Sentinelles
Réseau Sentinelles

Sorbonne Public Health Institute

Inserm

SANTÉ SORBONNE UNIVERSITÉ

Estimation d'une incidence commune des cas vus en consultation de médecine générale à partir de plusieurs sources de données :

méthode utilisée pour l'analyse conjointe des données de médecine générale du réseau Sentinelles et de l'entrepôt de données EMR-IQVIA

1. Présentation

Ce document présente une méthode pour analyser de façon conjointe des données de surveillance en médecine générale provenant de sources de données différentes.

Cette méthode est utilisée actuellement en France hexagonale pour analyser de façon conjointe les données collectées d'une part par le réseau Sentinelles (Sorbonne Public Health Institute, Inserm – Sorbonne Université) et d'autre part par l'entrepôt « Electronic Medical Records » (EMR) de la société IQVIA :

- les données du réseau Sentinelles sont collectées auprès des médecins généralistes participant. Il leur est demandé de rapporter chaque semaine le nombre de cas vus en consultation pour chacun des indicateurs surveillés selon des définitions de cas précises.
- les données de l'entrepôt EMR-IQVIA sont issues d'extractions automatisées des logiciels médicaux d'un échantillon de médecins généralistes. Ces médecins décrivent dans leur logiciel le(s) motif(s) de consultation de leurs patients à l'aide de codes diagnostics issus d'un thesaurus propre à l'entrepôt EMR-IQVIA. Ces codes de diagnostic sont ensuite retranscrits par la société IQVIA en codes CIM-10. Pour chaque indicateur suivi par le réseau Sentinelles, une liste spécifique de codes CIM-10 a été sélectionnée afin d'identifier les cas correspondant aux définitions utilisées par le réseau Sentinelles. Il est ainsi possible d'identifier, chaque semaine, les cas vus en consultation par les médecins de l'entrepôt EMR-IQVIA, correspondant aux définitions cliniques du réseau Sentinelles.

Un premier travail de comparaison des incidences hebdomadaires estimées séparément, à partir des données du réseau Sentinelles d'une part et de celles de l'entrepôt EMR-IQVIA d'autre part, a été effectué pour plusieurs indicateurs (infections respiratoires aiguës (IRA), syndromes grippaux (SG), diarrhée aiguë (DA), varicelle et zona) sur la période allant de janvier 2016 à décembre 2023. Les incidences ont été estimées selon la méthodologie utilisée au réseau Sentinelles (cf. Estimation des incidences à partir des données des médecins généralistes du réseau Sentinelles).

Pour chaque indicateur, il a été observé une très bonne corrélation entre les deux séries obtenues. Les incidences estimées à partir des données de l'entrepôt EMR-IQVIA étaient toutefois le plus souvent supérieures à celles estimées à partir des données du réseau Sentinelles. Par ailleurs, l'ampleur de ces écarts variait au cours du temps. Plusieurs différences méthodologiques entre les deux sources de données sont susceptibles d'expliquer les variations observées entre les deux séries :

- les critères d'inclusion des cas (définitions spécifiques pour les cas Sentinelles *vs.* une liste de codes CIM-10 pour les cas identifiés dans l'entrepôt EMR-IQVIA) : la liste de codes peut être à la fois plus large que la définition de cas Sentinelles et incomplète selon l'évolution de la nomenclature CIM-10 et des habitudes de codage ;
- le mode de calcul de la participation des médecins : en « équivalent temps plein » (ETP) pour les médecins Sentinelles *vs.* en « semaine complète » pour les médecins de l'entrepôt EMR-IQVIA) ;
- les profils des médecins participants dans chacun des deux réseaux : les modes d'exercice peuvent différer, de même que les profils dans les files actives de patients suivis;
- la couverture des deux réseaux en fonction des régions ou des départements.

