

Estimations départementales d'incidence des cancers à partir du SNDS et des registres

Édouard Chatignoux¹ Laurent Remonet² Marc Colonna^{3,4} Zoé Uhry^{1,2}

¹Santé publique France

²Service de biostatistique,
Hospices civils de Lyon

³Réseau français
des registres de cancer Francim

⁴Registre du cancer
de l'Isère

Congrès SFSPM, 10 novembre 2021



Déroulé de la présentation

La surveillance des cancers en France

Les registres des cancers

Données non spécifiques pour la surveillance

Méthode

Application

Discussion

Déroulé de la présentation

La surveillance des cancers en France

Les registres des cancers

Données non spécifiques pour la surveillance

Méthode

Application

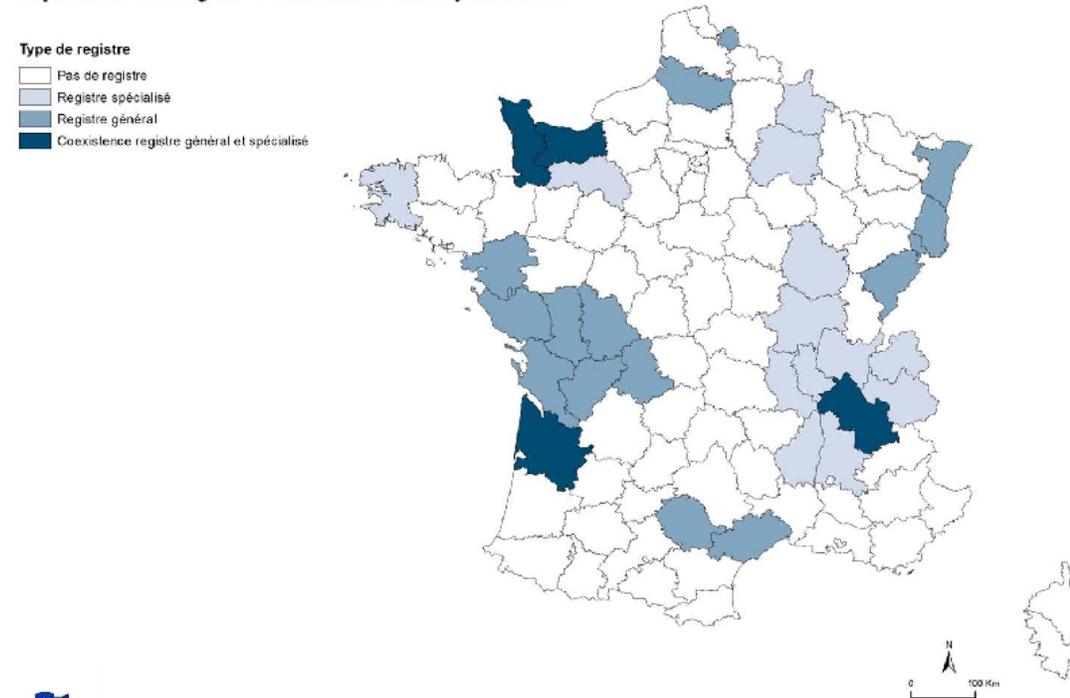
Discussion

Les registres des cancers

Objectifs des registres

Répertorier tous les nouveaux cas de cancer diagnostiqués parmi les habitants d'une zone géographique donnée pendant une période donnée quel que soit le lieu de prise en charge diagnostique et thérapeutique

Répartition des registres de cancers au 1er janvier 2014



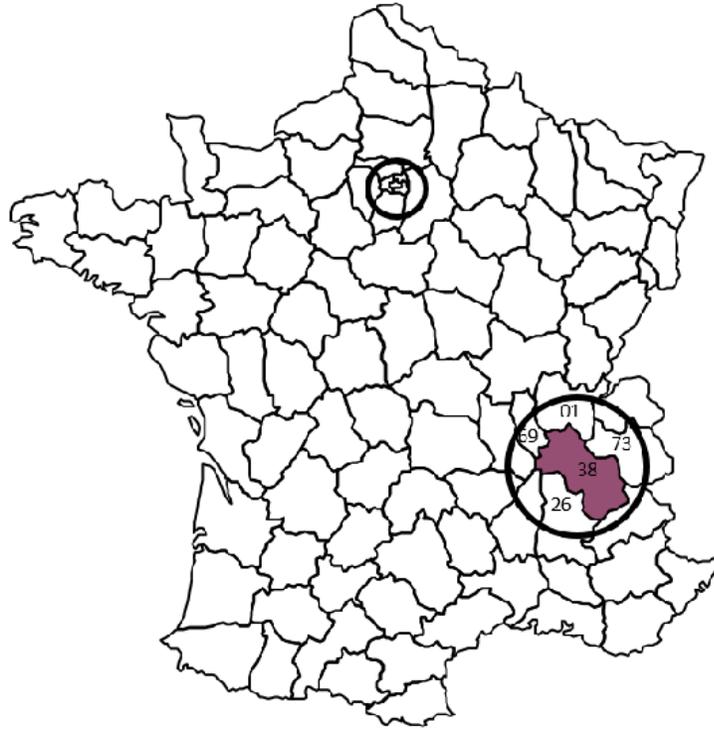
Sources : INVS, Mars 2014 (fond de carte : IGN, GéoFile 2008)

- 14 registres généraux, 10 spécialisés, 2 nationaux pédiatriques
- Couverture \simeq 20% de la population
- **Gold standard**

Les registres des cancers

Recherche active de cas, système multi-sources

Exemple du registre du cancer du département de l'Isère



- Existence d'autres établissements en dehors de Rhône-Alpes : IGR, Institut Curie, ...
- Echange entre registres car sources info communes

Secteur administratif de la Source Information

- Secteur public
- Secteur privé

Type de données des Sources Information

- Base administrative (ALD)
- Base médico-administrative (PMSI)
- Base anatomie-pathologique
- Base clinique : dossier médical

Mode de recueil des données

- Passif (réception de l'information)
- Actif (déplacement du personnel)

Isère, 2014 :

-> 7 300 cas de cancer invasif

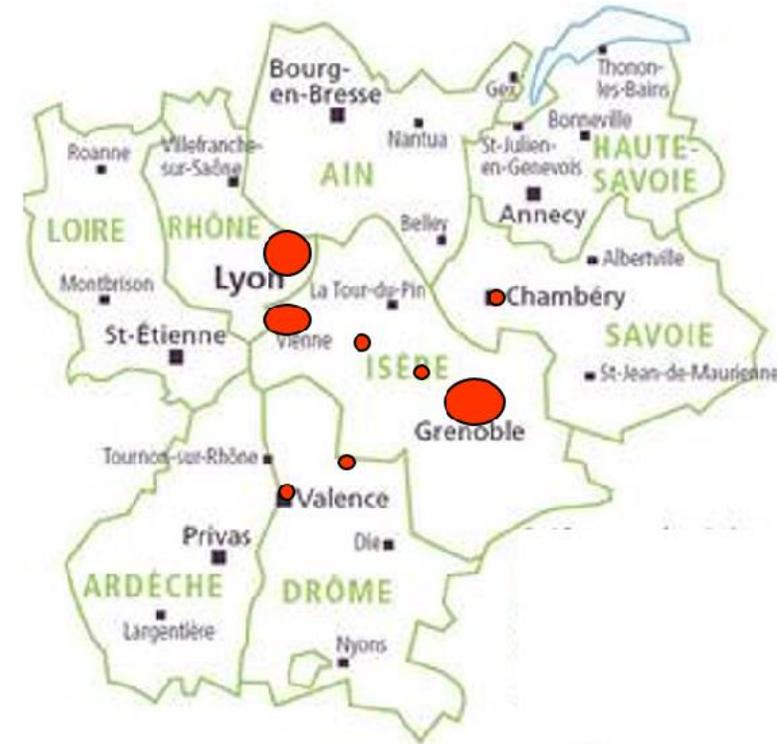
-> 1 100 cas de cancer in situ

Les registres des cancers

Illustration : registre de l'Isère

84 sources d'information (en 2010)

- 26 laboratoires d'anatomopathologie
- 45 établissements de santé
- Cabinets médicaux
- 3 Caisses d'assurance maladie
- 1 Réseau régional de de cancérologie
- 2 Registres nationaux de l'enfant



Les registres des cancers

Un gold standard

- Notification + recherche active de cas → exhaustivité
- Croisement de plusieurs sources + validation des cas
- Codage spécifique et précis (topologie, morphologie... → CIM-O-3)
- Information détaillée sur les cas

Insuffisant pour répondre à tous les besoins de surveillance

- Couverture non exhaustive
- Inadapté à la surveillance spatiale de l'incidence
 - Variation de l'incidence sur tout le territoire

⇒ Nécessité d'avoir recours à une information externe (indicateurs proxys de l'incidence)

- Tout indicateur corrélé à l'incidence (big data ?)
- Données médico-administratives et mortalité (SNDS)

Déroulé de la présentation

La surveillance des cancers en France

Les registres des cancers

Données non spécifiques pour la surveillance

Méthode

Application

Discussion

Données non spécifiques pour la surveillance

Le système national des données de Santé - SNDS

Identifiant anonyme unique

Données individuelles de l'Assurance Maladie (depuis 2006)

Informations

sociodémographiques Mois et année de naissance, CMUc, régime d'affiliation
Commune de résidence
Niveau de défaveur social de la commune de résidence

Consultations et actes Date de soins
Type de soins (infirmier, médecins...)

Médicaments :

Nom, date de délivrance, nombre de boîtes, dose

Acte de biologie

date, type

Affection longue durée (ALD)

Date de décès (Insee)

PMSI 2006-2018

Séjours hospitaliers :

Date d'entrée
Date de sortie
Durée d'hospitalisation
Mode d'entrée (domicile, transfert...)
Mode de sortie (décès, domicile...)

Diagnostics hospitaliers (principal, relié, associé)

Actes classants

Causes médicales de décès 2006-2016

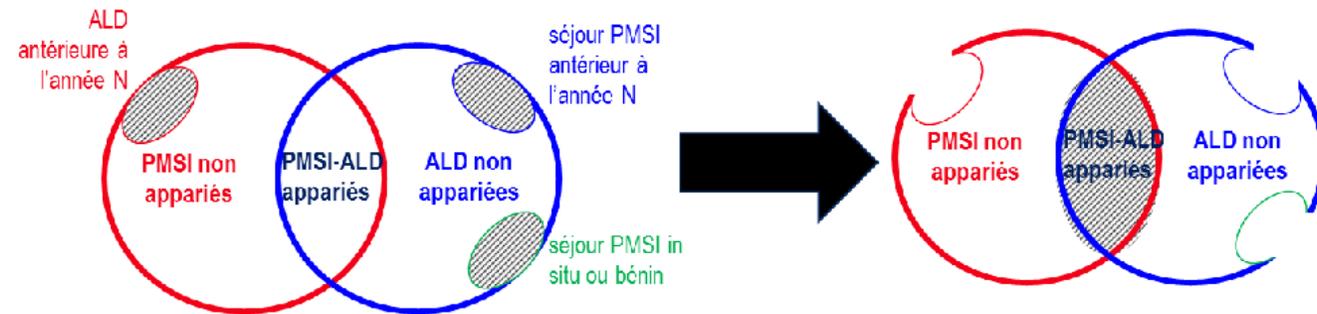
Certificat de décès :

Cause initiale
Cause associée
Age au décès
Code postal de domicile

Indicateurs SNDS

Proxys de l'incidence (i.e. nombre de nouveaux cas de cancers invasifs survenus chez les habitants d'un territoire pour une période donnée).

