

Estimations régionales et départementales d'incidence et de mortalité par cancers en France, 2007-2016

Matériel et méthodes

Auteurs

Édouard Chatignoux
Laurent Remontet
Marc Colonna
Pascale Grosclaude
Mélanie Cariou
Alice Billot-Grasset
Zoé Uhry

Estimations régionales et départementales d'incidence et de mortalité par cancers en France, 2007-2016

Matériel et méthodes

Étude collaborative

Réseau français des registres des cancers (réseau Francim), Service de Biostatistique-Bioinformatique des Hospices Civils de Lyon (HCL), Santé publique France, Institut national du cancer, dans le cadre du programme de travail partenarial relatif à la surveillance et à l'observation épidémiologiques des cancers à partir des registres.

Réalisation de l'étude

Extraction de la base commune des registres des cancers : Zoé Uhry (Santé publique France, Hospices Civils de Lyon).

Gestion et extraction des bases CnamTS : Mehdi Gabbas (CnamTS), Marjorie Boussac-Zarebska (Santé publique France).

Data-management : Elsa Decool (Santé publique France).

Analyse : Édouard Chatignoux (Santé publique France), Laurent Remontet (Hospices Civils de Lyon), Marc Colonna (Francim), Pascale Grosclaude (Francim), Alice Billot-Grasset (Francim), Zoé Uhry (Santé publique France, Hospices Civils de Lyon).

Coordination de la publication

Santé publique France : Florence de Maria, Olivier Catelinois, Philippe Pépin, Philippe Bouvet de la Maisonneuve, Édouard Chatignoux.

Francim : Alice Billot-Grasset, Marc Colonna, Pascale Grosclaude, Mélanie Cariou.

Institut national du cancer : Lionel Lafay, Philippe-Jean Bousquet.

Remerciements

Les auteurs remercient l'ensemble des sources de données qui contribuent à l'enregistrement des cancers par les registres, en particulier les laboratoires d'anatomie et de cytologie pathologiques, les Départements de l'information médicale des établissements de soins publics et privés, les échelons locaux des services médicaux de l'Assurance maladie, les médecins généralistes et spécialistes.

Conception et réalisation graphique

Les couvertures ont été réalisées par Anne-Sophie Mélard (Santé publique France).

Financement

Santé publique France, Institut national du cancer.

TABLE DES MATIÈRES

Abréviations	4
1 Introduction	5
2 Matériel	5
2.1 Définition des localisations de cancer étudiées	5
2.2 Données d'incidence observées	5
2.3 Données d'incidence prédites	7
2.4 Données de mortalité	9
2.5 Données de population	9
3 Lissage spatial de l'incidence et de la mortalité pour la cartographie	9
3.1 Objectifs	9
3.2 Modèle de lissage	10
3.3 Illustration de l'effet du lissage sur l'estimation des rapports standardisés d'incidence et de mortalité	12
4 Contenu des profils régionaux	15
4.1 Indicateurs présentés dans les profils	15
4.2 Présentation des profils et guide de lecture	16
Références	17

ABRÉVIATIONS

ALD	Affection de longue durée
ALD/I	ALD/incidence
ATIH	Agence technique de l'information sur l'hospitalisation
AUP	ALD union PMSI
AUP/I	AUP/incidence
CIM-10	Classification internationale des maladies 10 ^e révision
CIM-O-3	Classification internationale des maladies pour l'oncologie 3 ^e édition
ER	Erreur relative de prédiction
Francim	Réseau des registres français des cancers
HCL	Hospices Civils de Lyon
IC	Intervalle de confiance
INCa	Institut national du cancer
Insee	Institut national de la statistique et des études économiques
LMNH	Lymphome malin non-hodgkinien
MA	Données médico-administratives
PMSI	Programme de médicalisation des systèmes d'information
PMSI/I	PMSI/incidence
I	Incidence
SIR	Rapport standardisé d'incidence
SMR	Rapport standardisé de mortalité

1. INTRODUCTION

L'incidence et la mortalité par cancers sont des indicateurs épidémiologiques indispensables 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 des territoires (régions et départements) est essentielle.

Cette étude vise à produire des profils régionaux présentant l'incidence et la mortalité par cancer pour une vingtaine de cancers sur la période 2007-2016. Ces profils ont un double objectif : i) situer la région et ses départements au sein de la France en termes d'incidence et de mortalité par cancer et ii) détailler ces indicateurs pour la région et ses départements. Ce document présente le matériel, les méthodes, ainsi que le contenu-type des profils régionaux et un guide de lecture.

Les données de mortalité couvrent l'ensemble du territoire, mais les données d'incidence des cancers ne sont disponibles que dans les départements couverts par un registre : aussi l'incidence au niveau départemental et régional des cancers doit être estimée. Ces estimations d'incidence ont été réalisées à partir de données médico-administratives (MA) en utilisant un modèle de calibration [1-8] : le rapport MA/Incidence observé dans les départements avec registre est utilisé pour redresser les données MA observées dans chaque département. Une évaluation préalable de la qualité de ces estimations a été réalisée au sein de la zone registre, et ce pour chaque cancer et sexe séparément, et fait l'objet d'un rapport spécifique¹. Trois indicateurs MA ont été testés, issus des Affections Longue Durée, des données hospitalisations du PMSI ou du croisement individuel de ces deux sources, afin de déterminer celui fournissant les meilleures prédictions d'incidence et d'évaluer si ces prédictions étaient d'une qualité suffisante. Pour les cancers pour lesquels la qualité était jugée suffisante, les estimations départementales et régionales d'incidence sont présentées dans les profils régionaux. Pour les cancers pour lesquels aucun des indicateurs ne permettaient une estimation satisfaisante de l'incidence, seule l'incidence observée dans les départements couverts par un registre est présentée.

Par ailleurs, afin de visualiser les gradients géographiques d'incidence et de mortalité par cancer, ces indicateurs ont été cartographiés après mise en œuvre d'un lissage spatial.

Ce document présente les sources de données et les méthodes utilisées pour produire les indicateurs régionaux d'incidence et de mortalité par cancer. Cette étude s'inscrit dans la continuité des travaux sur les estimations infra-nationale de l'incidence, et a pour objectif de fournir des indicateurs valides d'incidence et de mortalité aux niveaux régional et départemental.

2. MATÉRIEL

La période d'étude a été déterminée par la disponibilité des données au moment de l'analyse. Elle porte sur la période 2007-2014 en ce qui concerne les données observées d'incidence et de mortalité, et sur la période 2007-2016 pour les incidences prédites.

2.1 Définition des localisations de cancer étudiées

Les codes diagnostics utilisés pour définir les 24 localisations cancéreuses étudiées sont détaillés dans le tableau 2-1. Ils correspondent aux codes utilisés dans les précédentes études portant sur les estimations de l'incidence régionale et départementale [8, 9].

2.2 Données d'incidence observées

Les données d'incidence ont été extraites de la base commune des registres du réseau Francim pour la période 2007-2014. Elles ont été sélectionnées dans les registres généraux et spécialisés suivants :

- 21 registres généraux : 14-Calvados, 16-Charente, 17-Charente-Maritime, 25-Doubs, 33-Gironde, 34-Hérault, 38-Isère, 44-Loire-Atlantique, 50-Manche, 59-Lille, 67-Bas-Rhin, 68-Haut-Rhin, 79-Deux-Sèvres, 80-Somme, 81-Tarn, 85-Vendée, 86-Vienne, 87-Haute-Vienne, 971-Guadeloupe, 972-Martinique, 973-Guyane.

