puis les hiérarchiser. Les localisations éligibles (une méthode au moins jugée valide) ont été distinguées. Des fourchettes d'estimations ont toutefois été fournies pour les localisations non éligibles.

Résultats – Chez les hommes, 8 localisations sur 19 étaient éligibles, dont 6 à partir des ALD et 2 du PMSI. Chez les femmes, 9 localisations sur 21 étaient éligibles, dont 3 à partir des ALD, 5 du PMSI et 1 de la mortalité.

Conclusion – Ce travail présente la première démarche structurée évaluant différentes approches d'estimation de l'incidence régionale. Il confirme l'apport des bases médico-administratives pour cette problématique.

Objective – Cancer registries in France cover only 20% of the population. Various correlates may be used to estimate the geographical regional incidence (I) of various cancers: mortality (M), hospital administrative data (PMSI, Programme de Médicalisation des Systèmes d'Information), or long-term illness health insurance data (ALD, Affection de Longue Durée). The present study aimed to evaluate the validity of these correlates and prioritize them in terms of their ability to provide regional estimations of cancer incidence over the period 2008-2010.

Method – The estimation principle was identical for all three correlates. It consisted in using the correlate (M, or PMSI or ALD, respectively) / I ratio of the registries area together with the corresponding observed M, or PMSI, or ALD regional data. As cancer registries only cover certain administrative districts (termed "Départements"), these methods cannot be evaluated at the administrative "Région" level but only at the "Département" level. The validity of each method has been thus evaluated in each "Département" with cancer registries. A decision tree was used to prioritize the accuracy of each method by cancer site. Eligible cancer sites (with at least one valid method) were determined. Estimate ranges were provided for non-eligible sites.

Results – In men, 8 out of 19 sites were found to be eligible, 6 of which using ALD data and 2 using PMSI data. In women, 9 out of 21 sites were found to be eligible, 3 of which using ALD data, 5 using PMSI data, and 1 using mortality data.

Conclusion – This study presents the first structured approach evaluating various incidence estimation methods at the regional scale. It confirms the value of using health administrative data to solve this problem.

Mots-clés : Cancer, Estimation régionale, Incidence, Registre, Affection de longue durée, PMSI, Mortalité

// **Keywords**: Cancer, Regional estimation, Incidence, Registry, Long-duration disease data, PMSI, Mortality

Introduction

L'incidence des cancers est un indicateur épidémiologique indispensable pour appréhender le poids de cette pathologie. Étant donné l'importance de l'échelon régional dans la mise en œuvre de politiques de santé, la connaissance de cette incidence au niveau régional est essentielle.

En France, l'incidence est mesurée par des registres départementaux des cancers qui couvrent environ 20% de la population métropolitaine adulte à l'heure actuelle. L'incidence nationale des cancers est donc estimée en multipliant la mortalité nationale par cancer par le rapport Incidence /Mortalité (I/M) observé sur l'ensemble des départements couverts par un registre (zone registre)¹. Cette méthode est adaptée au niveau national², mais présente des limites au niveau régional³. Elle suppose, en effet, un rapport I/M identique dans toutes les régions, autrement dit, qu'une même incidence produit une même mortalité d'une région à l'autre. Plusieurs facteurs peuvent mettre à mal cette hypothèse³, en particulier des disparités régionales potentielles de survie. En l'absence d'une alternative aboutie à l'époque de ces études, des estimations régionales ont néanmoins été réalisées avec cette méthode, accompagnée d'un avertissement³.

Depuis, différents travaux ont été menés afin d'évaluer l'intérêt d'utiliser des données médico-administratives comme corrélat de l'incidence départementale : Affection de longue durée (ALD)⁴⁻⁶ et Programme de médicalisation des systèmes d'information (PMSI)⁷⁻¹⁰. Notre étude s'inscrit dans la continuité de ces travaux. Elle présente pour la première fois une évaluation conjointe, pour 23 localisations cancéreuses, des trois méthodes : PMSI/Incidence (PMSI/I), ALD/Incidence (ALD/I) et Mortalité/Incidence (M/I), dans une perspective d'estimation de l'incidence régionale des cancers. Les objectifs de notre étude étaient de :

- évaluer, par localisation cancéreuse, chaque méthode selon une méthodologie commune, dans les départements avec registre ;
- hiérarchiser les méthodes par localisation cancéreuse pour identifier les localisations éligibles (une méthode au moins jugée valide) et la méthode la plus appropriée ;
- estimer l'incidence régionale en France métropolitaine sur la période 2008-2010 pour les localisations éligibles.

Dans cet article, seuls les deux premiers objectifs sont abordés, l'ensemble des résultats étant disponibles en ligne sur le site Internet de l'Institut de veille sanitaire (<http://www.invs.sante.fr>).

Matériel et méthode

En préambule, notons que la méthode M/I repose sur un lissage préalable de l'incidence et de la mortalité à partir de 1975, selon une méthode adaptée de

celle utilisée au niveau national¹. Il s'agit donc d'une approche longitudinale, qui n'a pas pu être appliquée aux méthodes I/PMSI et I/ALD en l'absence d'antériorité suffisante des données des bases médico-administratives (BMA). Dans ce cas, l'approche est transversale, sans lissage préalable des données.

Matériel

La période d'étude a été déterminée par la disponibilité des données.

Les données d'incidence ont été extraites de la base commune des registres du réseau Francim (France-cancer-incidence et mortalité) et couvraient la période 1975-2009 (11 registres généraux : Calvados, Doubs, Isère, Hérault, Loire-Atlantique, Manche, Bas-Rhin, Haut-Rhin, Somme, Tarn, Vendée et 9 registres spécialisés : Ardennes et Marne : cancer de la thyroïde ; Gironde, Orne : localisations hématologiques ; Saône-et-Loire, Finistère : localisations digestives ; Côte-d'Or : cancer du sein, localisations digestives et hématologiques). Elles sont codées selon la Classification internationale des maladies pour l'oncologie, 3^e édition (CIM-O-3).