- Mortalité : nombre de décès (cause principale)
- ALD : nombre de nouvelles mises en ALD
- PMSI : nombre d'hospitalisations incidentes (sélection en DP, élimination des cas prévalents à N-1 et N-2 - DP, DR ou DAS)
- AUP : nombre de nouvelles mises en ALD ou d'hospitalisations incidentes

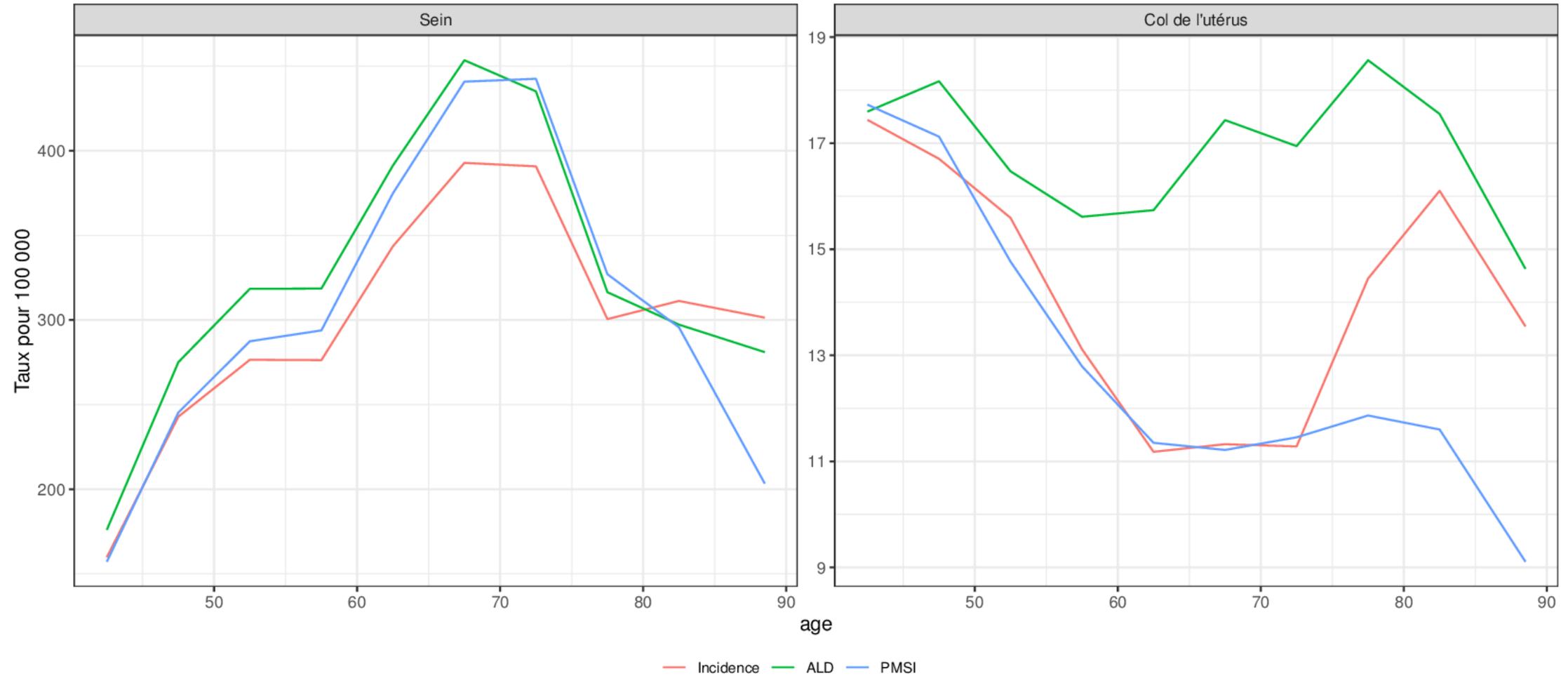


- Données (hors mortalité) produites dans une perspective de gestion/remboursement et non de surveillance
- Écarts avec l'incidence registre

Données non spécifiques pour la surveillance - illustration

Les proxys ne reflètent pas directement les niveaux d'incidence...

Taux d'incidence, de mise en ALD et d'hospitalisation pour les cancers du sein et du col de l'utérus, dans les départements couverts par des registres de cancer - période 2007-2015.

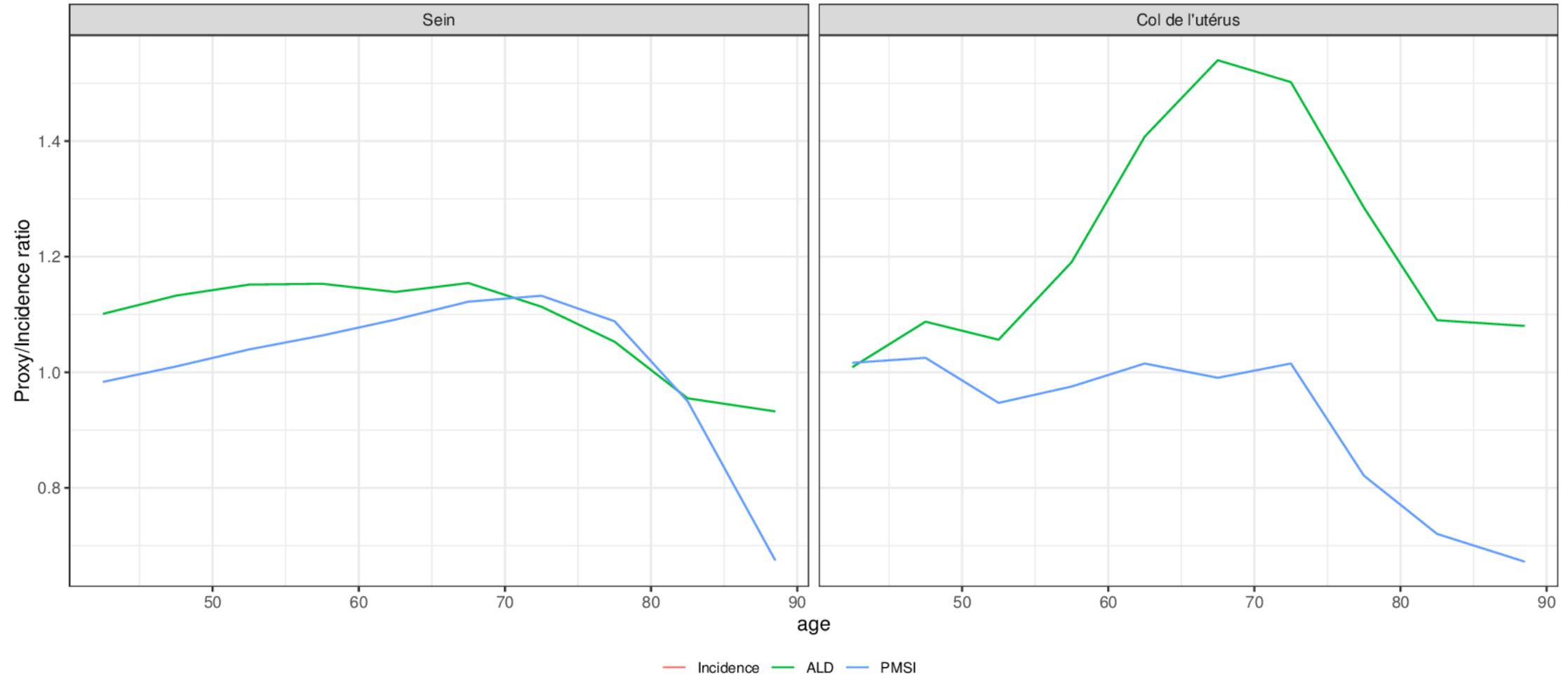


ALD, affections longue durées ; PMSI, hospitalisations

Données non spécifiques pour la surveillance - illustration

Les proxys ne reflètent pas directement les niveaux d'incidence...

Taux d'incidence, de mise en ALD et d'hospitalisation pour les cancers du sein et du col de l'utérus, dans les départements couverts par des registres de cancer - période 2007-2015.

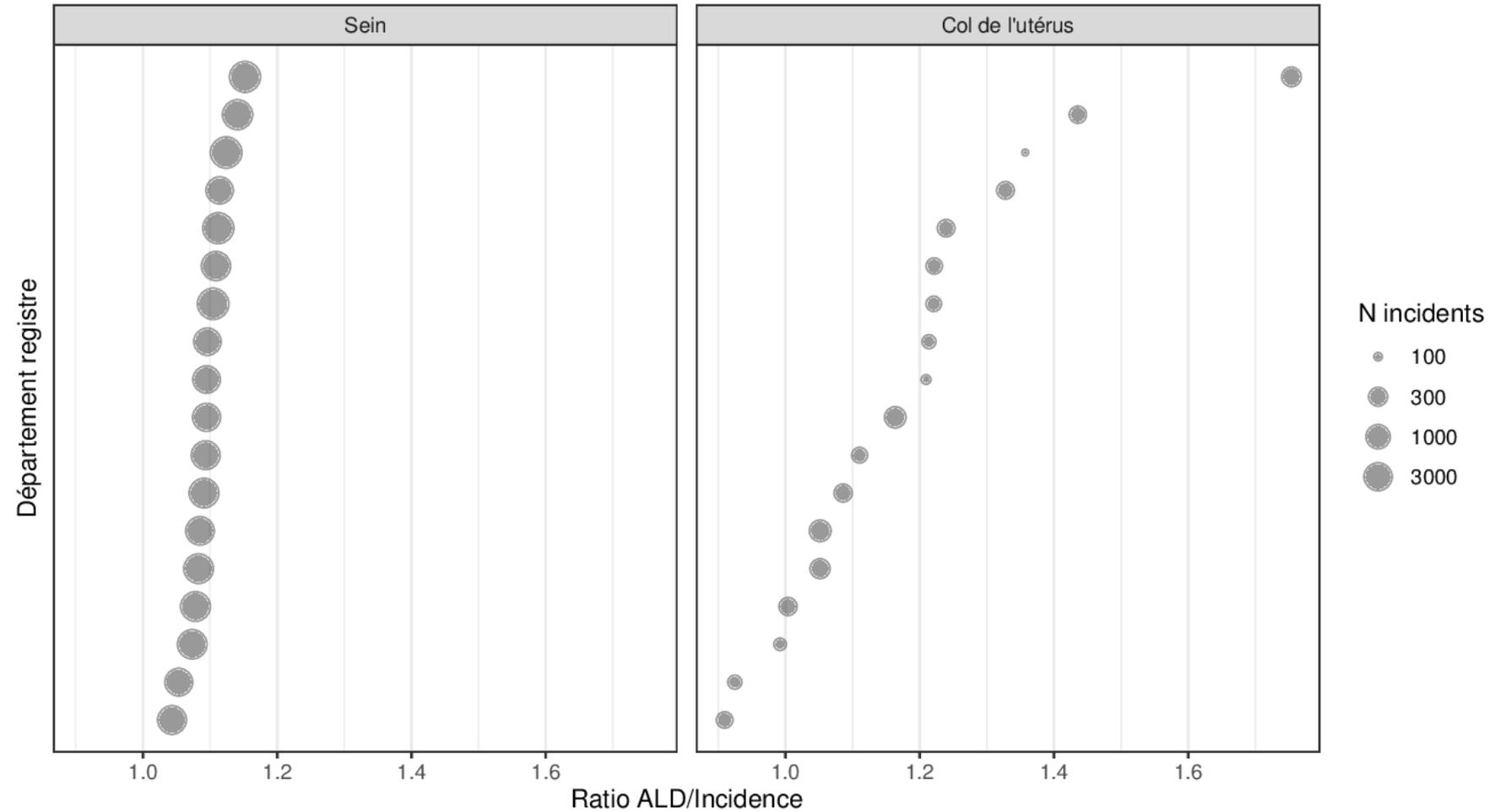


ALD, affections longue durées ; PMSI, hospitalisations

Données non spécifiques pour la surveillance - illustration

... et les différences varient selon les départements

Ratio entre le nombre total de cas incidents et de mise en ALD pour les cancers du sein et du col de l'utérus, par département couvert par un registres de cancer - période 2007-2015.