Suite à ces constatations, il a été proposé :

- de calculer un facteur de mise à l'échelle des incidences EMR-IQVIA pour les ajuster à celles du réseau Sentinelles. Le calcul de ce facteur est réalisé de façon hebdomadaire et évolue avec la consolidation des données ;
- puis d'estimer une incidence conjointe en appliquant ce facteur de mise à l'échelle aux données de l'entrepôt EMR-IQVIA pour les indicateurs IRA, SG et DA.

Ce choix s'appuie sur l'hypothèse que cette mise à l'échelle permet d'estimer ce qu'aurait été l'incidence estimée à partir des données de l'entrepôt EMR-IQVIA si les cas de l'entrepôt avaient été sélectionnés sur la définition stricte suivie par le réseau Sentinelles.

2. Estimation d'un facteur μ de mise à l'échelle

Principe général

Pour une semaine t, on a :

$i_{RS}(t)$: incidence hebdomadaire Sentinelles pour un indicateur donné

$i_{IQVIA}(t)$: incidence hebdomadaire EMR-IQVIA pour un indicateur donné

$n_{RS}(t)$: participation hebdomadaire Sentinelles exprimée en nombre d'ETP (cf. Estimation des incidences à partir des données des médecins généralistes du réseau Sentinelles).

$n_{IQVIA}(t)$: participation hebdomadaire EMR-IQVIA, qui correspond au nombre de médecins « actifs » (médecins ayant réalisé au moins une consultation dans la semaine *t*)

$\alpha(t)$: incidence hebdomadaire conjointe « Sentinelles-IQVIA »

$\mu(t)$: facteur de mise à l'échelle

On pose :

- $i_{RS}(t) \sim \mathcal{N}(\alpha(t), \frac{\sigma^2}{n_{RS}(t)})$
- $i_{IQVIA}(t) \sim \mathcal{N}(\mu(t) \times \alpha(t), \frac{\sigma^2 \mu^2}{n_{IQVIA}(t)})$

On veut estimer $\mu(t)$ pour obtenir l'incidence EMR-IQVIA mise à l'échelle, $icor_{IQVIA}(t) = \frac{1}{\mu(t)} i_{IQVIA}(t)$, et l'incidence conjointe $\alpha(t)$.

Par moindres carrés :

$$\alpha(t) = \frac{n_{RS}(t)}{n_{IQVIA}(t) + n_{RS}(t)} i_{RS}(t) + \frac{n_{IQVIA}(t)}{n_{RS}(t) + n_{IQVIA}(t)} \frac{1}{\mu(t)} i_{IQVIA}(t)$$

En utilisant le poids de participation : $w(t) = \frac{n_{RS}(t) \times n_{IQVIA}(t)}{n_{RS}(t) + n_{IQVIA}(t)}$, nous pouvons calculer $\mu(t)$ par la méthode du maximum de vraisemblance : $\mu(t) = \frac{\sum_t i_{IQVIA}(t)^2 \times w(t)}{\sum_t i_{IQVIA}(t) \times i_{RS}(t) \times w(t)}$.

Estimation de $\mu(t)$ et $\alpha(t)$ pour une semaine *t* au niveau régional

Pour avoir le plus de précision possible, sachant que la dynamique des épidémies peut être différente d'une région à l'autre, pour une semaine donnée, un facteur $\mu(t,r)$ est calculé pour chaque région chaque semaine (selon les 23 « anciennes » régions avant le regroupement administratif effectué en 2016).

Pour calculer le facteur $\mu(t,r)$ au niveau régional, plusieurs étapes de calcul sont mises en œuvre (Figure 1). Elles visent à prendre en compte le fait que les incidences estimées sont plus bruitées au niveau régional qu'au niveau national en raison d'un nombre plus faible de médecins participants, ce qui peut entrainer une plus grande variabilité des incidences $\mu(t,r)$ d'une semaine à l'autre.

a) Calcul d'un $\mu(t,r)$ historique « $\mu_{hist}(t,r)$ » pour chaque semaine

Dans chaque région *r*, pour chaque semaine *t*, un $\mu_{hist}(t,r)$ est calculé selon la méthode présentée ci-dessous. Cette méthode prend en compte plusieurs semaines de données (incidences et participations) afin d'obtenir une estimation plus stable de facteur de mise à l'échelle.