Les données de mortalité provenaient du Centre d'études sur les causes médicales de décès (CépiDc-Inserm) et couvraient la période 1975-2009. Elles sont codées selon la CIM-8 de 1975 à 1978, la CIM-9 de 1979 à 1999, puis la CIM-10 depuis 2000. Les données de mises en ALD de la période 2007-2010 provenaient des trois principales caisses d'assurance maladie : le régime général (CnamTS), le régime des indépendants (RSI) et le régime agricole (MSA), qui couvrent 96% de la population. Les données du PMSI de la période 2007-2010 ont été extraites des bases nationales. L'indicateur utilisé est le nombre de patients avec au moins un séjour et diagnostic principal de cancer sur la période d'étude. Ce choix est moins restrictif qu'une sélection intégrant également les actes médicaux, mais permet une sélection plus homogène⁹. Les données d'ALD correspondent au nombre de premières demandes de mise en ALD sur la période d'étude. Les ALD et le PMSI sont codés selon la CIM-10. Les données de la période 2007-2009, disponibles pour les différentes approches, servent à évaluer et hiérarchiser les méthodes par localisation. Les estimations portent ensuite sur la période plus récente 2008-2010.

Méthode

Le principe de l'estimation était identique pour les trois méthodes (PMSI, ALD et mortalité) et reposait sur l'application des rapports PMSI/I (ou ALD/I ou M/I, respectivement) de la zone registre aux données régionales du PMSI (ou ALD ou M, respectivement). Ces méthodes supposent que le rapport PMSI/I (ou ALD/I ou M/I, respectivement) était identique dans toutes les régions. La validité de cette hypothèse ne peut être examinée qu'au niveau géographique couvert par les registres des cancers, c'est-à-dire au niveau départemental.

Méthodes d'estimation de l'incidence

Modélisation du rapport PMSI/I (ou ALD/I ou M/I, respectivement), période 2007-2009

Pour la méthode PMSI/I (ou ALD/I), le rapport entre le nombre de patients issus du PMSI (ou nombre de nouvelles demandes d'ALD, respectivement) et le nombre de cas incidents dans les différents registres au cours de la période 2007-2009 a été modélisé en fonction de l'âge (en continu) et du département (effet aléatoire) dans un modèle de Poisson à effets mixtes¹¹. Le modèle s'écrit :

$$P_{a,j} | b_j \sim \text{Poisson}(\mu_{a,j}, C_{a,j}) \text{ avec } \text{Log}(\mu_{a,j}) = g(a) + b_j \text{ et } b_j \sim \mathcal{N}(0, \sigma^2)$$

Où $P_{a,j}$ est le nombre de patients issus du PMSI (ou d'ALD) de la classe d'âge a et du département j , $C_{a,j}$ le nombre de cas incidents correspondant, g une fonction continue de l'âge (spline) et b_j l'effet aléatoire du département j . Le paramètre σ^2 mesure la variabilité départementale du rapport. Ce modèle fournit le rapport moyen $f(a)$ utilisé ensuite pour estimer l'incidence par âge, égal à $\exp(g(a))$. L'analyse a été réalisée en utilisant la fonction *glmmPQL* de la librairie *mass* de S-Plus®, version 7.0.

Pour la méthode M/I, l'incidence et la mortalité dans la zone registre étaient chacune modélisées au préalable sur la période 1975-2009 (modèles âge-période-cohorte, voir Belot et coll.¹ pour la méthode détaillée). Le rapport M/I lissé $h(a)$ ainsi obtenu était extrait pour la période 2007-2009 dans la zone registre, par classe d'âge. Cette modélisation a été réalisée en utilisant la fonction *gam* de S-Plus®, version 7.0.

Estimation du nombre de cas incidents par région, 2008-2010

Pour la méthode PMSI/I (ou ALD/I), le nombre de cas incidents par région pour la période 2008-2010 était estimé à partir du nombre de patients PMSI de chacune des régions (ou ALD) et du rapport moyen $f(a)$ issu du modèle (1), sous l'hypothèse d'un rapport moyen 2007-2009 et 2008-2010 identiques. Ceci permet de réaliser des estimations d'incidence pour une période aussi récente que possible compte-tenu des données disponibles :

$$\hat{C}_{a,r} = P_{a,r} / f(a)$$

où $P_{a,r}$ est le nombre de patients PMSI (ou ALD) de la période 2008-2010 pour la classe d'âge a dans la région r .

Pour la méthode M/I, l'estimation du nombre de cas incidents par région était obtenue par classe d'âge, à partir du rapport M/I lissé $h(a)$ obtenu par modélisation, comme décrit au paragraphe précédent, et du nombre estimé de décès par région pour la période 2008-2010, par âge. Ce nombre était estimé à partir d'une modélisation préalable (âge-période-cohorte) de la mortalité régionale sur la période 1975-2009, avec une projection à un an pour l'année 2010.

Intervalle de prédiction des estimations régionales

Afin d'évaluer la variabilité des estimations, les estimations ponctuelles (nombre total de cas par région, taux standardisés sur la population mondiale et taux bruts) sont fournies avec leurs intervalles de prédiction (IP)¹². La variabilité départementale des rapports ALD/I et PMSI/I est prise en compte dans ces IP, en supposant que les erreurs de prédiction ne sont pas corrélées entre les différents départements d'une même région. À noter que cette hypothèse n'est pas vérifiable en raison du faible nombre et de la dispersion spatiale des départements couverts par un registre. Cette hypothèse induit une sous-estimation des variances des estimations régionales dans le cas où la corrélation serait, en fait, positive. Les intervalles de prédiction ne pouvaient être calculés de cette façon pour la méthode I/M et ne sont pas fournis pour cette méthode.