ALD, affections longue durées

Données disponibles

- Un gold standard : les données registres → dénombrement précis et spécifiques des cas de cancers sur ~20% de la population
- Des proxys de l'incidence : SNDS, mortalité, autres ?
 - ↪ Exhaustifs (100% de la population, tout le territoire)
 - ↪ Disponibles
 - ↪ Non spécifique de l'incidence, non dédiés à la surveillance

Utiliser les données registres pour corriger les données proxy

- Calibrer les proxys pour prédire l'incidence dans les zones non couvertes par un registre + mesurer les erreurs
- Un modèle développé à l'échelle départementale : le modèle ratio^{1 2}

1. Zoe UHRY 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.

2. Édouard CHATIGNOUX et al. "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". *eng. Biostatistics (Oxford, England)* 20.3 (juil. 2019), p. 452-467.

Déroulé de la présentation

La surveillance des cancers en France

Méthode

Application

Discussion

Principe de la méthode "ratio"

Pour une localisation cancéreuse, un sexe, un proxy P donné et un nombre de cancers incidents K , un département j :

Rationnel

1. On observe dans la zone registre, en moyenne, des écarts entre P et $K \rightarrow$ rapport $K/P \neq 1$
 \Rightarrow Calibration des proxys par ce rapport pour estimer l'incidence : $\hat{K}_j = (K/P)_{ZR} P_j$
2. Le rapport K/P varie d'un département à l'autre
 \Rightarrow Mesurer ces variations et en tenir compte dans le calcul des variances

Les départements couverts par des registres sont par définition les seuls départements où les cas de cancers incidents et le proxy est observé. Le rationnel ci-dessus est applicable aux départements hors zone registre si :

Hypothèses

1. Le rapport K/P mesuré dans les départements couverts par un registre mesure correctement le rapport K/P de la France entière
2. Les variations du rapport K/P entre les départements hors de la zone registre sont bien mesurées par les variations dans la zone registre

Validation

Les données des registres permettent de mesurer les écarts entre les incidences réelles (registres) et les incidences prédites avec la méthode ratio

Validation des prédictions

Les prédictions sont elles suffisamment précises pour être informatives ?

Critères

1. ERA : erreur de prédiction pour le nombre total de cas incidents τ_j

$$\rightarrow ERA_j = |\hat{\tau}_j^{CV} - K_j| / K_j .$$

Arbre de décision



Taille des erreurs (ERA)

Validation des prédictions

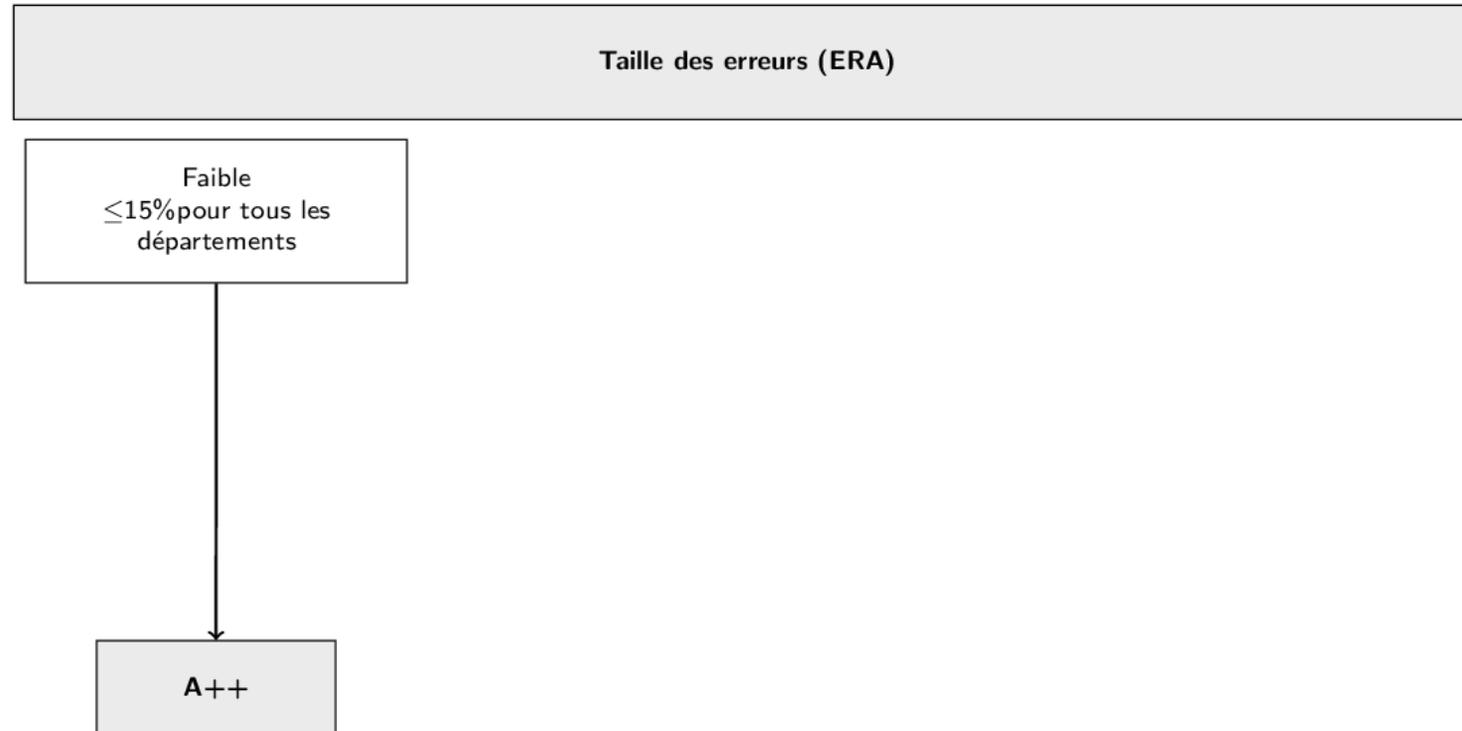
Les prédictions sont elles suffisamment précises pour être informatives ?

Critères

1. ERA : erreur de prédiction pour le nombre total de cas incidents τ_j

$$\rightarrow ERA_j = |\hat{\tau}_j^{CV} - K_j| / K_j .$$

Arbre de décision



Validation des prédictions

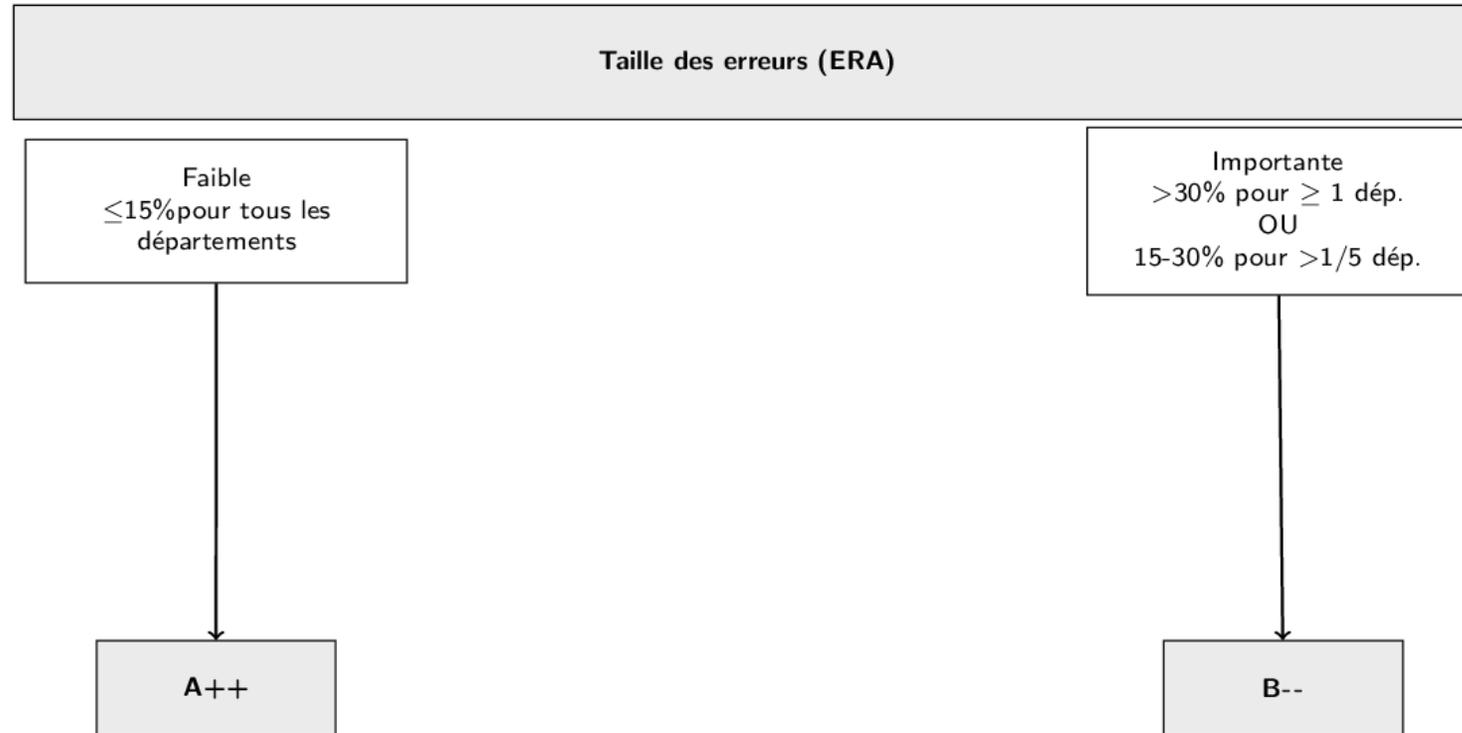
Les prédictions sont elles suffisamment précises pour être informatives ?

