

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 du réseau Sentinelles et de l'entrepôt EMR-IQVIA

## 1. Présentation

Ce document présente une méthode de calcul pour analyser de façon commune les données de surveillance en médecine générale collectées par des réseaux différents.

Cette méthode est utilisée actuellement en France hexagonale pour analyser de façon commune les données collectées d'une part par le réseau Sentinelles (Iplesp, 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 par les médecins généralistes Sentinelles selon des définitions de cas clinique spécifiques à chaque indicateur surveillé. Il leur est demandé de déclarer chaque semaine le nombre de cas vus en consultation pour chacun des indicateurs suivis.
- les données de l'entrepôt EMR-IQVIA sont issues d'extractions automatiques des logiciels métiers d'un échantillon de médecins généralistes. Ces médecins utilisent des codes diagnostics issus d'un thesaurus propre à l'entrepôt EMR-IQVIA pour décrire le motif de consultation. Ces codes de diagnostic sont ensuite retranscrits par la société IQVIA en codes CIM-10. Pour chacun des indicateurs suivis par le réseau Sentinelles, une liste spécifique de codes CIM-10 a été sélectionnée, afin d'identifier les cas vus en consultation chaque semaine par les médecins de l'entrepôt EMR-IQVIA.

Un premier travail de comparaison des incidences hebdomadaires estimées séparément, à partir des données du réseau Sentinelles et de celles de l'entrepôt EMR-IQVIA, a été effectué pour plusieurs indicateurs (infections respiratoires aiguës (IRA), syndromes grippaux (SG), 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](#)).

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 sont toutefois le plus souvent supérieures à celles estimées à partir des données du réseau Sentinelles. Plusieurs différences méthodologiques peuvent expliquer ces variations entre les deux séries :

- critères d'inclusion des cas (définitions spécifiques pour les cas Sentinelles vs. liste de codes CIM-10 pour les cas de l'entrepôt EMR-IQVIA), la liste de code pouvant 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 ;

- 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) ;
- profils des médecins participants dans chacun des deux réseaux avec de potentielles différences de profils dans les files actives respectives des patients suivis ;
- couverture des deux réseaux en fonction des régions ou des départements.

Suite à ces constatations, il a été proposé :

- de calculer un facteur hebdomadaire de mise à l'échelle des incidences EMR-IQVIA pour les ajuster à celles du réseau Sentinelles ;
- puis de calculer une incidence commune en appliquant ce facteur de mise à l'échelle aux données brutes de l'entrepôt EMR-IQVIA.

Ce choix s'appuie sur l'hypothèse que cette mise à l'échelle permet d'estimer ce qu'aurait été l'incidence 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 $\mu$ 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 commune « Sentinelles-IQVIA »

$\mu(t)$  : facteur de mise à l'échelle

On pose :

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

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

Par moindres carrés :

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

### Estimation de $\mu(t)$ pour une semaine (t)

Avec le poids de participation :  $w(t) = \frac{n_{RS}(t) \times n_{IQVIA}(t)}{n_{RS}(t) + n_{IQVIA}(t)}$

Une fois le poids de participation calculé pour chaque semaine  $t$ , nous pouvons estimer le  $\mu(t)$  par la méthode du maximum de vraisemblance :

$$\mu(t) = \frac{i_{IQVIA}(t)^2 \times w(t)}{i_{IQVIA}(t) \times i_{RS}(t) \times w(t)}$$

En temps réel, l'incidence Sentinelles  $i_{RS}(t)$  n'est pas connue de façon consolidée, il n'est donc pas possible de calculer un  $\mu(t)$  consolidé. Pour estimer  $\mu(t)$ , on calcule un  $\mu(t)$  historique, qui considère les données des semaines précédentes.

$$\widehat{\mu}_{hist}(t) = \frac{\sum_{k=1}^{\theta} (i_{IQVIA}(t-k)^2 \times w(t-k))}{\sum_{k=1}^{\theta} (i_{RS}(t-k) \times i_{IQVIA}(t-k) \times w(t-k))}$$

Avec :

$\theta \in \mathbb{Z}$  : nombre de semaines considérées (période glissante avant la semaine t)

Suite aux travaux préliminaires réalisés,  $\theta$  a été fixé à 3 pour l'estimation du  $\mu(t)$  historique. En conséquence,  $\widehat{\mu}_{hist}(4)$ , pour la semaine 4, prendra en compte les valeurs  $i_{IQVIA}(t)$ ,  $i_{RS}(t)$  et  $w(t)$  pour les semaines 1 à 3.

### Estimations d' $\alpha(t)$ pour une semaine (t)

L'incidence hebdomadaire commune « Sentinelles-IQVIA » est estimée de la façon suivante :

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

avec  $\mu(t) = \widehat{\mu}_{hist}(t)$ , où  $\theta = 3$

### 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  $\mu(t, r)$  spécifique est calculé pour chaque région (selon les anciennes régions avant le regroupement administratif effectué en 2016 = NUTS2).

Pour calculer le facteur  $\mu(t, r)$  au niveau régional, des étapes supplémentaires de calcul ont été ajoutées par rapport à la méthode présentée ci-dessus. Cela fait suite à la constatation que :

- les estimations des incidences régionales Sentinelles et EMR-IQVIA sont plus bruitées qu'au niveau national du fait d'un nombre de médecins participants plus faible ;
- que cela peut entraîner une variabilité plus importante des  $\mu(t)$  d'une semaine à l'autre.