Évaluation de la qualité des estimations dans les départements avec registre, période 2007-2009

Cette évaluation reposait sur trois éléments : l'adéquation entre incidence observée et prédite au niveau départemental, la variabilité départementale des rapports PMSI/I, ALD/I et M/I et l'ampleur des disparités géographiques d'incidence.

Mesure de l'adéquation entre incidence observée et prédite

Pour évaluer l'adéquation entre incidence observée et prédite à partir du PMSI, des ALD et de la mortalité, une comparaison entre nombre de cas observés et prédits a été effectuée par validation croisée⁸ dans les différents départements couverts par un registre. L'erreur relative a été calculée pour chacun des départements couverts par un registre :

$$ER = \frac{E}{O} - 1$$

Où O est le nombre de cas incidents observés dans un département et E est le nombre de cas prédits. Pour évaluer la qualité de l'adéquation, la valeur absolue de l'erreur relative (ERA), exprimée en pourcentage a été utilisée.

Mesure de la variabilité départementale des rapports PMSI/I, ALD/I et M/I

Cette mesure (σ^2) est tirée directement du modèle (1) pour les rapports PMSI/I et ALD/I. Un modèle similaire a été réalisé pour le rapport M/I (où le nombre de décès est substitué au nombre patients PMSI), afin de fournir cette même mesure σ^2 . Cette mesure témoigne de la validité de l'hypothèse de ces méthodes, c'est-à-dire de la validité de l'application d'un même rapport à toutes les régions. Elle est étroitement liée à la qualité des estimations.

Mesure de la variabilité départementale de l'incidence observée

La variabilité départementale de l'incidence observée, notée σ_k , est obtenue par un modèle similaire à (1), dans lequel le taux d'incidence est analysé

en fonction de l'âge et du département en effet aléatoire. Elle témoigne de l'ampleur des disparités d'incidence entre les départements de la zone registre.

Classification et hiérarchisation des méthodes par localisation

Classification de chaque méthode, par localisation et par sexe

Afin d'évaluer la pertinence de chaque méthode pour les estimations régionales, une règle de décision a été établie (figure), conduisant au classement des localisations en 4 catégories : A++ et A+ (estimations acceptables) et B- ou B-- sinon (estimations non acceptables). Cette règle de décision est une évolution des règles adoptées précédemment dans le cadre des données d'ALD⁶ et du PMSI¹³.

Le classement reposait en premier lieu sur l'adéquation entre incidence observée et prédite. Si celle-ci était jugée satisfaisante (ERA faible dans tous les départements), la méthode était classée d'emblée en A++. Si elle était médiocre (ERA élevée dans un nombre important de départements), la méthode était classée en B--.

Pour les situations intermédiaires, un deuxième critère intervenait, comparant la variabilité départementale σ_d du rapport PSMI/I (ou ALD/I ou M/I, respectivement) à la variabilité de l'incidence (σ_k). Ce critère revient à considérer que des estimations,

bien qu'imparfaites, sont acceptables (A+) lorsque les erreurs de prédictions sont « modérées » au regard de l'ampleur des disparités d'incidence. Une méthode était retenue et classée A- lorsque la variabilité interdépartementale de l'incidence était 2 fois plus élevée que celle de du rapport PMSI/I (ou ALD/I ou M/I, respectivement). Sinon, la méthode était classée B-.

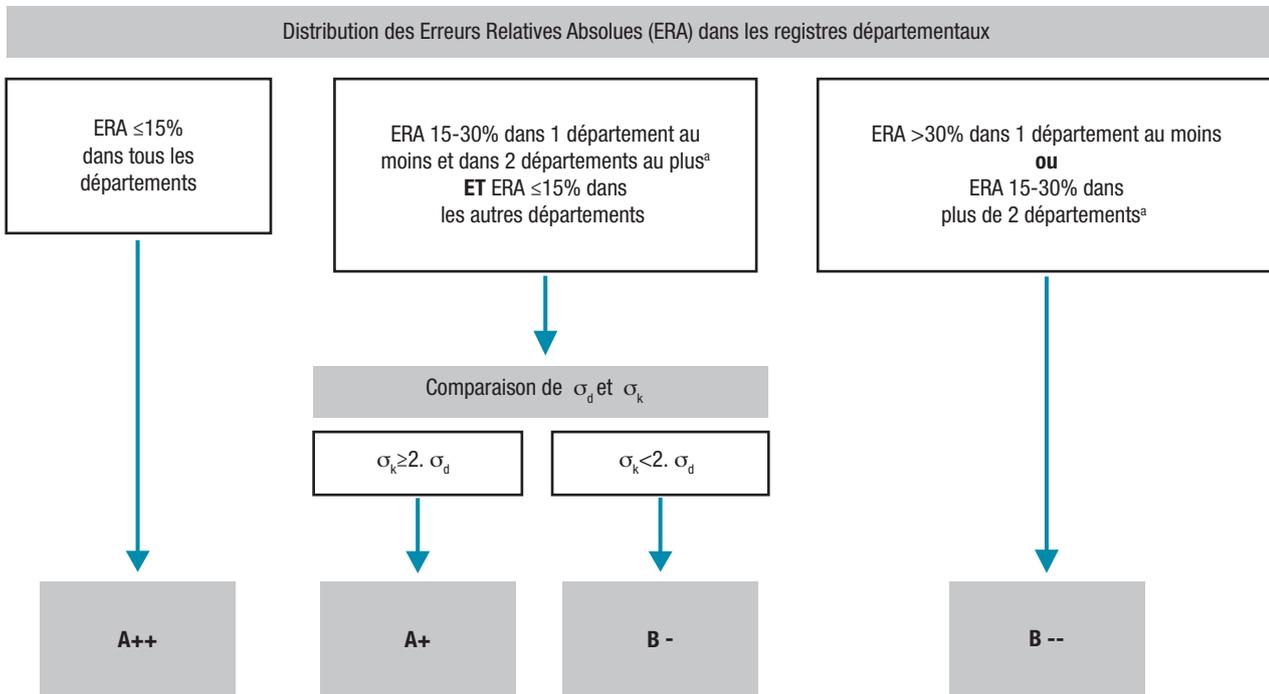
Les localisations éligibles avec au moins une méthode jugée satisfaisante (A++) ou acceptable (A+) ont été considérées comme éligibles pour une estimation régionale. Il faut noter que la significativité statistique des ERA n'intervient pas dans cette règle afin de ne pas écarter les localisations fréquentes lorsqu'elles ont des ERA significatives, mais mineures, et d'écarter au contraire les localisations avec des effectifs insuffisants pour évaluer la qualité des estimations. Ainsi, les classements B- et B-- regroupent à la fois des localisations pour lesquelles l'erreur de prédiction est jugée trop importante et des localisations avec des effectifs insuffisants pour évaluer la qualité des estimations.