$$\mu_{hist}(t,r) = \frac{\sum_{k=0}^{\eta-1} (i_{IQVIA}(t+\delta-k,r)^2 \times w(t+\delta-k,r))}{\sum_{k=0}^{\eta-1} (i_{RS}(t+\delta-k,r) \times i_{IQVIA}(t+\delta-k,r) \times w(t+\delta-k,r))}$$

Avec :

- le poids de participation : $w(t,r) = \frac{n_{RS}(t,r) \times n_{IQVIA}(t,r)}{n_{RS}(t,r) + n_{IQVIA}(t,r)}$
- $\eta \in \mathbb{N}$: nombre de semaines considérées dans le calcul du μ_{hist}
- $\delta \in \mathbb{N}$: délai (en semaines) par rapport à la semaine de calcul t

Suite aux travaux préliminaires réalisés, les paramètres η et ∂ pour le calcul de $\mu_{hist}(t,r)$ ont été fixés à $\eta = 3$ et $\partial = 1$. Ainsi, $\mu_{hist}(t=0,r)$, prendra en compte les valeurs $i_{IQVIA}(t,r)$, $i_{RS}(t,r)$ et $w(t,r)$ des semaines t=1, t=0 et t=-1.

Deux étapes sont ensuite appliquées afin de : (b) limiter le bruit de l'estimateur, (c) diminuer la variation de l'estimation entre les régions.

b) Calcul d'un $\mu(t,r)$ médian « $\mu_{md}(t,r)$ » pour chaque semaine

La valeur médiane $\mu_{md}(t,r)$ est calculée en prenant en compte η' valeurs de $\mu_{hist}(t,r)$, ces valeurs ayant un décalage temporel de δ' semaines entre la semaine t du calcul de $\mu_{md}(t,r)$ et les semaines des $\mu_{hist}(t,r)$ considérées.

$$\mu_{md}(t,r) = Mediane(\mu_{hist}(t+\delta',r);\ \mu_{hist}(t+\delta'-1,r)\ ;\ \ldots;\ \mu_{hist}(t+\delta'-\eta'+1,r))$$

Suite aux travaux préliminaires réalisés, les paramètres η' et ∂' pour le calcul de $\mu_{md}(t,r)$ ont été fixés à $\eta = 3$ et $\partial = 1$. Ainsi, $\mu_{md}(t=0,r)$ prendra en compte les valeurs $\mu_{hist}(t,r)$ des semaines t=1, t=0 et t=-1.

c) *Calcul d'un $\mu(t,r)$ final pour chaque semaine*

On applique une régression linéaire mixte sur les valeurs régionales de $\mu_{md}(t,r)$ permettant de prédire les valeurs du facteur de mise à l'échelle $\mu(t,r)$, en diminuant les variations inter-régionales :

$$\widehat{\mu(t,r)} \sim \mu_{md}(t,r) + b(r) + \varepsilon$$

Avec :

- r : région
- $b(r)$: l'effet aléatoire propre à la région

d) *Estimation de l'incidence conjointe « $\alpha(t,r)$» pour chaque semaine*

Finalement, l'incidence conjointe « Sentinelles-IQVIA » au niveau régional est estimée de la façon suivante :

$$\alpha(t,r) = \frac{n_{RS}(t,r)}{n_{IQVIA}(t,r)+n_{RS}(t,r)}\, i_{RS}(t,r) + \frac{n_{IQVIA}(t,r)}{n_{RS}(t,r)+n_{IQVIA}(t,r)}\, \frac{1}{\widehat{\mu(t,r)}}\, i_{IQVIA}(t,r)$$

Selon la méthode présentée, la semaine d'étude et les paramètres fixés, l'estimation de $\mu(t,r)$ prend en compte les données $i_{IQVIA}(t,r), i_{RS}(t,r)$ et $w(t,r)$ de cinq semaines (t+2 à t-2). La méthode de calcul est illustrée dans la Figure 1.