1. Rapport "Évaluation de l'utilisation des données médico-administratives pour estimer l'incidence départementale : comparaison de l'incidence observée et prédite dans les registres sur la période 2007-2014"

- 9 registres spécialisés : 21-Côte d'Or (Sein), 21-Côte d'Or (Hématologie), 08-51-Marne-Ardennes (Thyroïde), 14-Calvados (Digestif), 21-71-Saône-et-Loire (Digestif), 29-Finistère (Digestif), 14-50-61-Basse Normandie (Hématologie), 33-Gironde (Hématologie), 33-Gironde (Système nerveux central).

Les données d'incidence sont codées selon la Classification internationale des maladies pour l'oncologie, 3^e édition (CIM-O-3).

| TABLEAU 2-1 |

Définition des localisations de cancer étudiées

Localisation	Incidence ¹ (CIM-O-3)	ALD, PMSI ¹ (CIM10)	PMSI non invasifs ² (CIM10)
Lèvre-bouche-pharynx	C00-C14	C00-C14	D000, D370
Œsophage	C15	C15	D001
Estomac	C16	C16	D002
Côlon-rectum	C18-C21	C18-C21	D01
Foie	C22	C22	D015
Pancréas	C25	C25	D017
Larynx	C32	C32	D020
Poumon	C33-C34	C33-C34	D021-D022
Mélanome de la peau	C44 ³	C43	D03
Sein	C50	C50	D05
Col de l'utérus	C53	C53	D06
Corps de l'utérus	C54	C54	D070
Ovaire ⁴	C56, C570, C571, C572, C573, C574	C56-C57	
Prostate	C61	C61	D075
Testicule	C62	C62	D076
Vessie	C67	C67	D090, D414
Rein	C64-C66, C68	C64-C66, C68	D091
Système nerveux central ⁵	C70-C72	C70-C72	D32, D33, D353, D42, D43, D444, G930
Thyroïde	C73	C73	D093
Lymphome malin non-hodgkinien	5	C82-C85	
Lymphome de Hodgkin	6	C81	
Myélome multiple et maladies immunoprolifératives	7	C88, C90	
Leucémies aiguës myéloïdes	8	C920, C924, C925, C930, C942	
Tous cancers	C	C	

¹ Définition issue de Belot *et al.* [10]

² Codes utilisés pour repérer les patients mis en ALD ayant une hospitalisation la même année pour un même cancer in situ ou bénin

³ Morphologies 97203 à 97803

⁴ Toutes les morphologies sauf 84423, 84513, 84613, 84623, 84723 et 84733

⁵ Morphologies de 9590 à 9597, de 9670 à 9719, de 9724 à 9729 et de 9832 à 9838

⁶ Toutes topographies, Morphologies de 9650 à 9667

⁷ Toutes topographies, Morphologies de 9731 à 9734 et de 9760 à 9764

⁸ Morphologies 9840, 9860, 9861, 9866, 9867, de 9870 à 9874, de 9891 à 9931, 9984, 9805, de 9806 à 9809, 9865, 9869, 9911 et 9898

2.3 Données d'incidence prédites

Cette partie décrit brièvement la méthodologie adoptée pour prédire l'incidence départementale et régionale des cancers et valider leur utilisation, et présente la méthode de prédiction retenue pour chaque localisation éligible.

Les méthodes de prédictions sont détaillées dans un article spécifique [1] et dans le rapport sur les évaluations des méthodes, qui présente également le détail des sources de données utilisées et des résultats des évaluations de la qualité des prédictions.

Pour ce travail, trois indicateurs basés sur des données médico-administratives (indicateurs MA) étaient utilisés pour réaliser les prédictions :

1. ALD : nombre de nouvelles mises en ALD.
2. PMSI : nombre d'hospitalisations incidentes après élimination des cas prévalents à N-1 et N-2.
3. AUP : nombre de nouvelles mises en ALD ou d'hospitalisations incidentes et exclusion des nouvelles mises en ALD ayant une correspondance avec un cancer non invasif dans le PMSI.

2.3.1 Principes de la prédiction et de la validation

Modélisation du rapport MA/I et prédiction de l'incidence selon le sexe

Le principe de prédiction de l'incidence dans un département consiste à appliquer le rapport MA/I moyen par âge mesuré dans la zone registre (i.e. départements couverts par un registre de cancers) aux données MA du département. Par exemple, si, dans la zone registre, on observe, pour un âge donné, 100 cas de cancers incidents et 200 patients hospitalisés pour ce cancer dans le PMSI, le nombre de cancers incidents pour cet âge sera prédit dans un département à partir du PMSI par le nombre de patients hospitalisés divisé par deux. Le rapport MA/I par âge est modélisé dans un modèle de Poisson à effets mixtes, qui permet de tenir compte de la variabilité du rapport MA/I dans les différents départements.

Les prédictions sont d'autant plus proches du nombre de cas incidents que les ratio MA/I de chaque département sont proches du ratio MA/I moyen, i.e. que le ratio MA/I varie peu entre les départements. Les modèles permettent de fournir une mesure σ_d de cette variabilité (σ_d est d'autant plus important que la variabilité des ratios est importante), prise en compte dans le calcul des variances et des intervalles de prédiction.

Évaluation de la qualité des prédictions

Afin d'évaluer la pertinence de chaque méthode pour chacun des cancers, une règle de décision a été établie en classant les localisations en 4 catégories : A++ et A+ (localisations éligibles), B- et B-- (localisations non-éligibles). Le classement repose en premier lieu sur la comparaison au niveau départemental entre l'incidence observée et l'incidence prédite (adéquation mesurée par l'erreur relative de prédiction (ER), qui est la différence dans un département couvert par un registre entre le nombre total de cas de cancers incidents observé et prédit). Si l'adéquation est jugée satisfaisante ($|ER| < 15\%$ dans tous les départements), la méthode est classée d'emblée en A++. Si elle est médiocre ($|ER| > 30\%$ dans un département, ou $> 15\%$ dans plus de 20% des départements), la méthode est classée en B--.

Dans les situations intermédiaires (adéquation ni totalement satisfaisante ni médiocre), un second critère intervient. Ce critère a pour objectif de tolérer des erreurs un peu plus importantes lorsque les disparités d'incidence (mesurées par un indicateur σ_k) sont fortes d'un département à l'autre (classement A+). Dans le cas contraire la méthode est classée B-. Le classement A+ correspond donc à la situation où malgré des erreurs non négligeables dans l'adéquation, les prédictions restent informatives car elles permettent de rendre compte d'importantes disparités géographiques d'incidence.