Hiérarchisation des méthodes

Pour les localisations éligibles, la méthode la plus satisfaisante était retenue pour réaliser l'estimation de l'incidence (A++ préféré à A+). En cas d'ex-æquo, la méthode retenue était celle ayant l'ERA médiane la plus faible. Dans le cas où une même méthode permettait une estimation pour les deux sexes, cette

Figure

Arbre décisionnel utilisé pour classer les méthodes d'estimation des incidences régionales de cancer, France



^a Ce nombre est porté à 3 lorsque le nombre de registres atteint ou dépasse 13 (présence de registres spécialisés).

ERA : valeur absolue de l'erreur relative.

Note : le paramètre σ_k mesure la variabilité départementale de l'incidence alors que le paramètre σ_d mesure la variabilité départementale du rapport PMSI/I (respectivement ALD/I et M/I), où M est la mortalité et I l'incidence.

méthode était préférée aux autres, même si elle était moins performante pour un des sexes.

Gestion des localisations non-éligibles

Pour répondre à la demande forte des pouvoirs publics de disposer d'informations pour toutes les localisations cancéreuses, il a été décidé de fournir des fourchettes d'estimations (intervalle de prédiction : IP) pour les localisations non éligibles. En l'absence d'IP pour la méthode I/M, la méthode était choisie parmi PMSI/I ou ALD/I. En cas d'*ex-aequo*, la méthode ayant l'ERA médiane la plus faible était retenue.

Résultats

Parmi les 19 localisations cancéreuses étudiées chez les hommes, 8 ont fait l'objet d'une estimation d'incidence, dont 6 à partir des ALD et 2 à partir du PMSI (tableau 1). Parmi les 21 localisations étudiées chez les femmes, 9 ont fait l'objet d'une estimation d'incidence dont 3 à partir des ALD, 5 du PMSI et

1 de la mortalité. Les autres localisations n'étaient pas éligibles pour une estimation ponctuelle. Parmi les 12 localisations non éligibles chez les hommes, 8 étaient classées B-- pour les 3 méthodes. Chez les femmes, toutes les localisations non éligibles étaient classées B- - pour les 3 méthodes.

Le tableau 2 présente, pour chaque localisation et chaque méthode, une synthèse des éléments pris en compte pour le classement de la méthode : σ_d , σ_k , nombre de départements avec une erreur relative absolue $\leq 15\%$ et $> 30\%$. La mortalité est écartée d'emblée pour la grande majorité des localisations (B--) en raison d'erreur de prédictions trop importantes (tableau 2). Pour les 6 localisations spécifiques à un sexe, au moins une des méthodes ALD/I ou PMSI/I a été jugée éligible. Pour les 17 autres localisations étudiées, 4 sont éligibles pour les 2 sexes, 8 non éligibles pour les 2 sexes et 4 éligibles pour un seul sexe. Plus spécifiquement, un classement différent selon le sexe est rare pour les méthodes PMSI/I et M/I (1 et 3 localisations concernées respectivement), et plus fréquent pour la méthode ALD/I

Tableau 1

Classement des méthodes d'estimation d'incidence régionale de cancer en fonction de la localisation et du sexe, France

Localisations cancéreuses ^a	Homme					Femme				
	M/I	ALD/I	PMSI/I	Localisations éligibles : méthode retenue	Localisations non-éligibles : choix complémentaire	M/I	ALD/I	PMSI/I	Localisations éligibles : méthode retenue	Localisations non éligibles : choix complémentaire
Lèvre-bouche-pharynx	B - -	A++	A+	PMSI/I		B - -	B - -	A+	PMSI/I	
Œsophage	B - -	A+	B - -	ALD/I		B - -	B - -	B - -		PMSI/I
Estomac	B - -	B - -	B-		PMSI/I	A+	B - -	B - -	M/I	
Côlon-rectum	A+	B-	A++	PMSI/I		A++	A++	A++	PMSI/I	
Foie	B - -	B - -	B - -		PMSI/I	B - -	B - -	B - -		PMSI/I
Pancréas	B - -	B - -	B - -		ALD/I	B - -	B - -	B - -		ALD/I
Larynx	B - -	A+	B-	ALD/I		B - -	B - -	B - -		ALD/I
Poumon	B-	A++	B-	ALD/I		B - -	A++	B - -	ALD/I	
Mélanome de la peau	B - -	B - -	B - -		PMSI/I	B - -	B - -	B - -		ALD/I
Sein						B - -	A++	A++	ALD/I	
Col de l'utérus						B - -	B - -	A+	PMSI/I	
Corps de l'utérus						B - -	A+	A++	PMSI/I	
Ovaire						B - -	A++	A++	ALD/I	
Prostate	B - -	A++	B - -	ALD/I						
Testicule	B - -	A+	A+	ALD/I						
Vessie	B-	B-	B-		ALD/I	B - -	B - -	A+	PMSI/I	
Rein	B - -	B-	B-		PMSI/I	B - -	B - -	B - -		PMSI/I
Système nerveux central	B - -	B-	B - -		ALD/I	B - -	B - -	B - -		PMSI/I
Thyroïde	B - -	B - -	B - -		PMSI/I	B - -	B - -	B - -		PMSI/I
Lymphome malin non Hodgkinien	B - -	A++	B - -	ALD/I		B - -	B - -	B - -		ALD/I
Lymphome de Hodgkin	B - -	B - -	B - -		ALD/I	B - -	B - -	B - -		PMSI/I
Myélome multiple	B - -	B - -	B - -		ALD/I	B - -	B - -	B - -		ALD/I
Leucémies	B - -	B - -	B - -		ALD/I	B - -	B - -	B - -		ALD/I