Critères

1. ERA : erreur de prédiction pour le nombre total de cas incidents τ_j

$$\rightarrow ERA_j = |\hat{\tau}_j^{CV} - K_j| / K_j .$$

Arbre de décision



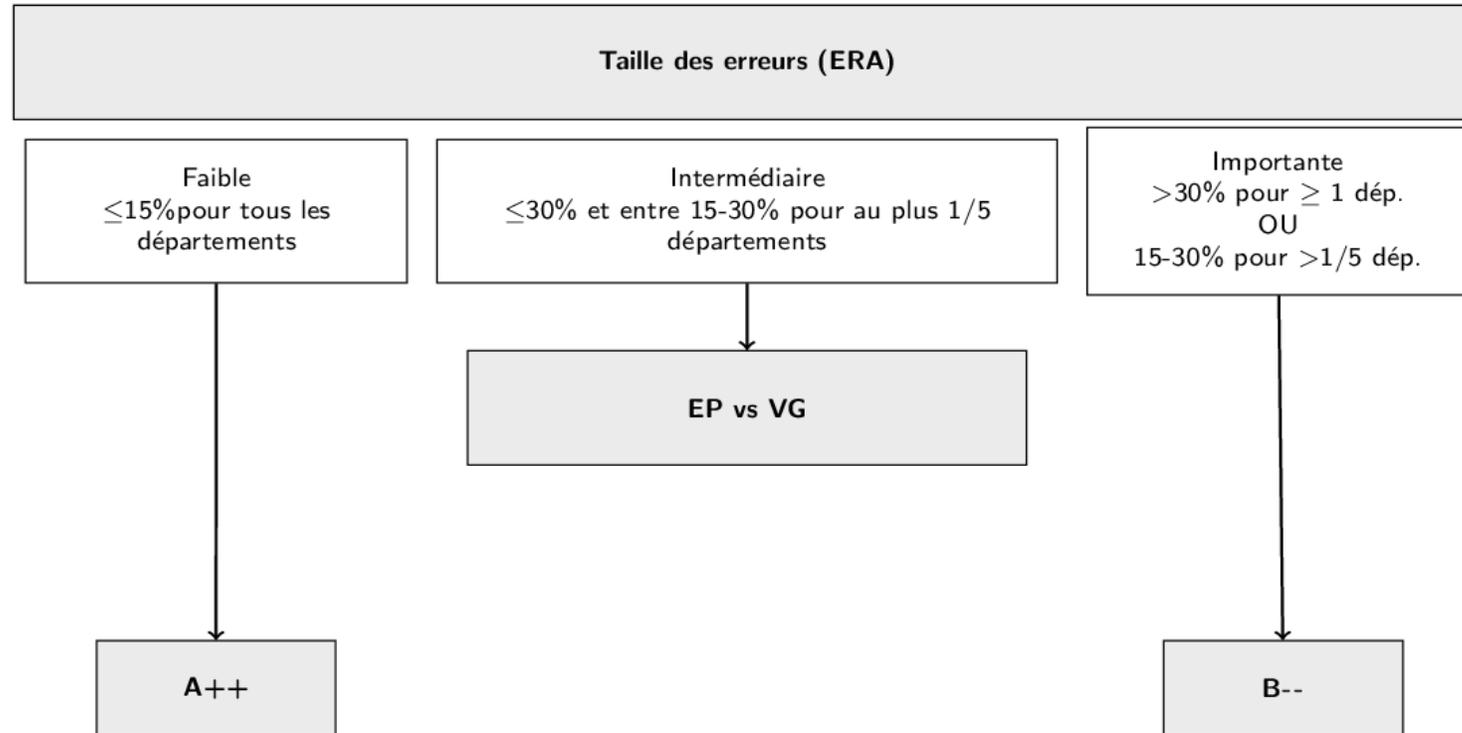
Validation des prédictions

Les prédictions sont elles suffisamment précises pour être informatives ?

Critères

1. ERA : erreur de prédiction pour le nombre total de cas incidents τ_j
→ $ERA_j = |\hat{\tau}_j^{CV} - K_j| / K_j$.
2. Importance des erreurs de prédiction (EP) vs variabilité géographique de l'incidence (VG)
→ On peut accepter des erreurs plus importantes si l'incidence vraie fortement sur le territoire

Arbre de décision



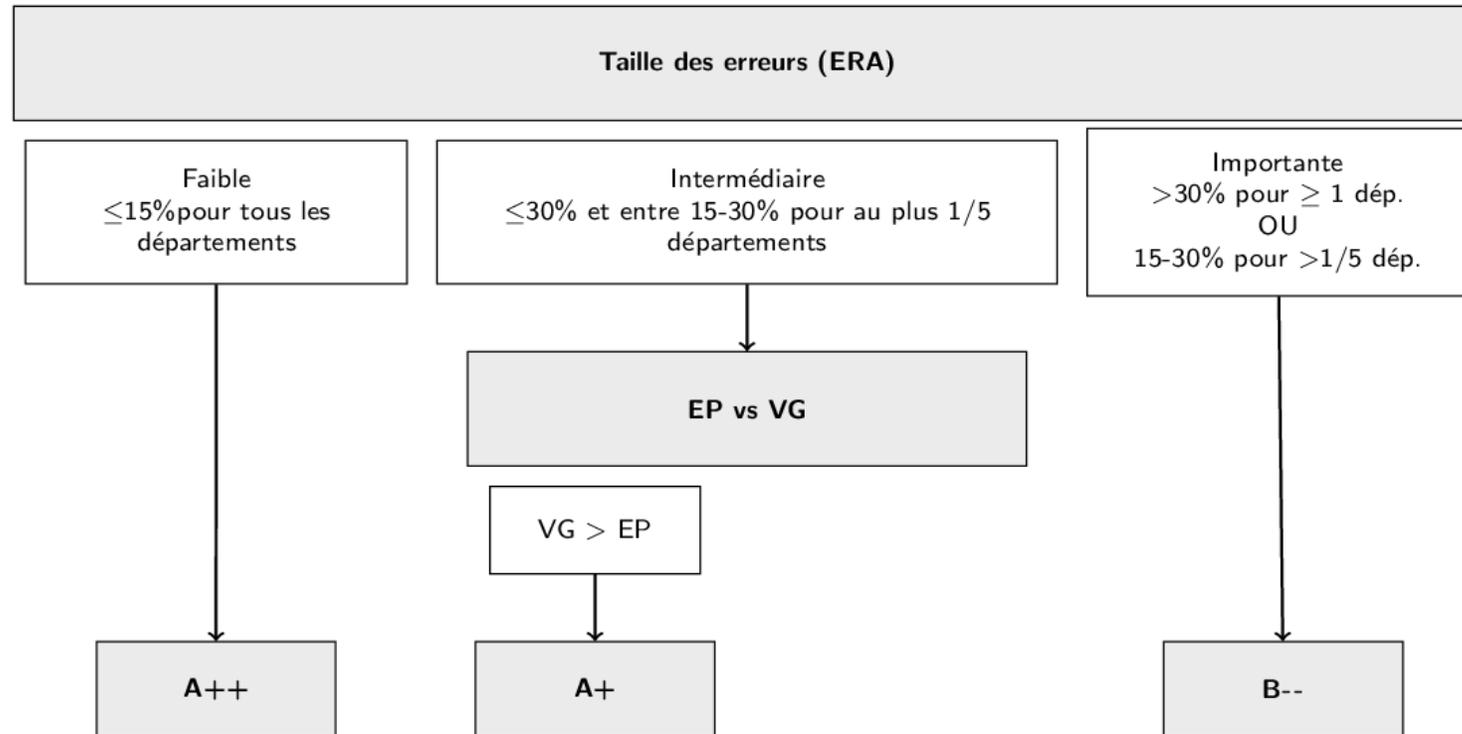
Validation des prédictions

Les prédictions sont elles suffisamment précises pour être informatives ?

Critères

1. ERA : erreur de prédiction pour le nombre total de cas incidents τ_j
→ $ERA_j = |\hat{\tau}_j^{CV} - K_j| / K_j$.
2. Importance des erreurs de prédiction (EP) vs variabilité géographique de l'incidence (VG)
→ On peut accepter des erreurs plus importantes si l'incidence vraie fortement sur le territoire

Arbre de décision



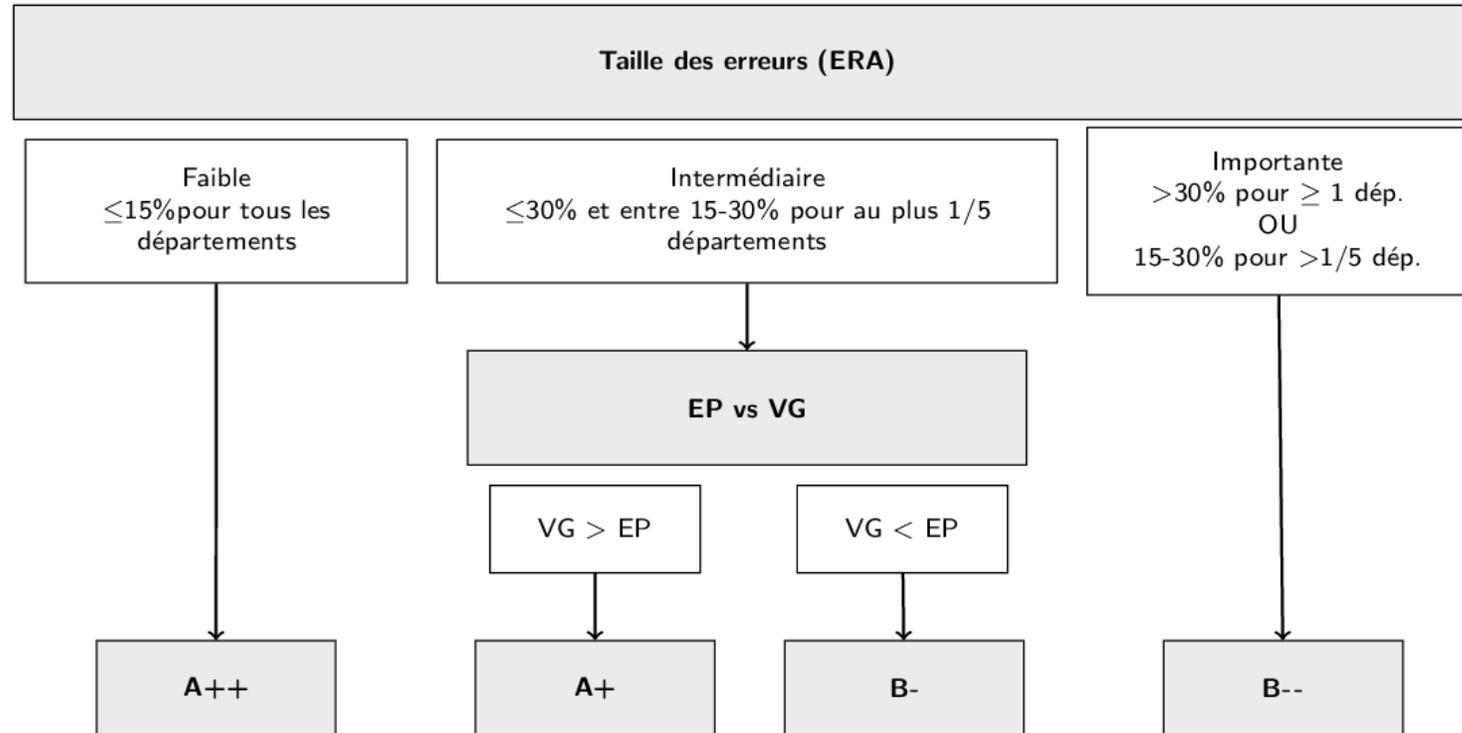
Validation des prédictions

Les prédictions sont elles suffisamment précises pour être informatives ?