Dans chaque région  $r$ , pour chaque semaine  $t$ , le  $\widehat{\mu}_{hist}(t, r)$  est calculé. Deux étapes sont ensuite appliquées afin de : (1) limiter le bruit de l'estimateur, (2) diminuer la variation de l'estimation entre les régions.

(1) La valeur médiane  $\widehat{\mu}_{md}(t, r)$  est calculée en prenant en compte les valeurs  $\widehat{\mu}_{hist}(t, r)$  entre  $t$  et  $t - 2$ ), permettant de limiter le bruit de l'estimateur.

$$\widehat{\mu}_{md}(t, r) = \text{Mediane}(\widehat{\mu}_{hist}(t, r); \dots; \widehat{\mu}_{hist}(t - 2, r))$$

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

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

Où  $r$  : région

L'incidence commune « Sentinelles-IQVIA » au niveau régional est :

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

Selon la méthode présentée, lorsque  $\theta = 3$  dans l'estimation du  $\mu(t)$  historique, alors, les données des 5 dernières semaines précédant la semaine  $t$  seront utilisées pour estimer  $\mu(t, r)$  (cf. exemple ci-dessous).

Au niveau départemental, la participation a été jugée insuffisante dans les deux bases de données pour calculer un facteur de mise à l'échelle  $\mu(t)$ , le  $\mu(t)$  de la région concernée (ancienne région, NUTS2) est donc utilisé pour estimer l'incidence  $\alpha(t)$  départementale.

### Exemple de l'estimation de $\mu(t)$ pour la semaine 6 ( $t=6$ ) au niveau régional

L'exemple est illustré dans la Figure 1 ci-dessous.

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 1 à 5 dans chaque région (22 régions)
2. Calcul de  $\widehat{\mu}_{hist}(t, r)$  pour les semaines 4 à 6 dans chaque région

Exemple pour une région :

$$\widehat{\mu}_{hist}(4) = \frac{(i_{IQVIA}(1)^2 \times w(1)) + (i_{IQVIA}(2)^2 \times w(2)) + (i_{IQVIA}(3)^2 \times w(3))}{(i_{RS}(1) \times i_{IQVIA}(1) \times w(1)) + (i_{RS}(2) \times i_{IQVIA}(2) \times w(2)) + (i_{RS}(3) \times i_{IQVIA}(3) \times w(3))}$$

Ainsi,  $\widehat{\mu}_{hist}(4)$  prend en compte les données des semaines 1 à 3,  $\widehat{\mu}_{hist}(5)$  les données des semaines de 2 à 4 et  $\widehat{\mu}_{hist}(6)$  les données des semaines 3 à 5.

3. Calcul de la valeur médiane  $\widehat{\mu}_{md}(6, r)$  qui prend en compte les  $\widehat{\mu}_{hist}(t, r)$  des semaines 4 à 6 dans chaque région.

$$\widehat{\mu}_{md}(6, r) = \text{Médiane}(\widehat{\mu}_{hist}(4, r); \widehat{\mu}_{hist}(5, r); \widehat{\mu}_{hist}(6, r))$$

4. Estimation de l'estimateur final  $\mu(6, r)$  pour chaque région à l'aide du modèle mixte

$$\mu(6, r) \sim \widehat{\mu}_{md}(6, r) + b(r) + \varepsilon$$

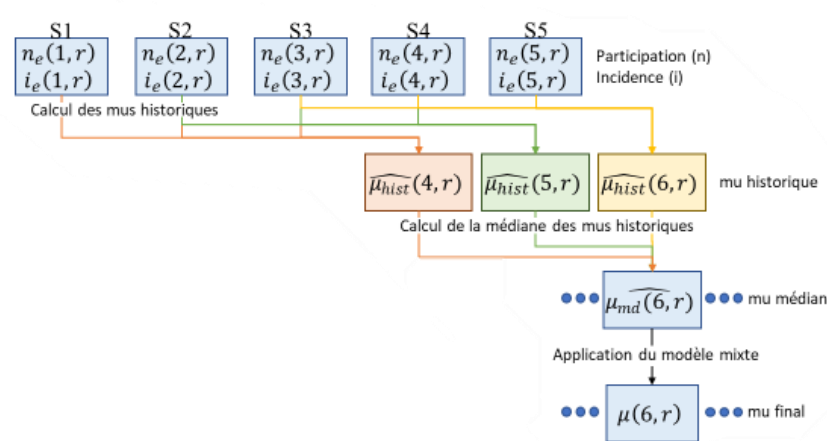


Figure 1 : Calcul du  $\mu$  de la semaine 6 pour une région  $r$  :  $\mu(6, r)$  ( $e$  : indice pouvant prendre la valeur IQVIA ou RS)

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

#### 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 :

- on calcule pour chaque région (NUTS 2, n=22 régions) 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  $\mu(t, r)$  au nombre de cas déclarés par les médecins IQVIA dans chaque département de la région concernée.

2) Estimation pour chaque département du nombre de cas vus par l'ensemble des médecins Sentinelles et IQVIA

- somme des cas vus par les médecins Sentinelles et des cas vus par les médecins de l'entrepôt EMR-IQVIA mis à l'échelle avec le facteur  $\mu(t, r)$  calculé précédemment

3) Estimation de la participation commune des médecins Sentinelles et IQVIA

- somme de la participation des médecins Sentinelles et des médecins de l'entrepôt EMR-IQVIA

4) Estimation de l'incidence commune « Sentinelles-IQVIA »

- Pour le département  $d$ , l'incidence est calculée 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 ci-contre : [Estimation des incidences à partir des données des médecins généralistes du réseau Sentinelles](#).