M : mortalité ; I : incidence ; ALD : affection de longue durée ; PMSI : Programme de médicalisation des systèmes d'information.

^a codes CIM (voir [1])

Tableau 2

Critères de classement par approche, localisation cancéreuse et sexe, France

	Nb depts	Sexe	PMSI						ALD					Mortalité				
			σ_k	σ_d	ERA15	ERA 15-30	ERA 30	Décision	σ_d	ERA15	ERA 15-30	ERA 30	Décision	σ_d	ERA15	ERA 15-30	ERA 30	Décision
Localisations cancéreuses ne concernant qu'un seul sexe																		
Sein	12	F	0,043	0,037	12	0	0	A++	0,032	12	0	0	A++	0,111	7	5	0	B-
Col utérus	11	F	0,108	0,000	9	2	0	A+	0,161	6	3	2	B-	0,166	5	4	2	B-
Corps utérus	11	F	0,157	0,014	11	0	0	A++	0,063	10	1	0	A+	0,143	8	2	1	B-
Ovaire	11	F	0,077	0,000	11	0	0	A++	0,000	11	0	0	A++	0,126	7	3	1	B-
Prostate	11	H	0,100	0,104	8	3	0	B-	0,056	11	0	0	A++	0,158	5	2	4	B-
Testicule	11	H	0,071	0,034	9	2	0	A+	0,000	10	1	0	A+	0,394	2	3	6	B-
Localisations cancéreuses éligibles pour les hommes et les femmes																		
Lèvre-bouche-pharynx	11	H	0,180	0,059	10	1	0	A+	0,000	11	0	0	A++	0,001	7	3	1	B-
		F	0,086	0,000	10	1	0	A+	0,005	8	2	1	B-	0,001	8	3	0	B-
Côlon-rectum	14	H	0,060	0,041	14	0	0	A++	0,076	12	2	0	B-	0,000	13	1	0	A+
		F	0,044	0,007	14	0	0	A++	0,031	14	0	0	A++	0,062	14	0	0	A++
Poumon	11	H	0,070	0,089	9	2	0	B-	0,036	11	0	0	A++	0,055	9	2	0	B-
		F	0,141	0,067	8	3	0	B-	0,027	11	0	0	A++	0,109	7	3	1	B-
Localisations cancéreuses éligibles pour un seul sexe																		
Œsophage	14	H	0,283	0,116	10	3	1	B-	0,102	11	3	0	A+	0,102	10	3	1	B-
		F	0,263	0,092	9	2	3	B-	0,001	7	5	2	B-	0,001	6	3	5	B-
Larynx	11	H	0,124	0,091	9	2	0	B-	0,000	9	2	0	A+	0,151	5	4	2	B-
		F	0,105	0,001	8	1	2	B-	0,034	7	2	2	B-	0,473	2	2	7	B-
Vessie	11	H	0,143	0,097	9	2	0	B-	0,080	9	2	0	B-	0,080	9	2	0	B-
		F	0,109	0,000	10	1	0	A+	0,000	10	0	1	B-	0,001	6	4	1	B-
Estomac	14	H	0,090	0,072	12	2	0	B-	0,073	10	4	0	B-	0,034	9	5	0	B-
		F	0,167	0,094	9	5	0	B-	0,055	10	4	0	B-	0,044	11	3	0	A+
Lymphome malin non Hodgkinien	14	H	0,114	0,147	7	5	2	B-	0,042	14	0	0	A++	0,103	8	6	0	B-
		F	0,104	0,130	10	2	2	B-	0,013	11	2	1	B-	0,071	9	4	1	B-
Localisations cancéreuses non éligibles																		
Foie	14	H	0,251	0,167	9	3	2	B-	0,141	6	6	2	B-	0,130	9	2	3	B-
		F	0,210	0,133	9	2	3	B-	0,002	7	2	5	B-	0,239	6	4	4	B-
Pancréas	14	H	0,102	0,102	10	2	2	B-	0,024	10	4	0	B-	0,089	9	5	0	B-
		F	0,145	0,135	10	2	2	B-	0,056	11	2	1	B-	0,131	8	5	1	B-
Mélanome	11	H	0,347	0,354	4	2	5	B-	0,215	4	2	5	B-	0,240	2	3	6	B-
		F	0,258	0,267	3	5	3	B-	0,252	4	3	4	B-	0,001	5	4	2	B-
Rein	11	H	0,058	0,051	10	1	0	B-	0,059	10	1	0	B-	0,125	7	4	0	B-
		F	0,097	0,073	8	3	0	B-	0,101	7	3	1	B-	0,061	7	3	1	B-
Système nerveux central	11	H	0,029	0,100	8	1	2	B-	0,036	9	2	0	B-	0,001	8	2	1	B-
		F	0,072	0,056	8	2	1	B-	0,001	7	4	0	B-	0,000	7	3	1	B-
Thyroïde	13	H	0,351	0,149	8	4	1	B-	0,138	6	5	2	B-	0,403	3	2	8	B-
		F	0,433	0,164	9	3	1	B-	0,176	7	4	2	B-	0,352	1	3	9	B-
Lymphome de Hodgkin	14	H	0,001	0,090	8	5	1	B-	0,009	11	1	2	B-	0,241	2	7	5	B-
		F	0,141	0,045	10	2	2	B-	0,000	10	2	2	B-	0,003	3	2	9	B-
Myélome multiple	14	H	0,107	0,178	6	3	5	B-	0,087	8	5	1	B-	0,000	9	3	2	B-
		F	0,108	0,164	6	7	1	B-	0,080	9	5	0	B-	0,001	8	2	4	B-
Leucémies	14	H	0,141	0,245	4	4	6	B-	0,117	9	4	1	B-	0,179	8	4	2	B-
		F	0,163	0,230	5	5	4	B-	0,138	9	4	1	B-	0,133	8	5	1	B-