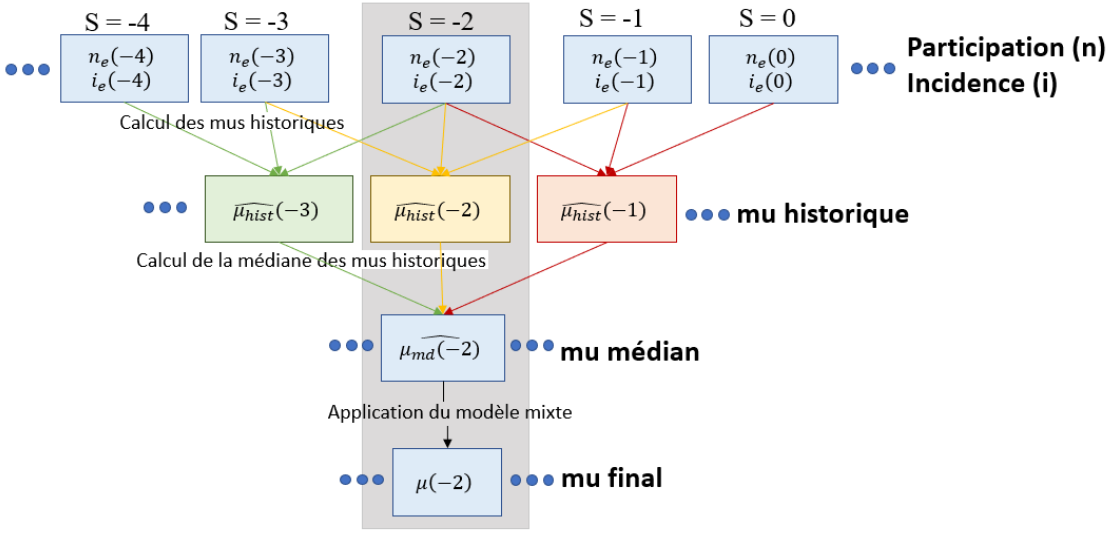


Figure 1. Calcul du $\mu(t,r)$ de la semaine t=-2 pour une région r : $\mu(t=-2,r)$

Au niveau départemental, la participation a été jugée insuffisante dans les deux bases de données pour calculer un facteur de mise à l'échelle $\mu(t,d)$. Pour estimer l'incidence conjointe $\alpha(t)$ au niveau départemental, le $\mu(t,r)$ de la région concernée (ancienne région) est utilisé.

Estimation de $\mu(t)$ et $\alpha(t)$ pour une semaine *t* au niveau régional en <u>temps réel</u>

En temps réel, la méthode présentée ci-dessus ne peut être utilisée pour les deux dernières semaines de données disponibles, car elle utilise des données d'incidence et de participation futures, donc non connues en temps réel. Cette méthode ne peut être appliquée que pour l'estimation de $\alpha(t)$ pour t ≤ -2 (avec t=0 semaine la plus récente de calcul).

Pour les semaines t=-1 et t=0, des travaux préliminaires ont permis de déterminer les valeurs optimales des paramètres η, ∂, η' et ∂'. Ces valeurs sont présentées dans le Tableau 1.

Tableau 1. Paramètres utilisés dans le calcul du facteur de mise à l'échelle historique (μ_{hist}) et dans le facteur de mise à l'échelle médian (μ_{md}) en fonction de la semaine de calcul de l'incidence en temps réel (s=dernière semaine de données disponibles)

Semaine de calcul = s	η	δ	η'	δ'
$\mu_{hist}(t = s)$	2	0	2	0
$\mu_{hist}(t = s - 1)$	2	1	3	1
$\mu_{hist}(t \leq s - 2)$	3	1	3	1

Ainsi, le calcul de :

- $\mu(t = s, r)$ prend en compte les données $i_{IQVIA}(t,r), i_{RS}(t,r)$ et $w(t,r)$ pour 3 semaines (s-2, s-1 et s), la méthode de calcul est illustrée dans la Figure 2.
- $\mu(t = -1, r)$ prend en compte les données $i_{IQVIA}(t,r), i_{RS}(t,r)$ et $w(t,r)$ pour 4 semaines (s-2 à s+1), la méthode de calcul est illustrée dans la Figure 3.