Critères

1. ERA : erreur de prédiction pour le nombre total de cas incidents τ_j
→ $ERA_j = |\hat{\tau}_j^{CV} - K_j| / K_j$.
2. Importance des erreurs de prédiction (EP) vs variabilité géographique de l'incidence (VG)
→ On peut accepter des erreurs plus importantes si l'incidence vraie fortement sur le territoire

Arbre de décision



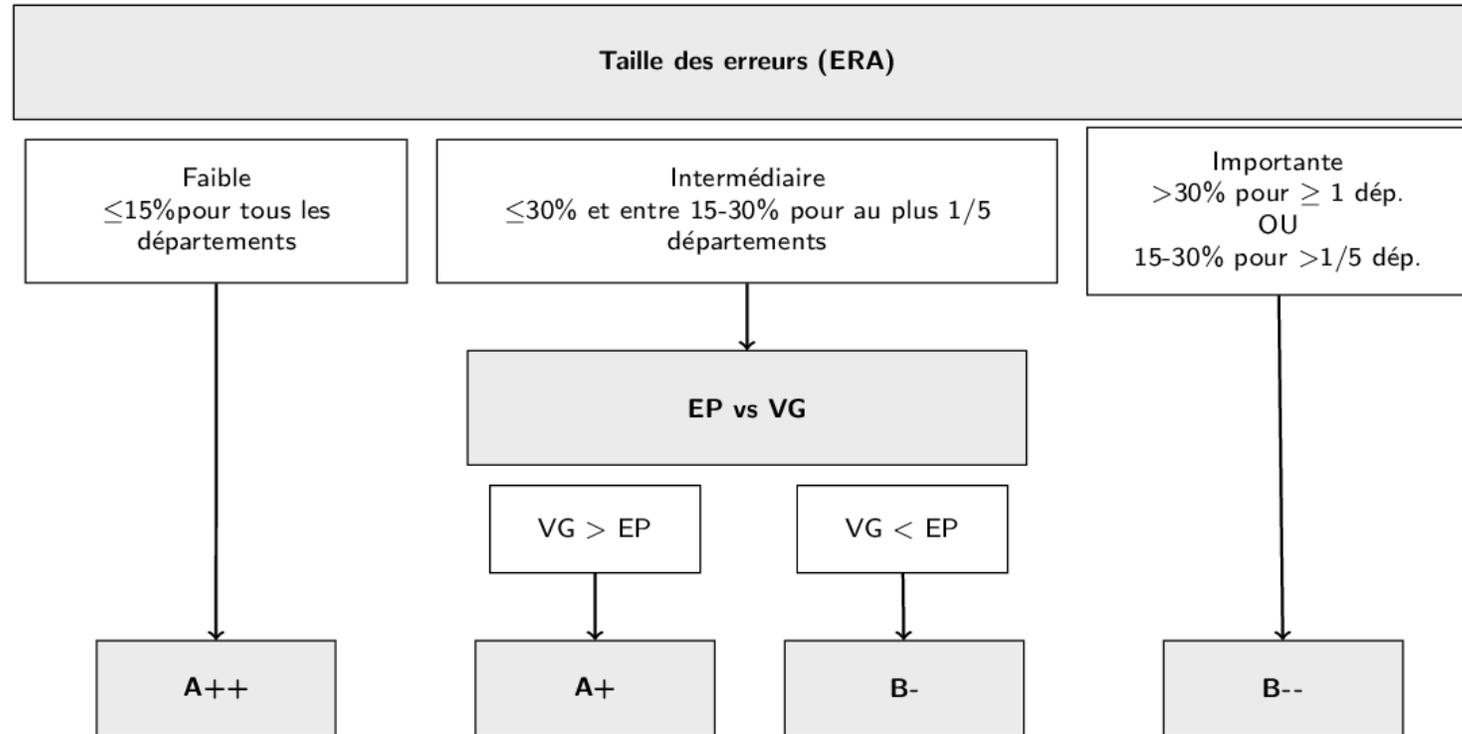
Validation des prédictions

Les prédictions sont elles suffisamment précises pour être informatives ?

Critères

1. ERA : erreur de prédiction pour le nombre total de cas incidents τ_j
→ $ERA_j = |\hat{\tau}_j^{CV} - K_j| / K_j$.
2. Importance des erreurs de prédiction (EP) vs variabilité géographique de l'incidence (VG)
→ On peut accepter des erreurs plus importantes si l'incidence vraie fortement sur le territoire

Arbre de décision



- $A++ > A+ > B- > B--$
- Prédictions intéressantes si A

Déroulé de la présentation

La surveillance des cancers en France

Méthode

Application

Discussion

Matériel et méthodes¹

Application de la méthode pour 3 localisations cancéreuses, sein, col de l'utérus et côlon rectum, période 2007-2015

Données

- Registres
 - Cancer du sein et du col de l'utérus : 14 registres généraux + 1 registre spécialisé \Rightarrow 18 départements couverts
 - Cancer du côlon-rectum-anus : 12 registres généraux + 3 registres spécialisés \Rightarrow 20 départements couverts
- Proxys
 - \hookrightarrow ALD, PMSI, AUP, DC issus du SNDS
- Codes CIM-10/CIM-O-3
 - Cancer du sein : C50
 - Cancer du col de l'utérus : C53
 - Cancer du colon-rectum-anus : C18-C21

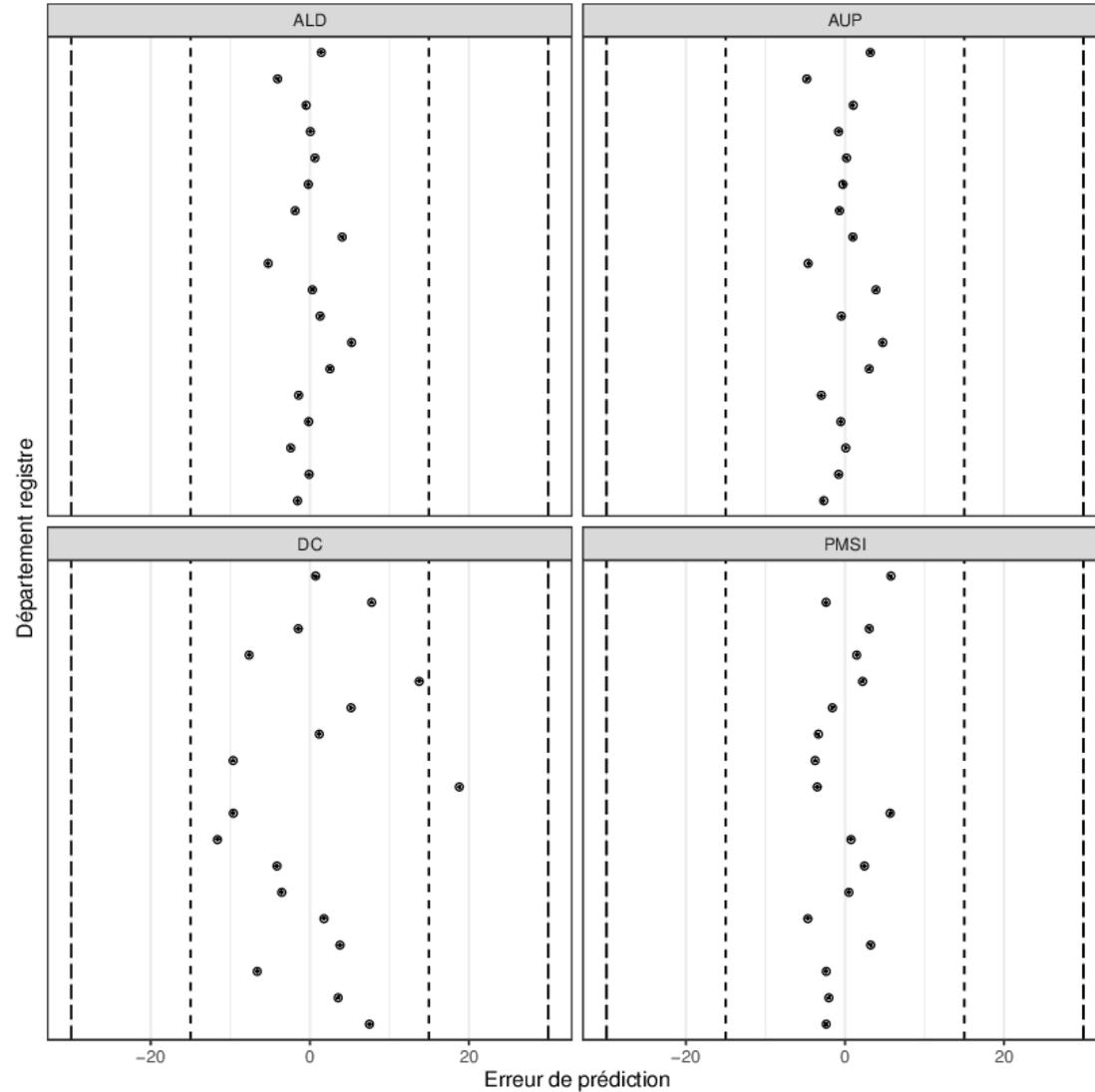
Ratios standardisés d'incidence - SIR

- Dans un département, rapport entre le nombre de cas prédit et celui attendus si les taux d'incidence par âge étaient identiques à ceux de la France entière
- \leftrightarrow SIR de 1,2 \simeq + 20% de cas par rapport à la moyenne nationale

1. [Edouard CHATIGNOUX et al.](#) "How to produce sound predictions of incidence at a district level using either health care or mortality data in the absence of a national registry : the example of cancer in France". *eng. International Journal of Epidemiology* (nov. 2020).