PMSI : Programme de médicalisation des systèmes d'information ; ALD : affection de longue durée ; ERA : valeur absolue de l'erreur relative. ERA15 : nombre de départements avec ERA < 15% ; ERA 15-30 : nombre de départements avec ERA > 15% et ≤ 30% ; ERA 30 : nombre de départements avec ERA > 30%.

Note : le paramètre σ_k mesure la variabilité départementale de l'incidence alors que le paramètre σ_d mesure la variabilité départementale du rapport PMSI/I (respectivement ALD/I et M/I), où M est la mortalité et I l'incidence.

(5 localisations concernées). Au final, 3 localisations sur ces 17 sont éligibles avec une même approche pour les deux sexes (lèvre-bouche-pharynx, côlon-rectum, poumon). Cinq localisations cancéreuses sont éligibles pour un seul sexe (œsophage, larynx et lymphomes malins non hodgkiniens (LMNH) chez les hommes, vessie et estomac chez les femmes). Les 9 localisations cancéreuses restantes sont non éligibles, avec dans la plupart des cas un classement défavorable (B--), c'est-à-dire avec des erreurs de prédictions importantes dans les départements couverts par un registre.

Discussion

En l'absence de registre national, la connaissance de l'incidence des cancers à un niveau infranational ne peut être obtenue qu'à partir d'une modélisation. Notre étude présente une évaluation des trois méthodes (PMSI/I, ALD/I, M/I) sur une unique période et selon une méthodologie commune. C'est la première démarche structurée réalisée au niveau régional, intégrant la méthode I/M utilisée jusqu'alors sans démarche d'évaluation. En adoptant des critères simples d'adéquation de l'incidence observée et prédite dans les départements avec un registre, et de comparaison de la variabilité de l'incidence observée avec celle du rapport PMSI/I (ou ALD/I ou M/I, respectivement), un nombre limité de localisations cancéreuses a été considéré comme éligible (8 localisations sur 19 chez les hommes et 9 sur 21 chez les femmes). Les localisations éligibles représentent toutefois plus de 70% des cas incidents¹.

Notre étude confirme les limites de l'utilisation de la mortalité pour estimer l'incidence régionale, limites qui avaient été énoncées mais non évaluées lors des précédentes estimations de l'incidence régionale en France³. Sur les 23 localisations cancéreuses étudiées, la mortalité n'est éligible que dans 3 cas et n'a été retenue que pour une seule localisation (estomac chez les femmes). Ceci est dû à une très mauvaise adéquation entre incidence observée et prédite dans les départements couverts par un registre. Pour 6 localisations chez les hommes (prostate, testicule, larynx, mélanome, thyroïde et maladie de Hodgkin) et 7 localisations chez les femmes (col de l'utérus, larynx, œsophage, foie, mélanome, thyroïde et maladie de Hodgkin), la majorité des départements présentent une erreur relative absolue supérieure à 15% dans l'étape de validation. Ces localisations cancéreuses présentant en majorité un pronostic favorable, le décalage temporel entre incidence et mortalité peut contribuer à la faible capacité prédictive de la mortalité. Un autre facteur explicatif probable est la variabilité géographique des pratiques diagnostiques pour les cancers dont l'incidence est très sensible à ces pratiques (*ie* : thyroïde, mélanome, prostate), ou encore le poids des causes compétitives de décès pour les cancers du sujet âgé (*ie* : prostate). Il est utile de préciser ici que si l'utilisation des données de mortalité pour estimer l'incidence régionale des cancers est inadaptée, son

utilisation pour l'estimation nationale de l'incidence n'est pas remise en cause².

Les BMA sont plus adaptées pour l'estimation de l'incidence régionale que les données mortalité, même si leur capacité prédictive reste limitée. Chez les hommes, 8 localisations sur 19 sont éligibles à partir des BMA et 9 sur 21 chez les femmes. Les raisons de la relativement faible capacité prédictive des ALD et du PMSI sont multiples : hétérogénéité potentielle dans le codage (*ie* : cancers hématologiques), utilisation de définitions différentes de la maladie (*ie* : cancer de la vessie), nature de la prise en charge thérapeutique (*ie* : surveillance pour un certain nombre de cancers de la prostate) et lieu de prise en charge (*ie* : mélanome de la peau). La faiblesse des effectifs pourrait aussi expliquer une partie des discordances selon le sexe (*ie* : cancers de l'œsophage et du larynx). Enfin, de façon générale, il y a nécessairement une certaine variabilité dans les résultats obtenus, en particulier pour les localisations se situant à la limite des seuils de décision. La reproductibilité des résultats sur une autre période devra être prise en considération pour consolider le classement.