Validation et sélection de la méthode de prédiction

Il s'agit, pour chaque localisation cancéreuse, de décider si une prédiction de l'incidence régionale et départementale peut être produite (localisation dite éligible), et dans l'affirmative, d'identifier la méthode la plus appropriée. Les règles suivantes ont été utilisées :

- aucune méthode classée A : pas de prédiction d'incidence (localisation non-éligible) ;
- une seule méthode classée A : prédiction avec cette méthode ;
- plusieurs méthodes classées A :

- * si une seule méthode classée A++, prédiction avec cette méthode ;
- * si des méthodes sont ex-aequo (e.g plusieurs méthodes classées A++), prédictions avec la méthode présentant l'|ER| médiane la plus petite. Toutefois, dans le cas où une même méthode permettrait une prédiction pour les deux sexes, cette méthode est préférée aux autres, même si elle est moins performante pour un des sexes.

2.3.2 Localisations éligibles et méthodes retenues

Les localisations retenues, la méthode utilisée, le classement (A+ ou A++), la variabilité géographique du rapport MA/I notée σ_d et la variabilité géographique de l'incidence dans la zone registre notée σ_k , sont reportées dans le Tableau 3-1. Le résultat détaillé des évaluations ayant conduit, ou non, au choix d'une méthode parmi les trois étudiées est disponible dans le rapport spécifique d'évaluation¹.

Aucune méthode n'a pu être validée au niveau départemental pour estimer l'incidence des cancers du larynx, des mélanomes, du myélome multiple et maladies immunoprolifératives et des leucémies aiguës myéloïdes. Quelques divergences étaient également observées entre sexes, une méthode étant validée uniquement chez les hommes pour les cancers du foie, du système nerveux central et pour les LMNH. Au total, 16 localisations ont donné lieu à des prédictions départementales d'incidence chez les hommes et 15 chez les femmes.

Les écarts entre l'incidence observée et prédite dans les départements couverts par un registre étaient limités (erreurs de prédictions <15%, méthode classée "A++") pour une majorité de localisations (LBP, œsophage (hommes), estomac (femmes), côlon-rectum, foie (hommes), pancréas, poumon, sein, corps de l'utérus, ovaire, prostate, testicule, vessie (hommes), rein, SNC (hommes) et tous cancers).

Les écarts étaient plus importants pour les cancers de l'œsophage (femmes), de l'estomac (hommes), du col de l'utérus, de la vessie (femmes), de la thyroïde, des LMNH (hommes) et du lymphome de Hodgkin mais jugés toutefois acceptables au regard de la variation géographique de l'incidence (méthode classée "A+").

TABLEAU 3-1

Localisations cancéreuses éligibles et méthodes de prédiction utilisées

Localisation	Hommes				Femmes			
	Méthode	Classement	$\sigma_d \times 100$	$\sigma_k \times 100$	Méthode	Classement	$\sigma_d \times 100$	$\sigma_k \times 100$
Lèvre-bouche-pharynx	PMSI	A++	2,2	16,4	PMSI	A++	0,0	10,2
Oesophage	AUP	A++	6,2	25,9	AUP	A+	5,7	18,0
Estomac	AUP	A+	4,8	10,8	AUP	A++	3,0	17,9
Côlon-rectum	ALD	A++	2,8	4,3	ALD	A++	1,8	4,3
Foie	AUP	A++	6,6	22,3	-	-	-	-
Pancréas	AUP	A++	2,4	7,9	AUP	A++	3,0	9,4
Poumon	ALD	A++	1,2	10,4	ALD	A++	1,9	13,4
Sein	-	-	-	-	AUP	A++	2,5	3,7
Col de l'utérus	-	-	-	-	PMSI	A+	5,4	14,8
Corps de l'utérus	-	-	-	-	PMSI	A++	0,0	11,5
Ovaire	-	-	-	-	AUP	A++	0,0	3,6
Prostate	ALD	A++	1,5	9,6	-	-	-	-
Testicule	AUP	A++	0,0	10,8	-	-	-	-
Vessie	ALD	A++	3,5	9,1	ALD	A+	0,5	7,8
Rein	AUP	A++	2,1	9,0	AUP	A++	4,2	9,7
Système nerveux central	PMSI	A++	3,8	4,8	-	-	-	-
Thyroïde	AUP	A+	6,5	36,5	AUP	A+	8,5	39,8
LMNH	ALD	A+	4,3	10,4	-	-	-	-
Lymphome de Hodgkin	ALD	A+	0,1	1,3	ALD	A+	0,0	2,8
Tous cancers	ALD	A++	1,9	4,9	ALD	A++	2,1	3,9

1. Rapport "Évaluation de l'utilisation des données médico-administratives pour estimer l'incidence départementale : comparaison de l'incidence observée et prédite dans les registres sur la période 2007-2014"

2.4 Données de mortalité

Les données de mortalité proviennent du Centre d'épidémiologie sur les causes médicales de décès (CépiDc) et couvrent la période 2007-2014. Elles sont codées selon la Classification internationale des maladies 10^e révision (CIM-10).

Du fait qu'il existe une proportion importante et variable de certificats de décès par cancer de l'utérus ne différenciant pas col et corps utérin [11], les indicateurs de mortalité n'ont pas pu être étudiés pour le col et le corps de l'utérus.

2.5 Données de population

Les données de population sont issues des estimations de l'Insee actualisées en janvier 2018, détaillées par sexe, âge révolu, année et département. À partir de ces données de population sont calculés les nombres de personnes-années utilisés pour le calcul des taux d'incidence.

