

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 de 1210 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 les médecins du réseau Sentinelles, il a été sélectionné une liste spécifique de codes CIM-10 correspondant, afin d'identifier les cas vus en consultation chaque semaine par les médecins de l'entrepôt.

Un premier travail de comparaisons 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 est 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 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 par semaine)

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

$\mu(t)$: facteur de mise à l'échelle

On pose :

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

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 :

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

L'estimateur de $\mu(t)$ par la méthode de maximum de vraisemblance :

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

Avec $w(t) = \frac{n_{IQVIA}(t) n_{RS}(t)}{n_{IQVIA}(t) + n_{RS}(t)}$

Estimations de $\mu(t)$ et $\alpha(t)$ pour une semaine donnée

Afin de tenir compte des variations de $\mu(t)$ d'une semaine à l'autre, pour une semaine T4, $\widehat{\mu(t)}$ de T4 est calculé sur une période glissante de 3 semaines à partir des 3 semaines précédentes (T1 à T3).

On calcule la moyenne des $\widehat{\mu(t)}$ à partir de la période T1 : T3, on l'applique à T4, et ainsi de suite.

Ainsi :

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

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

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

Pour calculer le facteur $\mu(t)$ au niveau régional, les calculs ont été un peu affinés 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.

On réalise les étapes suivantes pour chacune des régions :

- Etape 1 = on utilise la même procédure que précédemment pour estimer $\mu(t)$;
- Etape 2 = on calcule la médiane des $\mu(t)$ d'une manière rétrospective sur les 3 dernières semaines (on calcule par exemple la médiane des $\mu(t)$ sur la période T4 : T6, et on l'applique à T6, et ainsi de suite) ;
- Etape 3 = on applique une régression linéaire mixte, qui va permettre de prédire un facteur de mise à l'échelle $\mu(t)$ qui prendra en compte les informations de toutes les régions ($\mu_i(t) \sim \widehat{\mu(t)} + b_i + \varepsilon$).

Avec :

- $\widehat{\mu(t)}$: estimateur de $\mu(t)$ par la méthode de maximum de vraisemblance ;

- $\mu_i(t)$: valeur prédit de $\widehat{\mu}(t)$ par le model de régression mixte en prenant la médiane sur les 3 dernières semaines ;
- $b_i: \frac{1}{id_geo}$, id_geo , variable identifiant le niveau géographique considéré

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

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

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) est donc utilisé pour estimer le $\alpha(t)$ départemental.

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 (ancienne région avant 2016) un facteur de mise à l'échelle régional hebdomadaire $\mu_i(t)$ des données EMR-IQVIA selon la méthode précédemment présentée ;
- on applique ce facteur de correction régional $\mu_i(t)$ au nombre de cas déclarés par les médecins IQVIA dans chaque département de la région concernée.

3) 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 $\mu_i(t)$ 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 »

- Au niveau départemental, l'incidence est calculée selon la formule suivante :

$$\alpha(t) = \frac{\sum(nbcas_{RS} + \mu_i(t) nbca_{IQVIA})}{\sum(participation_{RS}(t) + participation_{IQVIA})} nb\ total\ MG$$

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.](#)