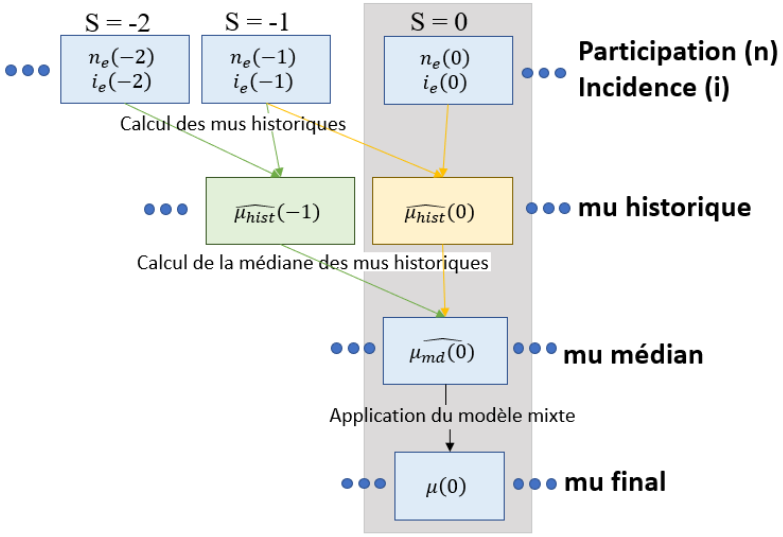


Figure 2. Calcul du $\mu(t,r)$ de la semaine t=0 pour une région r : $\mu(t = 0, r)$

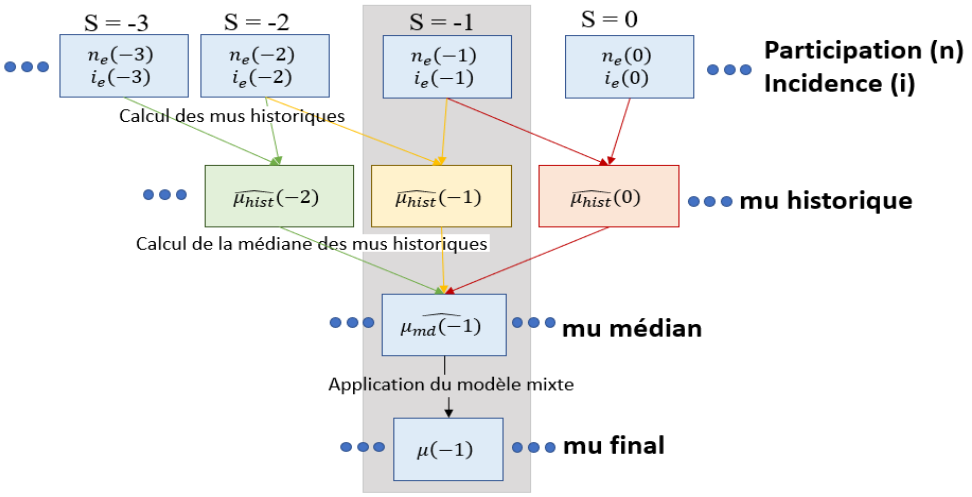


Figure 3. Calcul du $\mu(t,r)$ de la semaine t=-1 pour une région r : $\mu(t=-1,r)$

Exemple de l'estimation en temps réel de $\mu(t,r)$ pour la semaine 0 (dernière semaine de données disponibles) au niveau régional illustré dans la Figure 2

Pour calculer la valeur de $\mu(t=0,r)$, les valeurs des paramètres sont les suivantes :