3. LISSAGE SPATIAL DE L'INCIDENCE ET DE LA MORTALITÉ POUR LA CARTOGRAPHIE

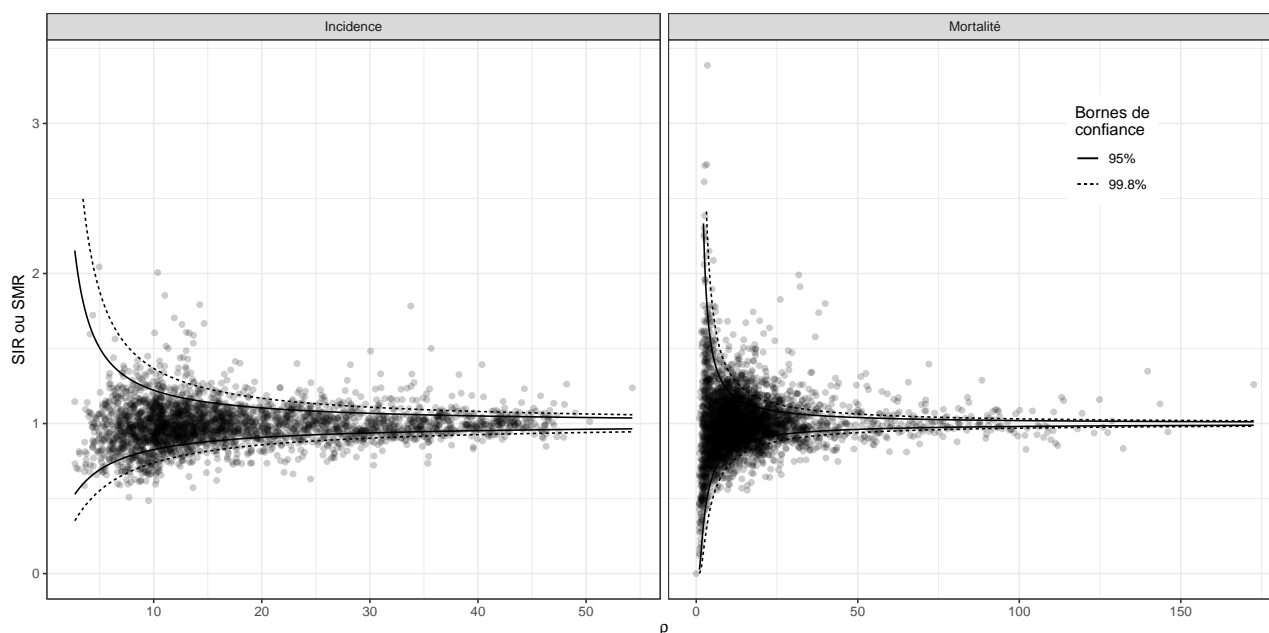
3.1 Objectifs

L'objectif de la représentation géographique de l'incidence et de la mortalité par cancer est de décrire leur distribution spatiale, afin de mettre en évidence de potentiels gradients géographiques de risque ou des zones à risque élevé. Pour ce faire, l'incidence et la mortalité sont habituellement représentées sous forme de rapports standardisés de morbidité (SIR) et de mortalité (SMR).

Une difficulté rencontrée avec ce type de représentations [12] est que les unités spatiales avec des estimations incertaines (e.g. petites unités géographiques, cancers rares...) ont plus de chance de présenter des valeurs extrêmes des SIR ou des SMR, du seul fait des fluctuations aléatoires. Il est dès lors difficile de distinguer les fluctuations aléatoires des estimations d'une vraie variation du niveau de risque entre les zones géographiques étudiées, d'autant plus que les estimations dans les différentes unités géographiques ont des niveaux de précision inégaux.

Ce phénomène est illustré sur la figure 1-1, qui représente les SIR et les SMR départementaux pour les localisations cancéreuses éligibles en fonction de leur précision ρ dans un funnel plot (ou graphe en entonnoir - voir encadré ci-après). La forme en entonnoir du funnel plot illustre bien les problèmes rencontrés lors de la cartographie des indicateurs d'incidence ou de mortalité : l'amplitude des valeurs des SIR et des SMR compatibles avec l'hypothèse nulle SIR=1 et SMR=1 est grande lorsque la précision est faible. Aussi, les cartes d'incidence et de mortalité auront tendance à être plus contrastées pour des données imprécises (localisations rares, petits départements), ces contrastes étant le plus souvent non significatifs. Inversement, plus la précision est grande, plus l'amplitude des valeurs des SIR et des SMR compatibles avec l'hypothèse nulle SIR=1 se réduit, et plus les contrastes représentés sont significatifs.

Variabilité des SIR 2007-2016 prédits avec la méthode (MA/I) retenue et des SMR 2007-2014 toutes localisations en fonction de leur précision



Note : Un point représente le SIR ou le SMR d'un département pour une localisation cancéreuse. Toutes les localisations et tous les départements sont représentés simultanément sur la figure.

Funnel plot

Le funnel plot (ou graphe en entonnoir) présente les rapports standardisés d'incidence (SIR) et de mortalité (SMR) départementaux (en ordonnée, un point par département) en fonction de leur précision ρ (en abscisse), avec des bornes de confiance à 95% et à 99,8% [13, 14]. La précision ρ est ici mesurée par l'inverse du coefficient de variation du SIR ou du SMR, estimé pour le SIR par exemple par $\rho = \hat{SIR} / \sqrt{Var(\hat{SIR})}$. Les bornes de confiance à 95% et à 99,8% sont également représentées. Elles sont calculées sous les hypothèses nulles SIR=1 ou SMR=1, en supposant une loi log-normale pour les SIR et une loi de Poisson pour les SMR. En d'autres termes, les bornes de confiance permettent de fournir les valeurs des SIR et des SMR compatibles avec les hypothèses nulles SIR=1 et SMR=1, en tenant compte des différents niveaux de précision des SIR et des SMR. Les bornes de confiance sont calculées, par exemple pour le niveau à 95% des SIR, de la manière suivante : $\exp(-1,96/\rho)$ et $\exp(1,96/\rho)$.

3.2 Modèle de lissage

Pour obtenir des données plus informatives et vraisemblables pour la cartographie, un lissage peut être utilisé [12, 15, 16]. L'objectif est de stabiliser les estimations des SIR et des SMR d'un département en utilisant les données de ce département mais également des autres départements. Des modèles hiérarchiques bayésiens sont habituellement utilisés pour réaliser ces lissages.

Le modèle utilisé dans cette étude est le modèle BYM de Besag, York et Mollié [17]. Dans ce modèle, la variabilité des rapports (SMR ou SIR) est décomposée en une composante spatiale et une composante non structurée spatialement, ou composante d'hétérogénéité. Ce modèle donne une estimation du rapport (SIR ou SMR) qui est, pour un département, un compromis entre le rapport moyen, le rapport moyen des départements voisins et le rapport observé du département.

Il est à noter que l'effet du lissage spatial sera d'autant plus important que les SIR ou les SMR sont imprécis, comme illustré sur la figure 3-1.

Lissage de la mortalité

Le nombre de décès D_j dans un département j est supposé suivre une loi de Poisson de moyenne δ_j :

$$\begin{cases} D_j \sim \mathcal{P}(\delta_j) \\ \log(\delta_j) = \log(E_j) + \lambda_j \end{cases} \quad (1)$$

où E_j est le nombre de décès attendus dans le département j et λ_j l'écart à l'attendu. $\exp(\lambda_j)$ mesure le risque relatif du département j par rapport à l'attendu (ou le SMR).

Le modèle BYM consiste à préciser une distribution *a priori* sur le log-risque λ_j qui s'exprime comme la somme d'un effet structuré spatialement θ_j , et d'un effet non structuré ϕ_j .

Plus précisément, le vecteur $\theta = (\theta_1 \dots \theta_J)^\top$ a une distribution autorégressive conditionnelle (CAR intrinsèque), définie par :

$$\begin{cases} \theta | \lambda_\theta \sim \text{CAR}(\lambda_\theta) \\ (\theta_j | \theta_i, i \neq j, i \sim j, \lambda_\theta) \sim \mathcal{N}\left(\frac{1}{n_i} \sum_{i \sim j} \theta_i, \frac{\lambda_\theta^2}{n_i}\right) \end{cases}$$

où $i \sim j$ indique que les départements i et j sont voisins (partagent une frontière commune), et n_i est le nombre de voisins du département i .

ϕ_j est un terme d'hétérogénéité, sans structure spatiale, distribué selon une loi normale de variance λ_ϕ^2 :

$$\phi_j | \lambda_\phi \sim \mathcal{N}(0, \lambda_\phi^2)$$

La spécification d'une loi *a priori* sur les hyperparamètres λ_θ et λ_ϕ complète la formulation du modèle.

Lissage des prédictions d'incidence

Les prédictions d'incidence fournissent un estimateur $\hat{\tau}_j$ de l'espérance du nombre de cas incidents τ_j dans le département j . Cet estimateur est distribué selon une loi log-normale :

$$\hat{\tau}_j \sim \mathcal{LN}(\tau_j, V(\hat{\tau}_j)) \quad (2)$$

où $V(\hat{\tau}_j)$ est connue.