Cancer du sein

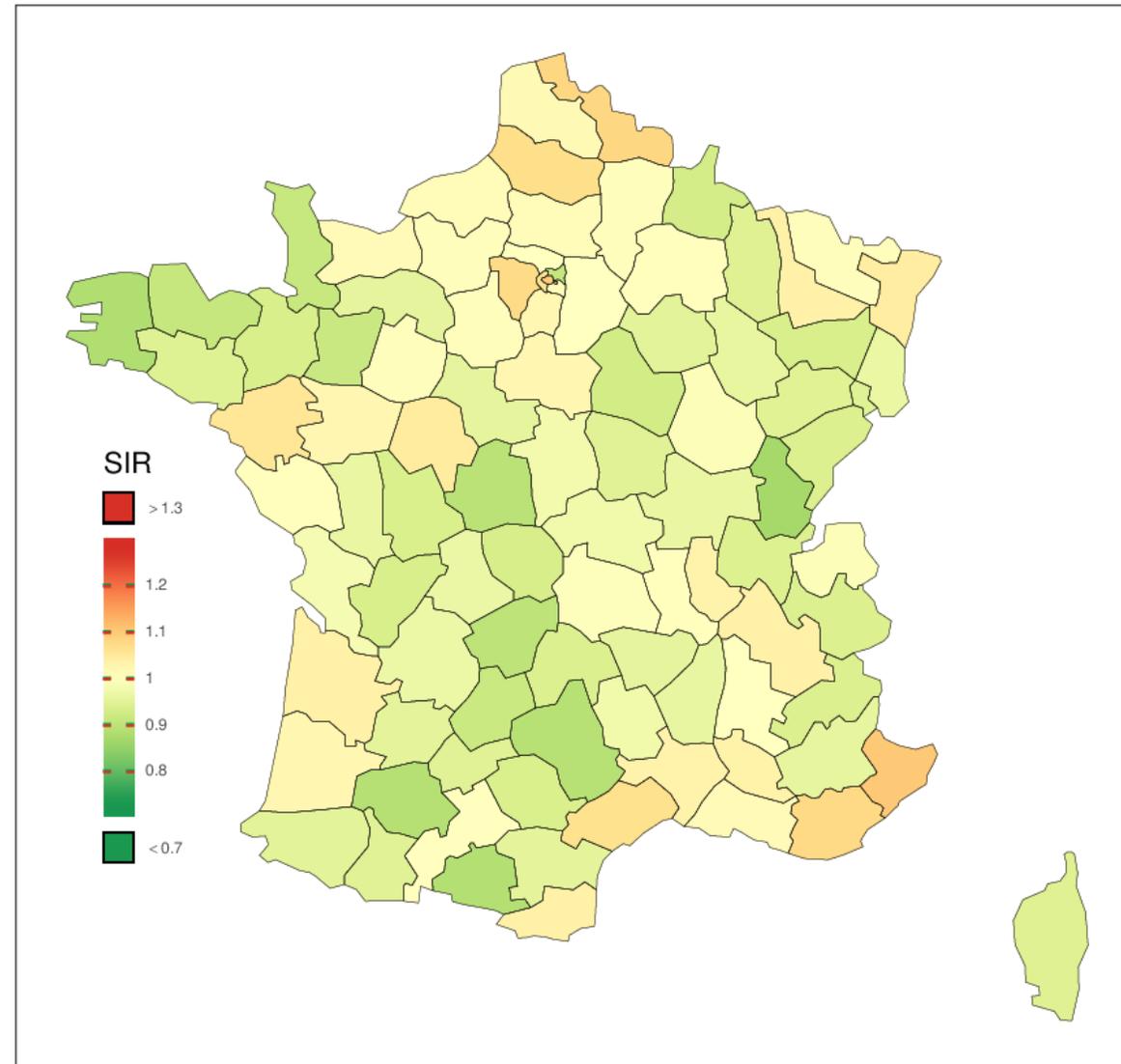
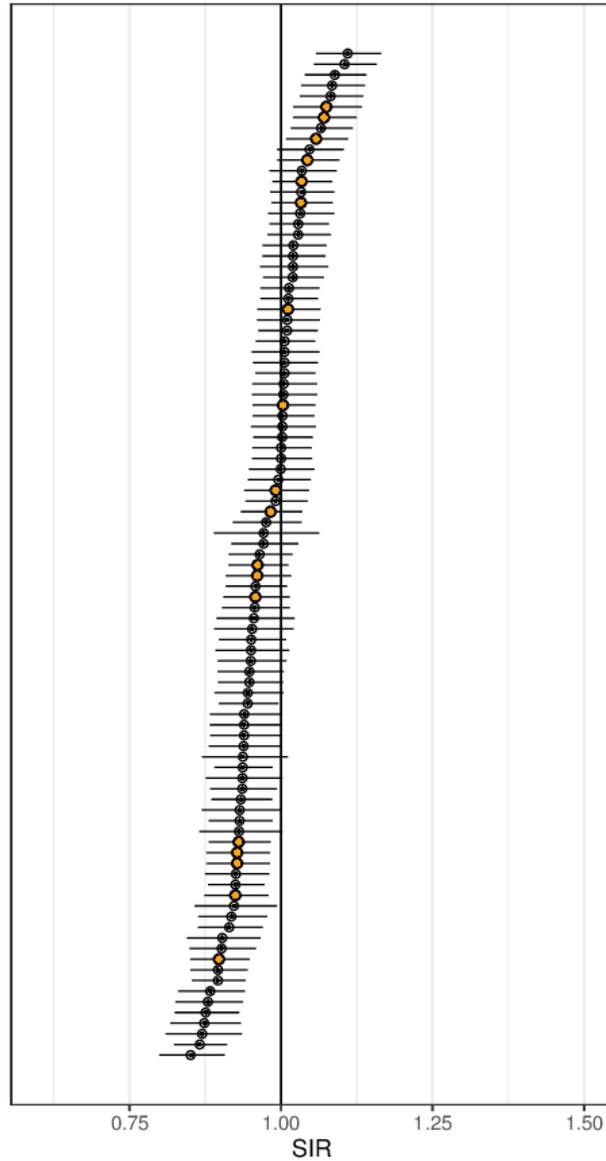
Validité des prédictions



- Faibles erreurs pour tous les proxys
- Prédictions à partir du proxy ALD

Cancer du sein

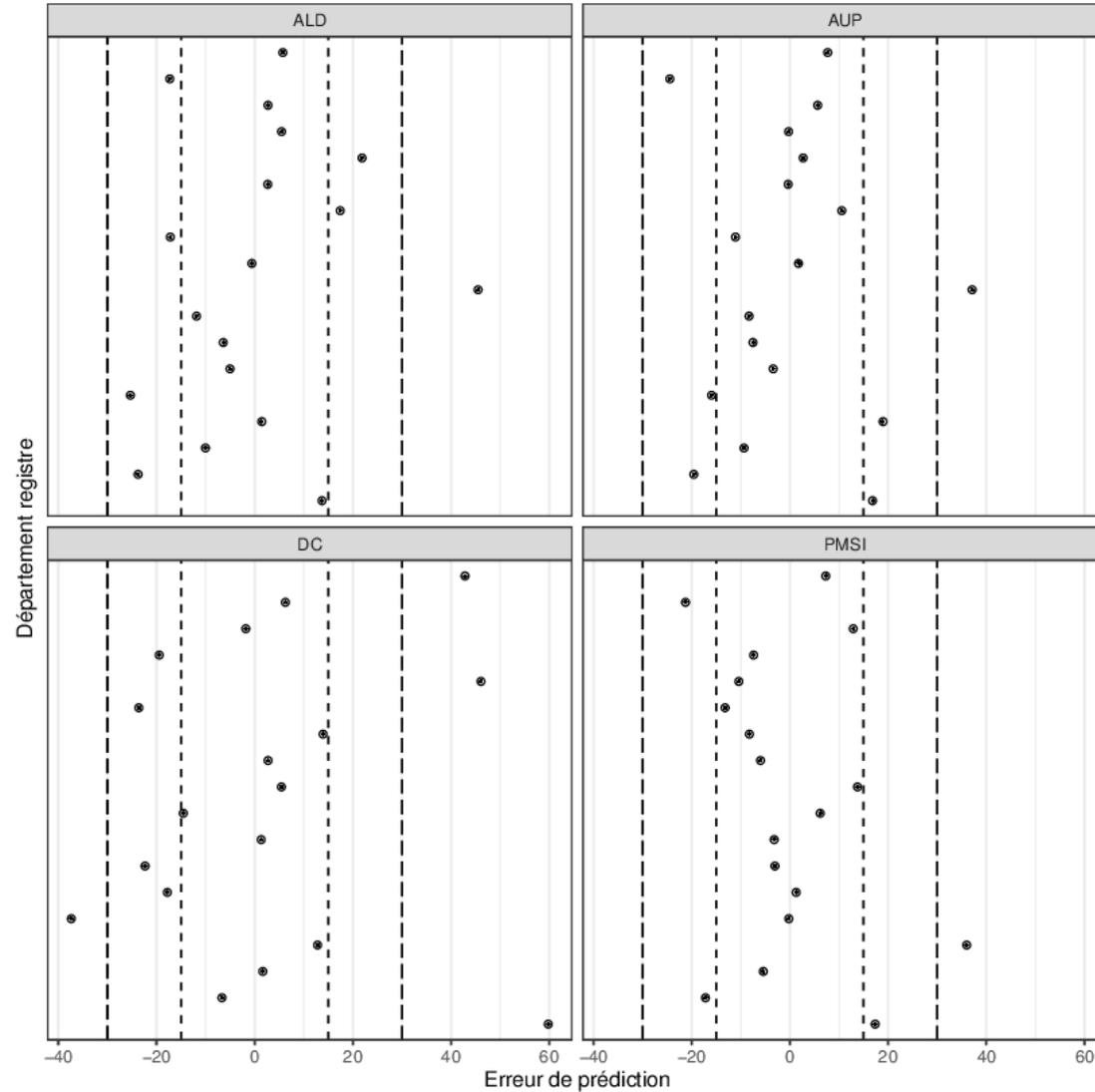
Variations départementales de l'incidence



- Prédictions précises
- Faibles variations géographiques de l'incidence ($\pm 10\%$)