- $\eta = \eta' = 2$
- $\delta = \delta' = 0$

Les étapes du calcul sont les suivantes :

1. Calcul du poids de participation $w(t,r)$, de l'incidence IQVIA $i_{IQVIA}(t,r)$ et de l'incidence Sentinelles $i_{RS}(t,r)$ pour les semaines t=0, t=-1 et t=-2 dans chaque région (23 régions)

2. Calcul de $\mu_{hist}(t,r)$ pour les semaines t=0 et t=-1 dans chaque région

Exemple pour une région :

$$\mu_{hist}(t=0) = \frac{\left(i_{IQVIA}(t=0)^2 \times w(t=0)\right)+\left(i_{IQVIA}(t=-1)^2 \times w(t=-1)\right)}{\left(i_{RS}(t=0)\times i_{IQVIA}(t=0)\times w(t=0)\right)+\left(i_{RS}(t=-1)\times i_{IQVIA}(t=-1)\times w(t=-1)\right)}$$

Ainsi, $\mu_{hist}(t=0)$ prend en compte les données des semaines 0 à -1 et $\widehat{\mu_{hist}}(t=-1)$ les données des semaines de -1 à -2.

3. Calcul de la valeur médiane $\widehat{\mu_{md}}(t=0,r)$ qui prend en compte les $\mu_{hist}(t,r)$ des semaines t=0 à t=-1 dans chaque région.

$$\mu_{md}(t=0,r) = Médiane(\mu_{hist}(t=0,r);\ \mu_{hist}(t=-1,r))$$

4. Estimation de l'estimateur final $\mu(t=0,r)$ pour chaque région à l'aide du modèle mixte

$$\mu(t=0,r) \sim \mu_{md}(t=0,r) + b(r) + \varepsilon$$

3. Application pratique : estimation d'une incidence commune « Sentinelles-IQVIA » au niveau départemental

Etapes réalisées pour une semaine donnée (t)

1) Estimation, pour chaque département, du nombre de cas que les médecins IQVIA auraient déclarés durant cette semaine (t) s'ils avaient suivi la définition du réseau Sentinelles et calcul du nombre de cas vus par l'ensemble des médecins Sentinelles dans chaque département pour la semaine t :

- on calcule pour chaque région (n=23, anciennes régions avant 2016) le facteur de mise à l'échelle régional hebdomadaire $\mu(t,r)$ des données EMR-IQVIA selon la méthode ci-dessus ;
- on applique ce facteur de correction régional $\mu(t,r)$ à la somme du nombre de cas déclarés par les médecins IQVIA dans chaque département de la région ($nbcas_{RS}(t,d)$) ;
- on réalise la somme des cas déclarés par les médecins Sentinelles ($nbcas_{RS}(t,d)$) dans chaque département pour la semaine t.

2) Estimation de la participation des médecins Sentinelles et IQVIA :

- on calcule la participation des médecins Sentinelles dans chaque département pour la semaine t ($n_{RS}(t,d)$) ;
- on calcule la participation des médecins de l'entrepôt EMR-IQVIA dans chaque département pour la semaine t ($n_{IQVIA}(t,d)$).

3) Estimation de l'incidence conjointe « Sentinelles-IQVIA » :

- Pour le département d, l'incidence conjointe est estimée selon la formule suivante :

$$\alpha(t,d) = \frac{nbcas_{RS}(t,d) + \frac{1}{\mu(t,r)}\, nbcas_{IQVIA}\,(t,d)}{n_{RS}(t,d) + n_{IQVIA}(t,d)} nMG(t,d)$$

Où d appartient à la région r et $nMG(t,d)$ est le nombre total de médecins généralistes libéraux exerçant dans le département d durant l'année de la semaine t.

Pour l'estimation des incidences régionales et nationales, la méthode suit la méthodologie utilisée historiquement au réseau Sentinelles et présentée dans le document téléchargeable suivant : Estimation des incidences à partir des données des médecins généralistes du réseau Sentinelles.