Un modèle BYM suivant la même formulation que pour la mortalité peut ainsi être utilisé pour l'incidence :

$$\begin{cases} \hat{\tau}_j \sim \mathcal{LN}(\tau_j, V(\hat{\tau}_j)) \\ \log(\tau_j) = \log(E_j) + \theta_j + \phi_j \\ \theta | \lambda_\theta \sim \text{CAR}(\lambda_\theta) \\ \phi_j | \lambda_\phi \sim \mathcal{N}(0, \lambda_\phi^2) \end{cases} \quad (3)$$

Hyperparamètres

Les lois *a priori* sur l'inverse des hyperparamètres (*précisions*) sont prise comme des lois Gamma, avec des paramètres a et b choisis pour refléter un *a priori* uniforme entre 0.01 et 0.4 pour λ_θ et λ_ϕ [18](ce qui correspond à l'intervalle des variations spatiales habituellement observées).

Implémentation

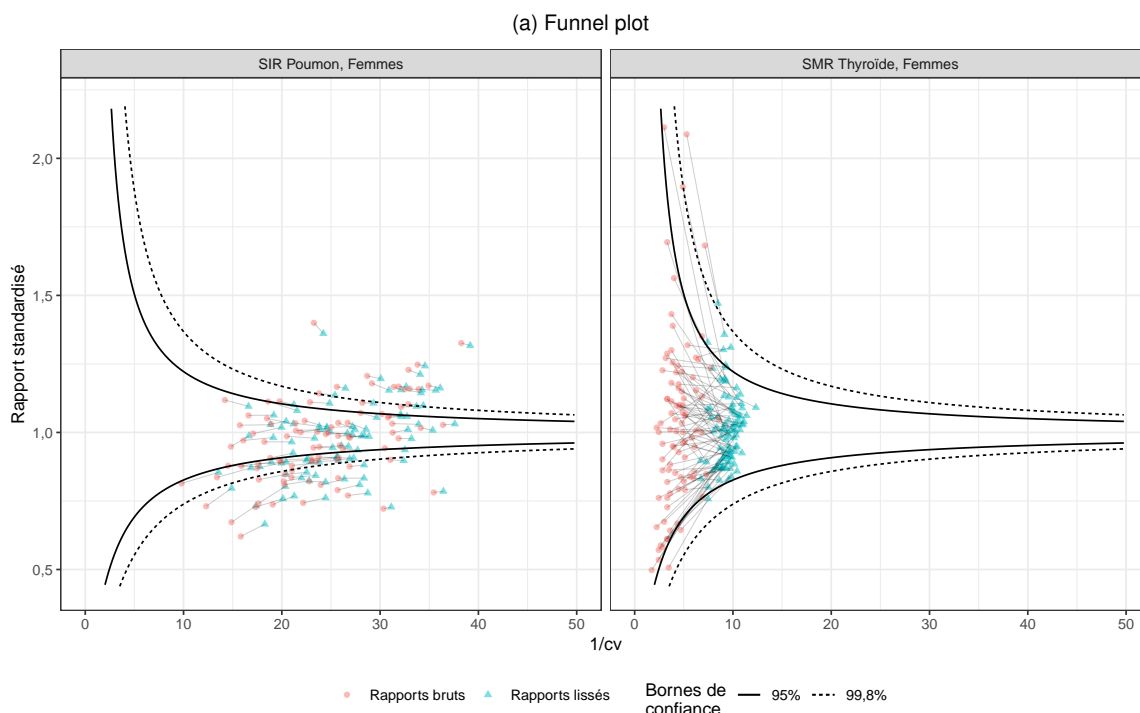
Les modèles de lissage ont été estimés en utilisant des approximations de Laplace [19], implémentées dans le logiciel R-INLA.

3.3 Illustration de l'effet du lissage sur l'estimation des rapports standardisés d'incidence et de mortalité

Les graphiques de la figure 3-1 illustrent l'effet du lissage sur deux indicateurs : l'incidence du cancer du poumon chez les femmes, et la mortalité par cancer de la thyroïde chez les femmes. Pour le premier, les estimations sont précises (le cancer du poumon touche plus de 11 000 femmes en moyenne par an entre 2007 et 2016 et la méthode de prédiction, classée A++, est fiable), et l'effet du lissage est peu marqué. Pour le second en revanche, les rapports de mortalité standardisés sont instables (246 cas par an en France en moyenne entre 2007 et 2014), et le lissage permet de stabiliser les SMR.

| FIGURE 3-1 |

Effet du lissage sur l'estimation des rapports standardisés d'incidence et de mortalité



Note de lecture

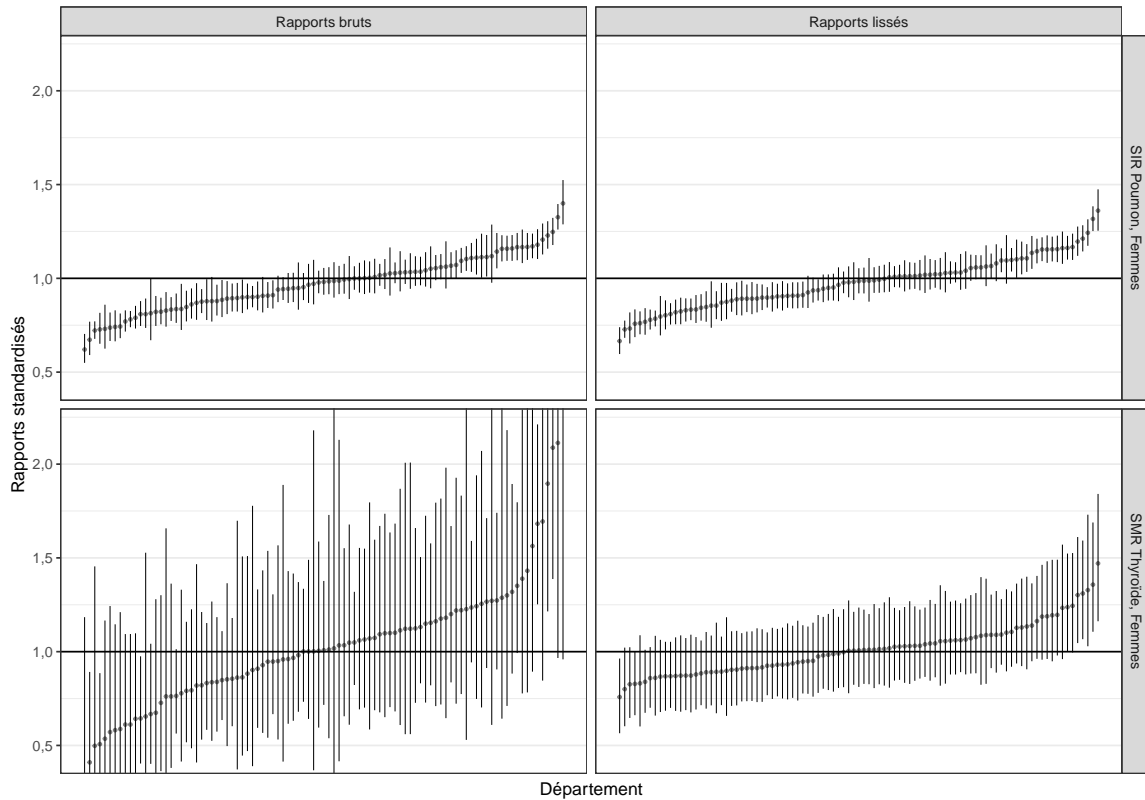
Dans ces graphiques, chaque point représente un rapport standardisé d'incidence ou de mortalité, soit issu des données brutes (mortalité observée ou incidence prédite - points rouges), soit estimé par le modèle de lissage (triangles bleus). Les points d'un même département sont reliés par un trait.