Cancer du col de l'utérus

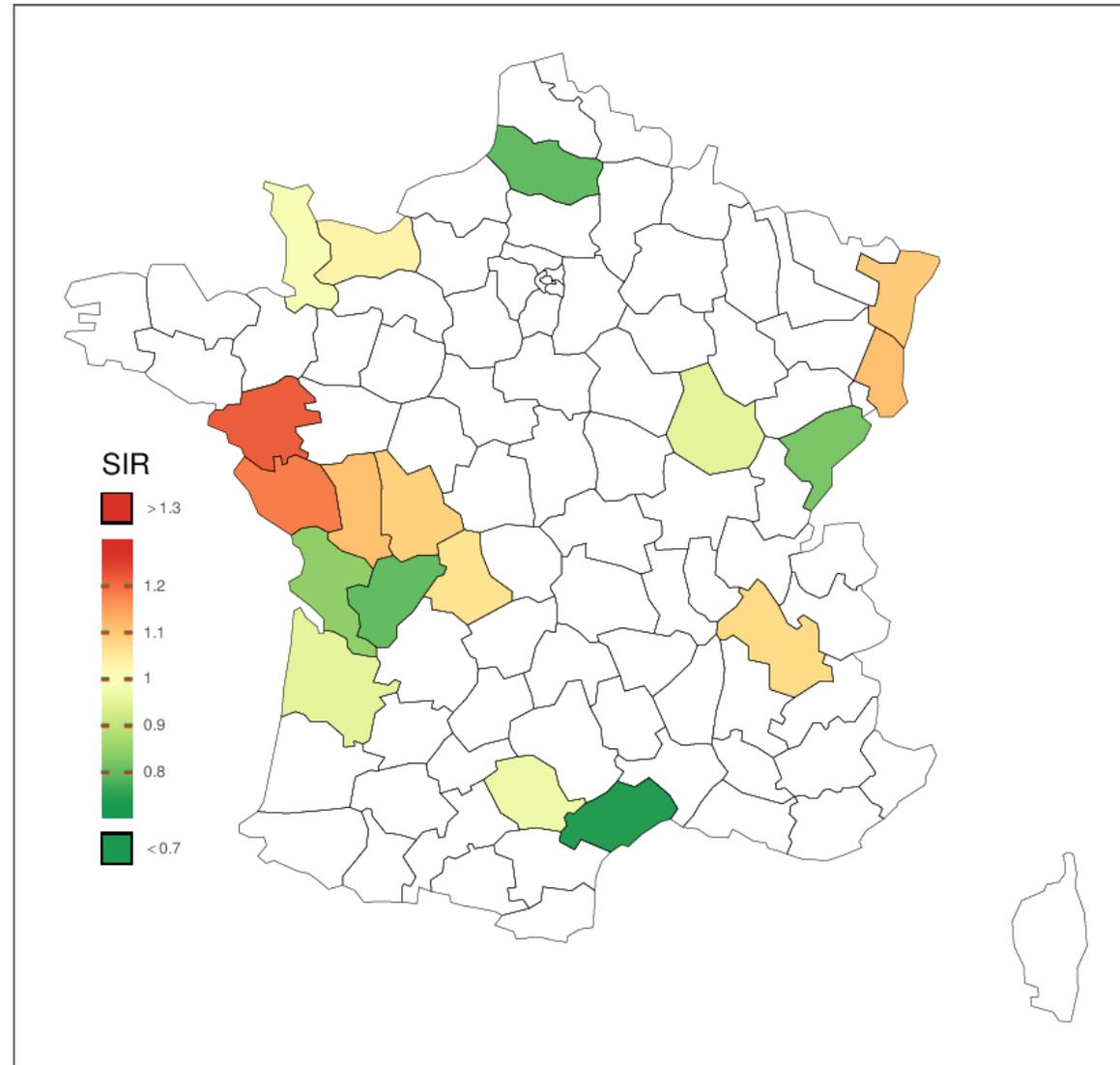
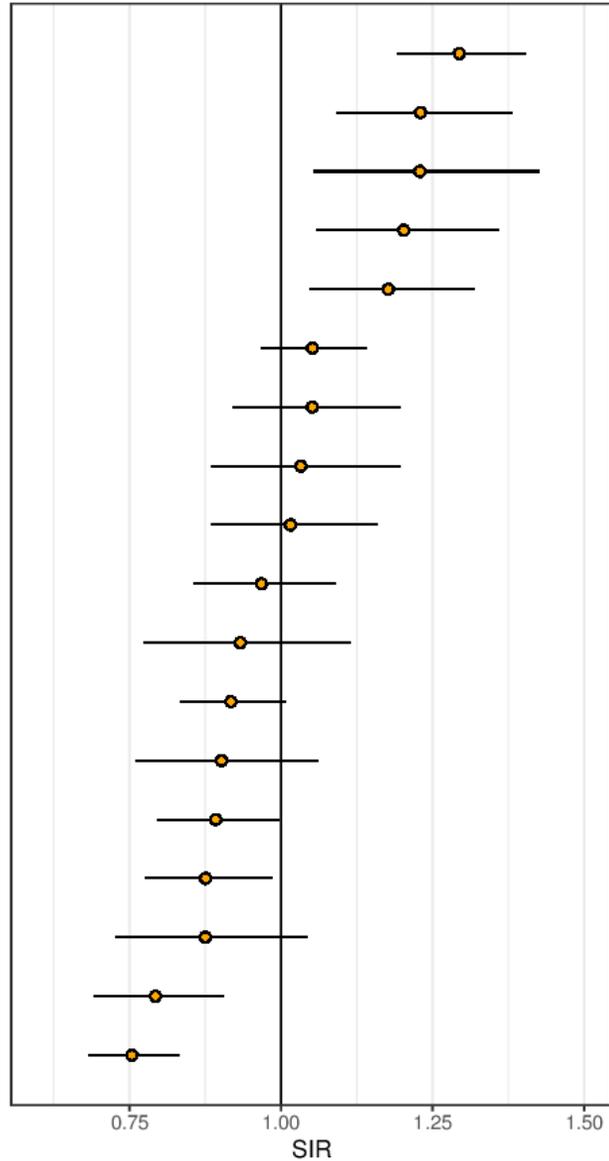
Validité des prédictions



- Erreurs importantes pour tous les proxys
- Pas de prédictions possibles (méthode B-)

Cancer du col de l'utérus

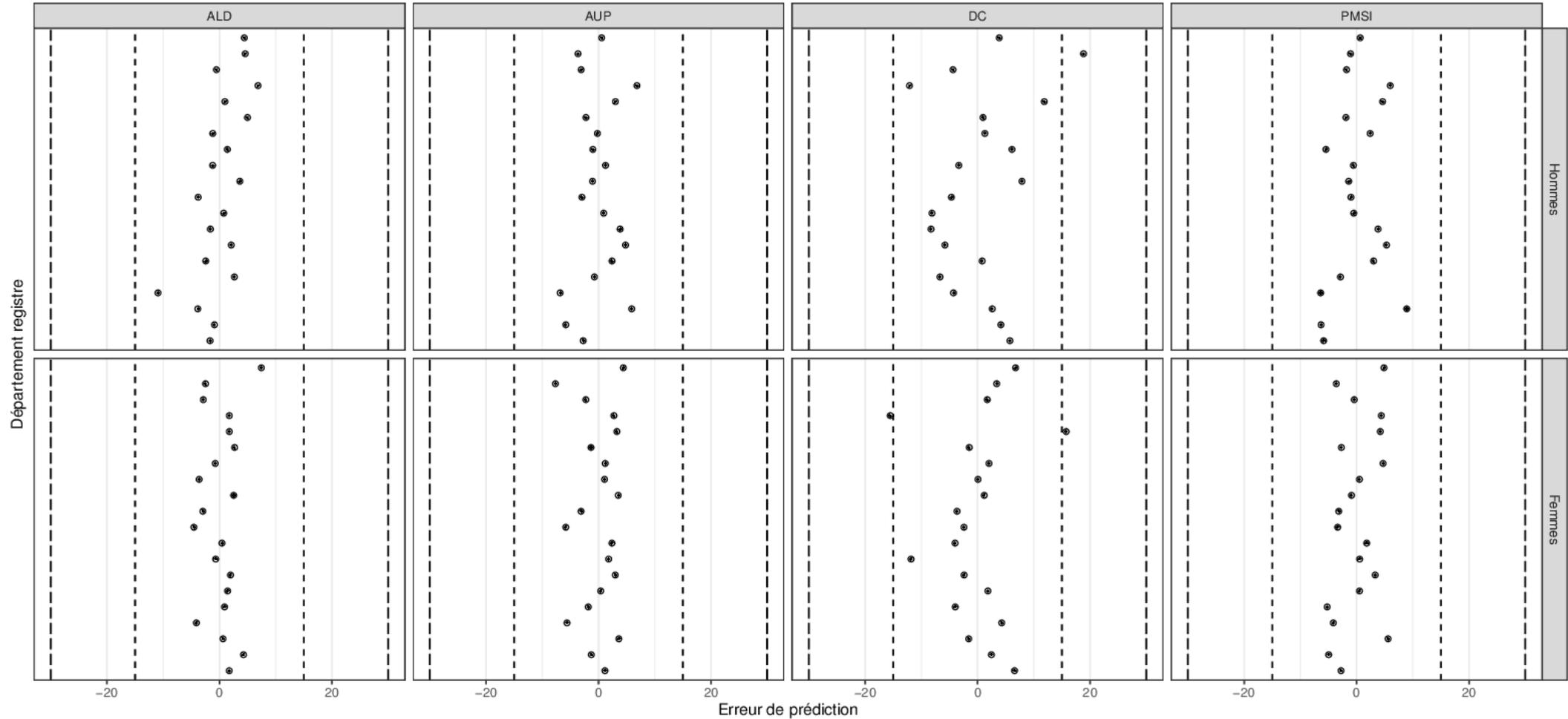
Variations départementales de l'incidence - Départements couverts par les registres uniquement



Variations géographiques de l'incidence entre les registres non négligeables ($\pm 25\%$)