La règle de décision adoptée ici pour classer les méthodes est une évolution de règles précédemment adoptées dans le cadre de l'utilisation des d'ALD⁶ et du PMSI¹³ à des fins d'estimation départementale des cancers, qui reposaient sur le même type d'éléments (erreur de prédiction, comparaison de σd et σk). Le niveau régional est cependant une échelle particulière pour l'estimation de l'incidence des cancers, car on ne peut évaluer la qualité des estimations directement à cette échelle, seul le niveau départemental étant accessible à la validation. Or, la qualité des estimations est sans doute moindre au niveau départemental que régional. Nous avons de ce fait assoupli les exigences sur la qualité des estimations pour le niveau régional par rapport aux précédents travaux visant à produire une estimation départementale de l'incidence. Des fourchettes d'estimations ont de plus été fournies, au niveau régional, pour les localisations non éligibles.

Par rapport aux travaux cités précédemment, nous avons également simplifié la règle adoptée, pour un énoncé plus facile, et nous avons appliqué une règle de décision stricte, sans faire intervenir l'expertise des professionnels du PMSI et des ALD dans le choix final du classement. Cette discussion avec les experts était en effet difficile à mettre œuvre dans le cadre d'une première confrontation de trois approches. Elle introduit aussi un élément qualitatif difficile à formaliser. Il serait intéressant toutefois de reconsidérer cet aspect à terme. La comparaison brute des classements obtenus ici à ceux des précédents travaux n'est pas d'interprétation immédiate, les règles décisionnelles ayant évolué avec des périodes d'études différant d'une étude à l'autre. Une comparaison fine, en appliquant la règle de décision adoptée pour le niveau régional aux précédents travaux, est prévue et permettra une réelle discussion des règles de décisions adoptées. Toutefois, on peut d'ores et déjà remarquer qu'un certain nombre de localisations

semblent particulièrement défavorables à l'utilisation des données médico-administratives pour une estimation d'incidence infranationale (mélanome, leucémie, myélome, cancer de la thyroïde).

Les critères que nous avons retenus pour la règle de décision ne reposent pas sur la significativité statistique des erreurs de prédiction. Une évaluation de la sensibilité de la règle de décision est donc très certainement à envisager afin d'identifier les localisations dont le classement serait changé avec de légères modifications des critères. Nous avons pris en considération non seulement la qualité des estimations, mais aussi un critère relatif mettant en regard qualité des estimations et ampleur des variations géographiques (comparaison de σd et σk). La « balance » entre ces deux critères est sujette à discussion. En abandonnant par exemple le critère d'exclusion basé sur une erreur relative absolue supérieure à 30%, mais en conservant la comparaison de σd et σk , le cancer du pancréas et le LMNH chez les femmes deviendraient éligibles. Le très mauvais ajustement d'un seul département pourrait avoir une origine très spécifique, ce qui ne devrait alors pas limiter l'utilisation des données d'ALD (ou PMSI) sur l'ensemble du territoire. Cette hypothèse d'une origine spécifique n'est toutefois pas vérifiable en pratique, car nous sommes rarement en mesure d'identifier les facteurs expliquant une telle configuration. De façon générale, il est extrêmement difficile d'identifier les facteurs expliquant les variations géographiques des rapports BMA/I. Ce point a été abordé dans les travaux sur l'utilisation des données d'ALD à des fins d'estimations départementales⁶. En particulier, on ne peut isoler des caractéristiques propres au département (eg défavorisation, caractère rural ou urbain, etc.), l'effet du département n'étant pas similaire selon les localisations. Les facteurs faisant varier les rapports BMA/I entre département sont très probablement multiples et résultent de spécificités locales, dépendantes de la localisation étudiée. Ils sont de ce fait très difficiles à mesurer.

L'utilisation des BMA pour une estimation de l'incidence régionale constitue une avancée par rapport à l'utilisation des données de mortalité, même si un nombre limité de localisations sont éligibles à ce jour. L'amélioration de la qualité du chaînage du PMSI depuis 2007 permettra à l'avenir d'exclure les cas prévalents de la sélection PMSI étudiée. Des perspectives méthodologiques sont de plus envisageables pour tenter d'améliorer l'utilisation des BMA et méritent d'être explorées: (i) l'évaluation des données appariées ALD-PMSI ; (ii) la prise en compte simultanément, dans un même modèle statistique, des données du PMSI, ALD et mortalité, qui nécessite un développement méthodologique important ; (iii) la prise en compte de données médico-administratives de manière longitudinale, sous réserve d'une homogénéité de la définition des cas au cours du temps. La modélisation permettrait dans ce cas de s'affranchir d'une partie des fluctuations des observations, qui sont importantes pour les localisations à faible incidence. Ces différentes pistes sont d'autant

plus importantes pour progresser sur le niveau départemental, échelon géographique tout aussi indispensable et pour lequel il est encore plus difficile d'obtenir des estimations de qualité. ■

Remerciements

À toutes les sources de données qui contribuent à l'enregistrement des cancers par les registres en particulier : les laboratoires et services d'anatomie et de cytologie pathologiques, les départements de l'information médicale (DIM) des établissements de soins publics et privés, les échelons locaux des services médicaux de l'assurance maladie, les cliniciens généralistes et spécialistes. Au Centre d'épidémiologie sur les causes médicales de décès (CépiDC-Inserm). Aux caisses d'assurance maladie : A. Weill (CnamTS), F. Frete (MSA) et M. Altana (RSI).

Financement

Cette étude a été en partie réalisée dans le cadre d'un financement INCa (Appel à projet) - Étude Oncepi, et s'inscrit dans le cadre du Programme de travail partenarial Francim-HCL-InVS-INCa (financement InVS-INCa).