En utilisant l'information de l'ensemble des départements pour estimer le rapport standardisé d'un département, le lissage permet de stabiliser les estimations (i.e. d'améliorer leur précision). Les points bleus sont de ce fait situés à droite des points rouges sur l'axe des x (meilleures précisions).

Pour le cancer du poumon chez les femmes, les estimations départementales de l'incidence sont précises (le nombre de cas est important et la méthode de prédiction fiable - A++). Les estimations issues du modèle de lissage s'écartent peu des prédictions brutes de l'incidence (proximité des points sur l'axe y), ce d'autant moins que les prédictions sont précises (précision élevée, à droite de l'axe des x). L'effet du lissage est donc peu marqué.

À l'inverse, les décès par cancer de la thyroïde chez la femme sont rares (246 cas par an en France en moyenne entre 2007 et 2014), et les estimations des rapports de mortalités par départements très imprécises (points rouges à gauche de l'axe des x). L'effet du lissage est plus marqué dans ce cas, avec des estimations issues du modèle de lissage pour les SMR les plus extrêmes qui se rapprochent de la moyenne nationale (SMR=1).

(b) Rapports (SIR,SMR) et intervalles d'incertitude



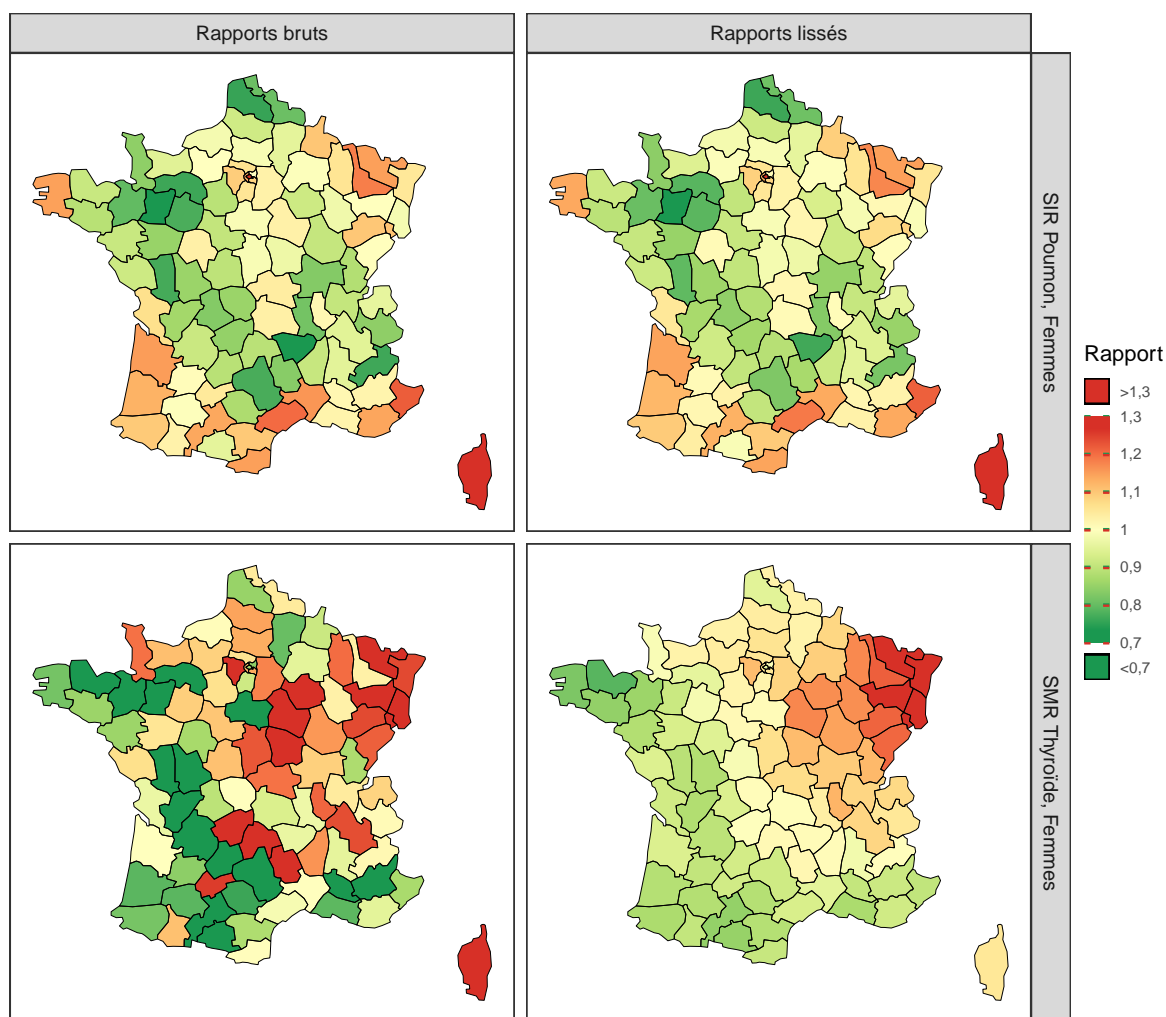
Note de lecture

Dans ces graphiques, chaque point représente un rapport standardisé d'incidence ou de mortalité dans un département, et les rapports sont représentés en ordre croissant. Les lignes verticales représentent les intervalles de confiance et de prédictions à 95%.

Pour le cancer du poumon chez les femmes, l'amplitude des variations des SIR sont comparables entre les prédictions brutes (entre 0,62 et 1,39) et les prédictions lissées (entre 0,66 et 1,35). De même, la taille des intervalles de prédiction à 95% varie peu entre les prédictions brutes et lissés. On retrouve là le fait que pour cette localisation, les prédictions sont précises, et donc l'effet du lissage peu marqué.

À l'inverse, pour les décès par cancer de la thyroïde chez la femme, les SMR bruts varient de manière importante (entre 0,33 et 2,39), mais les estimations sont très instables, avec des intervalles de confiance très larges. Le lissage permet dans ce cas de stabiliser les SMR, qui sont ramenés vers la moyenne des départements voisins et la moyenne nationale (SMR entre 0,75 et 1,47), avec des intervalles de confiance bien plus étroits.

(c) Cartes des rapports standardisés d'incidence et de mortalité



Note de lecture

Ces cartes représentent les rapports standardisés d'incidence du cancer du poumon chez la femme, et rapports standardisés de mortalité par cancer de la thyroïde chez la femme, avant et après lissage. La couleur verte est utilisée pour les départements en sous-incidence ($SIR < 1$) ou sous-mortalité ($SMR < 1$) par rapport à la moyenne nationale, la couleur rouge pour les départements en sur-incidence ($SIR > 1$) ou surmortalité ($SMR > 1$). L'intensité de la couleur varie entre ces teintes de façon continue, les départements avec une incidence ou une mortalité proche du niveau national (SIR et $SMR = 1$) étant en jaune.