Cancer du côlon-rectum-anus

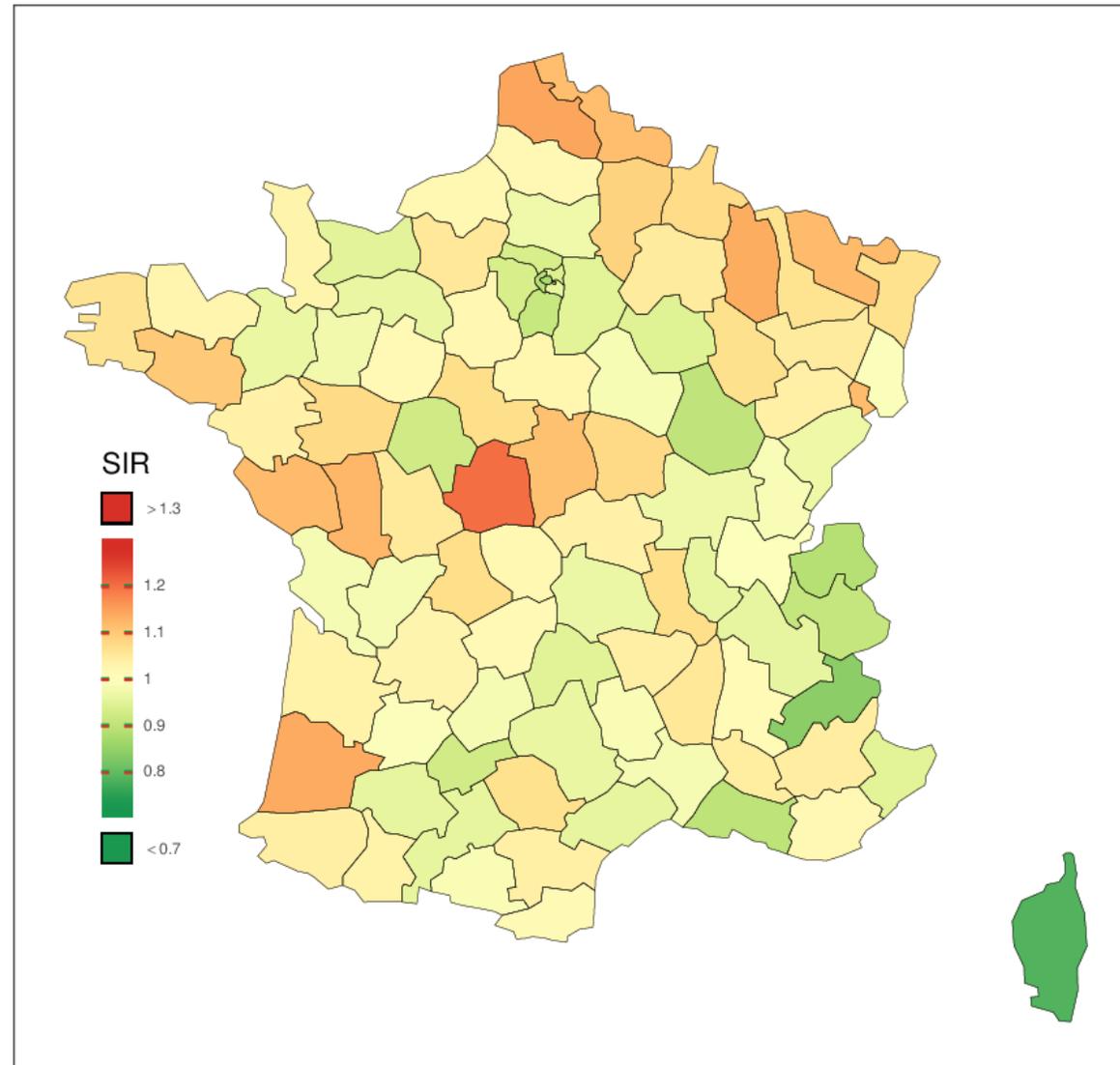
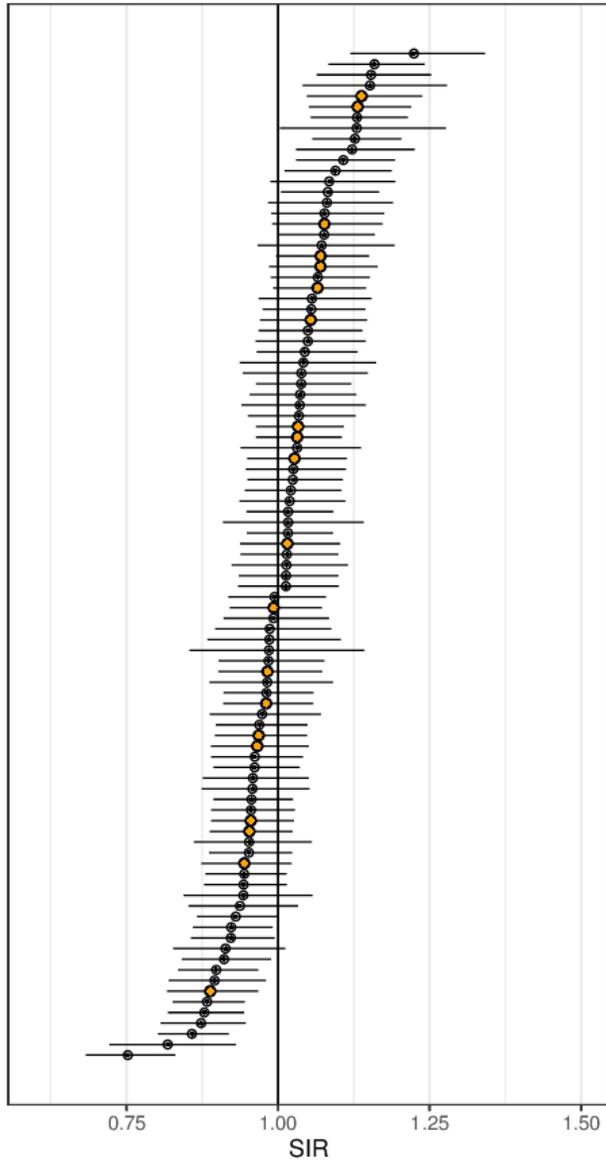
Validité des prédictions



- Faibles erreurs de prédiction pour les deux sexes
- Prédications à partir du proxy ALD

Cancer du côlon-rectum-anus

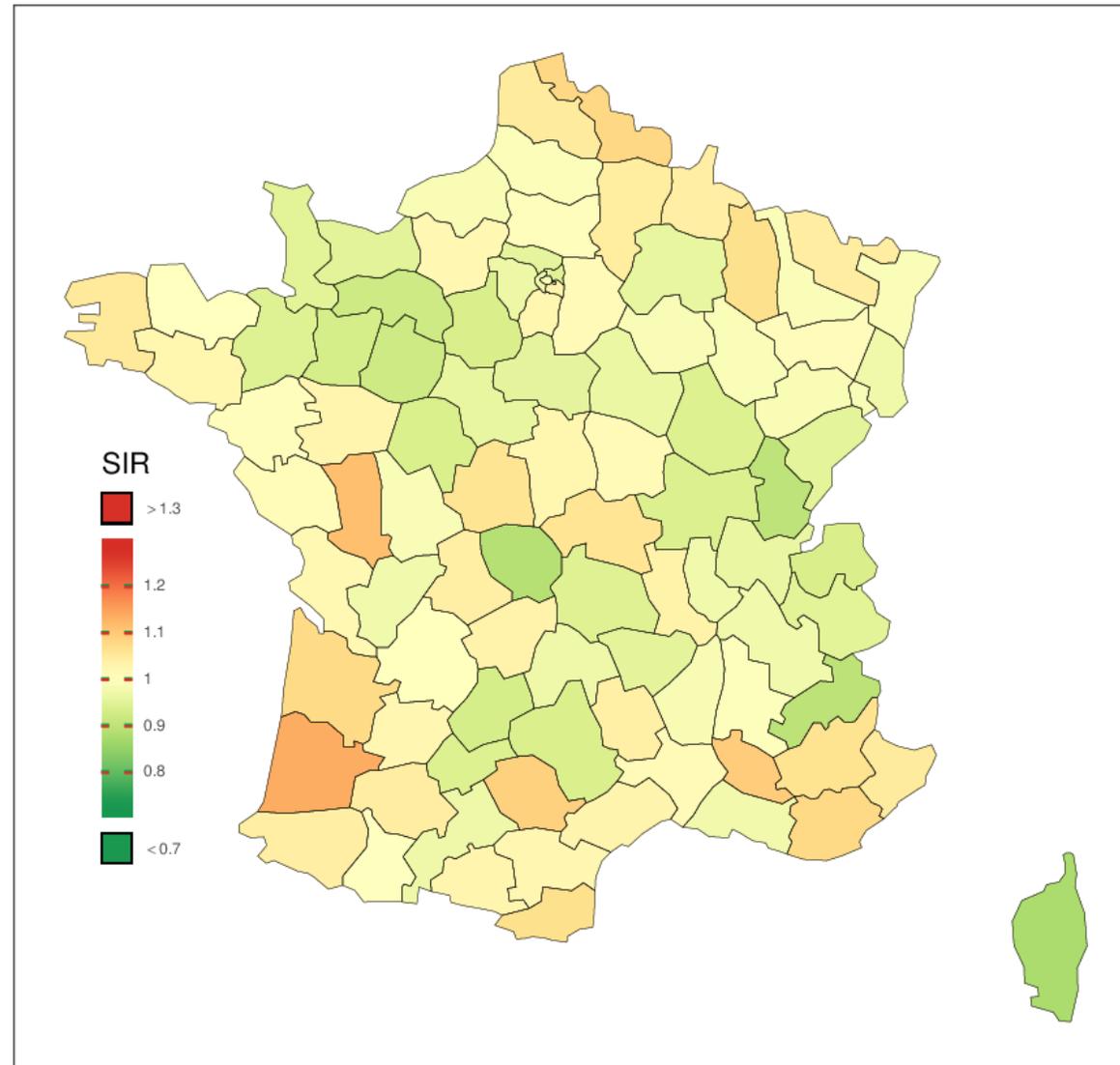
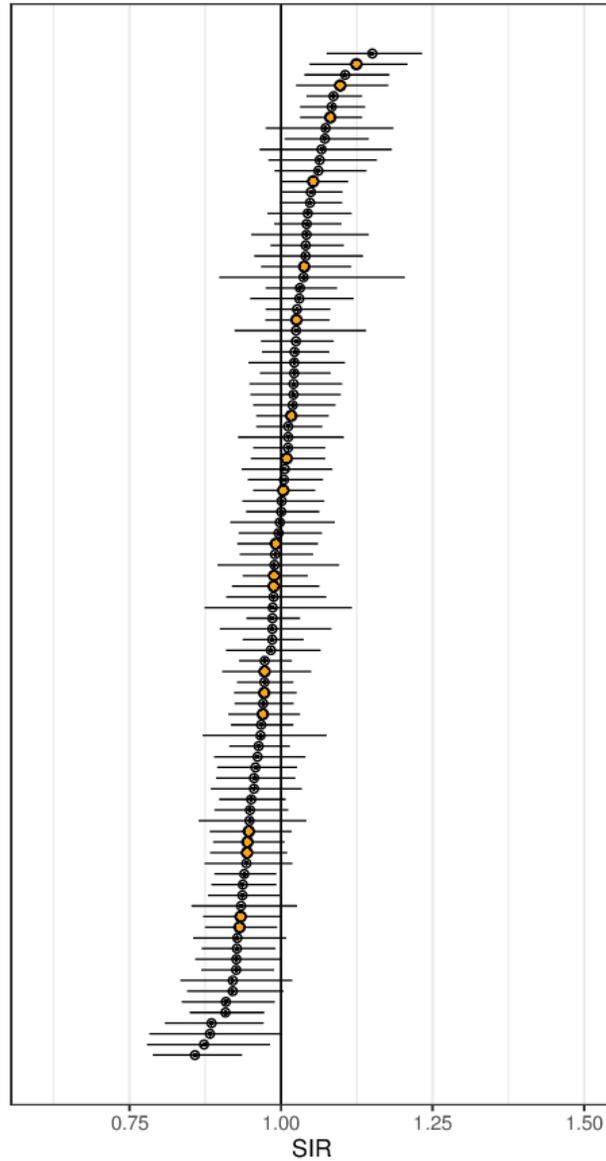
Variations départementales de l'incidence - hommes



- Prédiction précise
- Faibles variations géographiques de l'incidence ($\pm 15\%$)

Cancer du côlon-rectum-anus

Variations départementales de l'incidence - Femmes



- Prédictions précises
- Faibles variations géographiques de l'incidence ($\pm 10\%$)

Déroulé de la présentation

La surveillance des cancers en France

Méthode

Application

Discussion

Avantages de la méthode

- Cadre d'évaluation de la qualité de l'information contenues dans un proxy pour estimer l'incidence
- Correction des différences grâce au ratio
- Mesure de la vraie variabilité des prédictions obtenues en utilisant un proxy
 - Sous l'hypothèse que les départements couverts par les registres sont représentatifs en termes de ratio P/K

Predictions de l'incidence départementale pour d'autres localisations

Production de fiches régionales d'incidence (et de mortalité) 2007-2016 en 2019¹



Perspectives

Développement d'une méthode pour prédire l'incidence à un niveau infra-départemental.

1. Édouard CHATIGNOUX et al. *Estimations régionales et départementales de l'incidence et de la mortalité par cancer en France, 2007-2016*. Saint-Maurice, 2019.

Merci de votre attention !