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] Uhry Z, Belot A, Colonna M, Bossard N, Rogel A, Iwaz J, *et al.* National cancer incidence is estimated using the incidence/mortality ratio in countries with local incidence data: is this estimation correct? *Cancer Epidemiol.* 2013;37(3):270-7.
- [3] Colonna M, Bossard N, Mitton N, Remontet L, Belot A, Delafosse P, *et al.* le réseau Francim. Some interpretation of regional estimates of the incidence of cancer in France over the period 1980-2005. *Rev Epidemiol Santé Publique.* 2008;56(6):434-40.
- [4] 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 pour 24 localisations cancéreuses, 2000-2005. Saint-Maurice: Institut de veille sanitaire; 2011. 101 p. http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10431
- [5] 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, 2000-2005. *Bull Epidemiol Hebd.* 2012;(5-6):71-7. http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10440
- [6] Uhry Z, Remontet L, Colonna M, Belot A, Grosclaude P, Mitton N, *et al.* Cancer incidence estimation at a district level without a national registry: A validation study for 24 cancer sites using French health insurance and registry data. *Cancer Epidemiol.* 2013;37(2):99-114.
- [7] Uhry Z, Colonna M, Remontet L, Grosclaude P, Carré N, Couris CM, *et al.* Estimating infranational and national thyroid cancer incidence in France from cancer registries data and national hospital discharge database. *Eur J Epidemiol.* 2007;22(9):607-14.
- [8] 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.
- [9] Mitton N, Colonna M, Trombert B, Olive F, Gomez F, Iwaz J, *et al.* A Suitable approach to estimate cancer incidence in area without cancer registry. *J Cancer Epidemiol.* 2011;2011:418968. <http://dx.doi.org/10.1155/2011/418968>

[10] Bossard N, Gomez F, Remontet L, Uhry Z, Olive F, Mitton N, *et al.* Utilisation des données du PMSI pour estimer l'incidence des cancers en France à l'échelon infranational : la démarche du groupe Oncepi. *Bull Epidemiol Hebd.* 2012;(5-6):77-80. http://opac.invs.sante.fr/index.php?lvl=notice_display&id=10441

[11] Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *JASA.* 1993;88(421):9-25.

[12] Hakulinen T, Dyba T. Precision of incidence predictions based on Poisson distributed observations. *Stat Med.* 1994;13(15):1513-23.

[13] Colonna M. Pour une observation du cancer au niveau départemental en France : estimation de l'incidence et de la prévalence. Rapport d'activité, Appel à projets, INCa, 2013.

Citer cet article

Colonna M, Mitton N, Remontet L, Belot A, Bossard N, Grosclaude P, *et al.* Méthode d'estimation de l'incidence régionale des cancers à partir des données d'incidence des registres, des données de mortalité par cancer et des bases de données médico-administratives. *Bull Epidemiol Hebd.* 2013;(43-44-45):566-74.

> ARTICLE // Article

COLLECTE ET UTILISATION DES STADES DE CANCERS : PROBLÈMES POSÉS DANS LES ÉTUDES EN POPULATION

// COLLECTING AND USING DATA ON CANCER STAGES: PROBLEMS POSED IN POPULATION STUDIES

Pascale Grosclaude^{1,2} (pascale.grosclaude@inserm.fr), Anne-Marie Bouvier^{1,3}, Brigitte Trétarre^{1,4}, Patricia Delafosse^{1,5}

¹ Réseau français des registres de cancer Francim, France

² Registre des cancers du Tarn, Toulouse, France

³ Registre bourguignon des cancers digestifs, Dijon, France

⁴ Registre des cancers de l'Hérault, Montpellier, France

⁵ Registre des cancers de l'Isère, Grenoble, France

Soumis le 02.07.2013 // Date of submission: 07.02.2013

Résumé // Abstract

Le classement en stades d'un cancer selon son extension anatomique doit permettre une description à la fois précise et condensée de l'extension tumorale, dans le cadre de la prise en charge pluridisciplinaire et des études cliniques ; son utilisation est également indispensable dans les études épidémiologiques décrivant l'évolution de l'incidence ou de la survie des cancers.

Une bonne mesure du stade présente des difficultés dans le cas d'un enregistrement systématique en population. Ces difficultés, et les biais potentiels qui en découlent, sont illustrés par des exemples issus d'une étude sur les cancers recto-coliques diagnostiqués en 1990, réalisée par six registres français. Les problèmes posés par l'évaluation des dimensions constitutives de la classification TNM (notamment les glissements de classification) sont présentés, ainsi que les difficultés rencontrées dans la reconstitution des stades.

La complexité d'une utilisation correcte des stades en situation d'observation explique la prudence avec laquelle les résultats doivent être interprétés et la nécessité d'une standardisation de l'enregistrement du stade par les registres.

To be of value to multidisciplinary care and clinical trials, the classification of the anatomical extension of a malignant tumour into various stages must involve stage descriptions which are both precise and concise. Such a classification is also essential in epidemiological studies to describe the evolution of cancer incidence and survival.

It is difficult to accurately evaluate cancer stage in population-based studies. These difficulties and the potential resulting biases are illustrated by findings from a study on colon cancers diagnosed in 1990 carried-out in six French registries. The problems encountered when evaluating the essential components of the TNM classification system are presented (in particular the classification shifts), as well as the difficulties met when rebuilding stages due to missing data.

Given the complexities encountered in using stage classification correctly in observational studies, results must be interpreted with caution. These findings also highlight the need for population-based registries to standardise how cancer stage is recorded.

Mots-clés : Cancer, Registre, Stade Epidémiologie

// **Keywords**: Cancer, Registry, Tumoural extension, Epidemiology