Les différences entre les cartes d'incidence brute et lissée du cancer du poumon chez la femme sont difficilement visibles. Les deux cartes montrent que l'incidence de ce cancer est globalement plus élevée dans le sud de la France, en Corse et à Paris, alors que des zones de sous-incidences sont visibles dans le Nord, et sur une ligne allant de la Normandie au sud du Massif central.

La carte brute des décès par cancer de la thyroïde chez la femme est très contrastée, avec des zones de sur et de sous-mortalité respectivement visibles dans l'est et l'ouest de la France. Des gradients ou des zones homogènes sont cependant difficiles à appréhender sur cette carte, des départements avec une sur-mortalité importante côtoyant des départements avec une sous-mortalité importante. Les SMR étant très variables pour cette localisation, il est difficile de voir sur cette carte si ces contrastes sont caractéristiques d'une vraie variation de l'incidence sur le territoire, ou s'ils sont la résultante d'aléas statistiques. La carte lissée confirme en revanche, de manière claire, un gradient ouest-est de la mortalité, avec une structuration spatiale forte.

4. CONTENU DES PROFILS RÉGIONAUX

4.1 Indicateurs présentés dans les profils

Les prédictions d'incidence portaient sur la période 2007-2016, l'incidence et la mortalité observées sur la période 2007-2014.

Taux d'incidence et de mortalité standardisés

Les taux standardisés d'incidence et de mortalité étaient calculés en utilisant la population mondiale de l'Organisation mondiale de la santé (OMS) de 1960 comme population de référence [20]. Ils correspondent aux taux d'incidence et de mortalité que l'on observerait dans cette population type, si elle était soumise à la force d'incidence et de mortalité observée. Ils sont exprimés pour 100 000 personnes-années.

Rapports standardisés d'incidence (SIR) et de mortalité (SMR)

Le SIR (respectivement SMR), dans une zone géographique donnée (département ou région), est le rapport entre le nombre de cas incidents estimés (resp. de décès) et le nombre attendu de cas (resp. de décès) si les taux d'incidence (resp. de mortalité) par âge dans cette zone géographique étaient identiques à ceux de la France entière. Par exemple, un SIR de 1,2 dans un département représente 20 % de cas supplémentaires par rapport à ce qu'on attendrait si les taux d'incidence par âge du département étaient égaux aux taux nationaux. Au contraire, un SIR de 0,8 représente 20 % de cas en moins.

Pour les localisations non éligibles aux prédictions d'incidence, le nombre de cas attendu dans un département était calculé en utilisant les taux d'incidence par âge de l'ensemble des départements couverts par un registre. Les SIR s'interprètent donc dans ce cas comme des écarts à l'ensemble des départements couverts par un registre.

Intérêt et limites des deux méthodes de standardisation

Les deux méthodes de standardisation : directe (Taux standardisés) et indirecte (SIR/SMR), fournissent des indicateurs synthétiques qui permettent de résumer les taux d'incidence et de la mortalité par âges observés dans une population. Ces deux résumés sont cependant de nature différente, et peuvent parfois donner des résultats discordants [21]. Ainsi, les taux standardisés présentés dans cette étude ont été calculés en utilisant la population standard de l'OMS comme référence, cette population étant la référence habituellement utilisée dans la production d'indicateurs d'incidence et de mortalité pour les cancers dans le monde [22]. Il est à noter cependant que la structure de la population mondiale est beaucoup plus jeune que celle de la population française. Aussi, les Taux standardisés monde (TSM) auront tendance à refléter davantage les taux observés dans les classes d'âges les plus jeunes que ceux observés dans les classes d'âges les plus âgées. Dans les cas (relativement rares) où les taux par âge des populations comparées se croisent, c'est-à-dire que les taux dans une des populations sont plus élevés chez les jeunes et inversement chez les personnes âgées, cet indicateur sera plus conséquent dans la population présentant les taux élevés aux âges jeunes. Dans le cas de la standardisation indirecte, le calcul des SIR et des SMR conserve les structures par âge des populations et mesure la différence relative d'incidence et de mortalité entre la France métropolitaine et le territoires étudié (département, région), étant donnée la répartition par âge de la population du territoire. La standardisation indirecte est particulièrement adaptée pour comparer la situation dans chacune des populations à celle de la population de référence.

Intervalles de confiance, de prédiction et de crédibilité

Les intervalles de confiance pour les taux de mortalité étaient calculés en utilisant la méthode de Fay et Feuer [23] ; les intervalles de confiance pour les effectifs et les taux bruts de mortalité en utilisant une loi de Poisson.

Pour les incidences prédites (nombre de cas, taux d'incidence bruts et standardisés), les intervalles de prédiction ont été calculés en supposant une distribution log-normale [1].

Le modèle de lissage présenté en section 3.2 fournit des estimations des SIR et des SMR, ainsi que leur distribution *a posteriori*. Les intervalles de confiance des estimations lissées des SIR et des SMR (appelés intervalles de crédibilité en statistique bayésienne) se déduisent directement de ces distributions.

Par souci de concision, nous utiliserons le terme intervalles de confiance pour désigner les intervalles de confiance, de prédiction et de crédibilité dans l'ensemble des profils.

4.2 Présentation des profils et guide de lecture

Les résultats sont présentés sous forme d'une synthèse régionale puis d'une fiche-type par localisation cancéreuse, comprenant une première partie intitulée "contexte national" décrivant des indicateurs sur la France métropolitaine, et une seconde partie intitulée "situation régionale et départementale" détaillant les indicateurs dans la région et ses départements.

Contexte national

Le contexte national comprend :

- Un tableau résumé présentant le nombre annuel moyen de cas prédit et de décès observés au niveau national, les taux d'incidence et de mortalité bruts nationaux, et les taux d'incidence et de mortalité standardisés sur la population mondiale. Ces indicateurs sont accompagnés de leur intervalle de confiance. L'étendue des variations des taux d'incidence et de mortalité standardisés départementaux (sous forme de 5^e et 95^e percentiles) est également reportée dans ce tableau ;
- Une carte des SIR et des SMR départementaux lissés. Lorsque la localisation cancéreuse n'est pas éligible, seuls les SIR observés dans les départements couverts par un registre de cancer sont reportés sur la carte ;
- Une figure représentant les SIR et les SMR lissés (axe des ordonnées) dans les départements de France métropolitaine (axe des abscisses), ordonnés par ordre croissant et accompagnés de leur intervalle de confiance à 95 %. Les départements de la région étudiée y sont représentés en surbrillance. Cette figure permet d'une part d'avoir une vision globale des variations géographiques¹ de l'incidence et de la mortalité et de leur précision, et d'autre part d'apprécier les différences des départements de la région entre eux et vis-à-vis des autres départements de France métropolitaine.

Situation régionale et départementale

Les résultats présentant la situation régionale et départementale comprennent :

- Un tableau résumé de la région présentant le nombre annuel moyen de cas prédits et de décès observés et les taux d'incidence et de mortalité standardisés sur la population mondiale dans la région et au niveau national. Les rapports standardisés d'incidence et de mortalité de la région sont également reportés. Ces indicateurs sont accompagnés de leur intervalle de confiance ;
- Un second tableau détaillant ces mêmes indicateurs par département de la région ;
- Une figure représentant les taux standardisés (axe des abscisses) dans les régions de France métropolitaine (axe des ordonnées), ordonnés par ordre croissant. Les intervalles de confiance entourent les estimations. Une ligne verticale est placée à la valeur du taux standardisé national. Cette figure permet d'avoir une vision des variations géographiques de l'incidence et de la mortalité entre les régions² et de la précision des estimations.

1. La même échelle a été utilisée pour toutes les localisations cancéreuses (SIR et SMR entre 0,5 et 2).

2. La même échelle de variations relatives a été utilisée pour toutes les localisations cancéreuses (taux compris entre 0,5 et 1,7 fois le taux de la France métropolitaine).

RÉFÉRENCES

- [1] Chatignoux, É., Remontet, L., Iwaz, J., Colonna, M. et Uhry, Z. "For a sound use of health care data in epidemiology : evaluation of a calibration model for count data with application to prediction of cancer incidence in areas without cancer registry". *Biostatistics* (mar. 2018).
- [2] Remontet, L., Mitton, N., Couris, C. M., Iwaz, J., Gomez, F. *et al.* "Is it possible to estimate the incidence of breast cancer from medico-administrative databases ?" *European Journal of Epidemiology* 23.10 (2008), p. 681-688.
- [3] Uhry, Z., Remontet, L., Grosclaude, P., Velten, M. et Colonna, M. "Estimations départementales de l'incidence du cancer colorectal en France à partir des données hospitalières, 1999-2003". *Revue d'Épidémiologie et de Santé Publique* 57.5 (oct. 2009), p. 329-336.
- [4] Mitton, N., Colonna, M., Trombert, B., Olive, F., Gomez, F. *et al.* "A Suitable Approach to Estimate Cancer Incidence in Area without Cancer Registry". *Journal of Cancer Epidemiology* 2011 (2011), p. 1-11.
- [5] Uhry, Z., Colonna, M., Remontet, L., Grosclaude, P., Carré, N. *et al.* "Estimating infra-national and national thyroid cancer incidence in France from cancer registries data and national hospital discharge database". *European Journal of Epidemiology* 22.9 (2007), p. 607-614.
- [6] Uhry, Z., Belot, A., Colonna, M., Bossard, N., Rogel, A. *et al.* "National cancer incidence is estimated using the incidence/mortality ratio in countries with local incidence data : Is this estimation correct ?" *Cancer Epidemiology* 37.3 (2013), p. 270-277.
- [7] Bossard, N., Gomez, F., Remontet, L. et Uhry, Z. "Utilisation des données du PMSI pour estimer l'incidence des cancers en France à l'échelon infranational : la démarche du groupe Oncepi". *Bull Epidémiol Hebd* Numéro thématique (5-6) (2012), p. 77-80.
- [8] Colonna, M., Chatignoux, E., Remontet, L., Mitton, N., Belot, A. *et al.* *Estimation de l'incidence départementale des cancers en France métropolitaine 2008-2010. Étude à partir des données des registres des cancers du réseau Francim et des bases de données médico-administratives*. Saint-Maurice : Institut de veille sanitaire, 2015, 53 p.
- [9] Colonna, M., Mitton, N., Remontet, L., 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 Epidémiol Hebd* 43-44-45 (2013), p. 566-574.
- [10] Belot, A., Grosclaude, P., Bossard, N., Jouglu, E., Benhamou, E. *et al.* "Cancer incidence and mortality in France over the period 1980-2005". *Revue d'Épidémiologie et de Santé Publique* 56.3 (juin 2008), p. 159-175.
- [11] Rogel, A., Belot, A., Suzan, F., Bossard, N., Boussac, M. *et al.* "Reliability of recording uterine cancer in death certification in France and age-specific proportions of deaths from cervix and corpus uteri." *Cancer epidemiology* 35.3 (juin 2011), p. 243-9.
- [12] Wakefield, J. "Disease mapping and spatial regression with count data". *Biostatistics* 8.2 (2007), p. 158-183.
- [13] Spiegelhalter, D. J. "Funnel plots for comparing institutional performance". *Statistics in Medicine* 24.8 (2005), p. 1185-1202.
- [14] Spiegelhalter, D. J. "Handling over-dispersion of performance indicators". *Quality and Safety in Health Care* 14.5 (2005), p. 347-351.
- [15] Gorla, S. et Le Tertre, A. *Les études locales autour d'un point source. Note méthodologique*. Saint-Maurice : Institut de veille sanitaire, 2010, 8 p.
- [16] David, S., Remontet, L., Bouvier, A. M., Faivre, J., Colonna, M. *et al.* "[How to choose in practice a model to describe the geographic variation of cancer incidence ? Example of gastrointestinal cancers from Côte-d'Or]." *Revue d'Épidémiologie et de santé publique* 50.5 (oct. 2002), p. 413-25.
- [17] Besag, J., York, J. et Mollié, A. "Bayesian image restoration, with two applications in spatial statistics". *Annals of the Institute of Statistical Mathematics* 43.1 (1991), p. 1-20.
- [18] Blangiardo, M. et Cameletti, M. *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley, 2015.
- [19] Rue, H., Martino, S. et Chopin, N. "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations". *Journal of the Royal Statistical Society. Series B : Statistical Methodology* 71.2 (avr. 2009), p. 319-392.
- [20] Ahmad, O. B., Boschi-Pinto, C., Lopez, A. D., Murray, C. J., Lozano, R. *et al.* *Age standardization of rates : a new WHO standard*. Geneva : World Health Organization, 2001. URL : <https://seer.cancer.gov/stdpopulations/>.
- [21] Bouyer, J. *Epidémiologie : principes et méthodes quantitatives*. Lavoisier, 2009, 498 p.
- [22] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C. *et al.* "Cancer incidence and mortality worldwide : sources, methods and major patterns in GLOBOCAN 2012." *International journal of cancer* 136.5 (mar. 2015), E359-86.
- [23] Fay, M. P. et Feuer, E. J. "Confidence intervals for directly standardized rates : A method based on the gamma distribution". *Statistics in Medicine* 16.7 (1997), p. 791-801.

Chatignoux É, Remontet L, Colonna M, Grosclaude P, Cariou M, Billot-Grasset A, Uhry Z.
*Estimations régionales et départementales d'incidence et de mortalité par cancers en France,
2007-2016. Matériel et méthodes.* Saint-Maurice : Santé publique France, 2019. 18 p.

Disponible à partir des URL :

<https://geodes.santepubliquefrance.fr>

<http://lesdonnees.e-cancer.fr/>

<http://www.santepubliquefrance.fr/>

<https://www.e-cancer.fr/>

ISBN-NET : 979-10-289-0516-3

ISSN : 1956-6964

RÉALISÉ PAR LA DIRECTION DE LA COMMUNICATION, SANTÉ PUBLIQUE FRANCE

Tous droits réservés - Siren : 130 022 338

DÉPÔT LÉGAL : JANVIER 2019