Sentinelles
Réseau Sentinelles

iPLesp Inserm SANTÉ SORBONNE UNIVERSITÉ

Estimation des incidences des cas d'IRA vus en consultation de médecine générale attribuables à un virus respiratoire donné

1. Objectifs

L'objectif principal est d'estimer l'incidence hebdomadaire en France métropolitaine des cas d'IRA vus en consultation de médecine générale attribuables à un virus respiratoire donné (c'est à dire dans le cadre de la surveillance Sentinelles, aux virus grippaux (tous types/sous-types confondus), au VRS et au Sars-CoV-2).

Les objectifs secondaires sont l'estimation de ces incidences par groupes d'âge et par inter-régions.

2. Données

Les données collectées par le réseau Sentinelles utilisées pour ces estimations sont :

- le nombre de cas d'infections respiratoires aiguës (IRA) déclarés par semaine par les MG Sentinelles ;
- la description individuelle de ces cas (80-95% des cas rapportés sont décrits) ;
- les taux de positivité par virus respiratoire (virus grippaux, VRS et SARS-CoV-2) des prélèvements réalisés sur un échantillon des cas d'IRA vus en consultation par les médecins Sentinelles.

Les hypothèses faites dans ce travail sont les suivantes :

- **H1/** Les cas d'IRA décrits sont représentatifs des cas d'IRA déclarés par les médecins Sentinelles ;
- **H2/** Les cas d'IRA prélevés sont représentatifs des cas d'IRA déclarés par les médecins Sentinelles.

3. Méthode d'estimation des incidences

Conventions :

- N = nombre total de MG exerçant une activité libérale
- I = incidence
- C = nombre de cas rapportés par les médecins Sentinelles
- P = participation des médecins Sentinelles
- z = zone géographique [1, ...]
- Z = ensemble des zones géographiques z
- a = classe d'âge [1, ...]
- A = ensemble des classes d'âge

a. Estimation de l'incidence des cas d'IRA vus en consultation de médecine générale

Le calcul de l'incidence des cas d'IRA vus en consultation de médecine générale utilisé au réseau Sentinelles permet de prendre en compte la participation des médecins. Il s'appuie sur la formule suivante pour une semaine s et une zone géographique Z :

$$\hat{I}(s) = \sum_{z=1}^{Z} \frac{C_{IRA}(s,z)}{P(s,\ z)} N_z^{med} \quad (1)$$

Où :

- $C_{IRA}(s,z)$ est le nombre de cas d'IRA rapportés par les médecins Sentinelles dans la zone *z*, pour la semaine *s*
- $P(s,z)$ est la participation des médecins Sentinelles pour la semaine *s*, dans la zone géographique *z*
- N_z^{med} est le nombre total de médecins généralistes dans la zone *z*

Des explications plus complètes et le détail du calcul de cette estimation sont apportées dans un autre document téléchargeable dans la page « Méthode » du site du réseau Sentinelles : Estimation des incidences à partir des données des médecins généralistes du réseau Sentinelles

b. Estimation de l'incidence des cas d'IRA vus en consultation de médecine générale attribuables au virus V

Le calcul de l'estimation de l'incidence des cas vus en consultation de médecine générale attribuables au virus V s'appuie sur deux estimations :

- l'incidence des cas d'IRA vus en consultation de médecine générale ;
- le taux de positivité, pour ce virus V, des cas d'IRA en consultation de médecine générale.

Ce taux de positivité peut être estimé à partir de deux sources de données :

- pour les virus grippaux, le VRS et le Sars-CoV-2, à partir des cas d'IRA prélevés par les médecins Sentinelles ;
- pour le Sars-CoV-2, à partir des questions intégrées aux formulaire de description des cas d'IRA déclarés par les médecins Sentinelles concernant les résultats des tests virologiques pour la Covid-19 réalisés en ville en dehors de la surveillance virologique Sentinelles. Ces résultats permettent de fournir une estimation du taux de positivité plus robuste pour le Sars-CoV-2 : le nombre de cas déclarés et décrits d'IRA pendant la saison est plus de 10 fois supérieur au nombre de prélèvements réalisés par semaine dans le cadre de la surveillance virologique. Le taux de positivité utilisé alors correspond au nombre de cas rapporté positif divisé par le nombre de cas décrits.

On estime donc,

$$\hat{C}_{v_{\langle type \rangle}}(s,z) = C_{IRA}(s,z) \cdot \hat{V}_{\langle type \rangle}(s,z) \quad (2)$$

Où :

- $\hat{C}_{v_{\langle type \rangle}}(s,z)$ est l'estimation du nombre de cas d'IRA attribuables au virus V vus en consultation par les médecins Sentinelles dans la zone *z* pour la semaine *s ;*
- $\hat{V}_{\langle type \rangle}(s,z)$ est le taux de positivité, pour le virus V, des cas d'IRA vus en consultation de médecine générale durant la semaine s pour la zone z.

L'incidence des cas vus en consultation de médecine générale attribuables au virus V peut alors être estimée par :

$$\hat{I}_{v_{\langle type \rangle}}(s) = \sum_{z \in Z} \frac{C_{IRA}(s,z) \cdot \hat{V}_{\langle type \rangle}(s,z)}{P(s,\ z)} N_z^{med} \qquad (3)$$

c. Estimation du taux de positivité des cas d'IRA pour le virus V

Trois estimateurs du taux de positivité, pour le virus V, des cas d'IRA vus en consultation de médecine générale durant la semaine s pour la zone z peuvent être utilisés :

- zonal = taux de positivité pour le virus V des prélèvements réalisés par les médecins Sentinelles durant la semaine s dans la zone z ;
- national = taux de positivité pour le virus V des prélèvements réalisés par les médecins Sentinelles durant la semaine s pour l'ensemble de la France métropolitaine ;
- zonal(tel) = taux de positivité pour le virus V des prélèvements réalisés par les médecins Sentinelles durant la semaine s dans l'ensemble d'une zone téléphonique donnée.

Suite au nombre limité de prélèvements réalisés par semaine par les médecins Sentinelles, la règle suivante est appliquée pour l'estimation du taux de positivité dans la zone z durant la semaine s :

- si pour la zone téléphonique comprenant la zone z, le nombre de prélèvements est supérieur ou égale à 30, alors le taux de positivité utilisé pour la zone z est le taux de la zone téléphonique, « zonal(tel) » ;
- sinon le taux de positivité utilisé pour la zone z est le taux national.

d. Estimation de l'incidence des cas d'IRA vus en consultation de médecine générale par classe d'âge

En complément des données utilisées précédemment, on considère les cas d'IRA de la semaine s dont l'âge a été renseigné par les médecins dans le formulaire de description des cas. Ces cas constituent un sous-échantillon de l'ensemble des cas déclarés par les médecins

Sentinelles. Ils permettent d'estimer $\widehat{D}(s,z,a)$, c'est-à-dire la proportion des cas déclarés appartenant au groupe d'âge a, pour la zone z et la semaine s

L'incidence des cas d'IRA vus en consultation de médecine générale par classe d'âge peut alors être calculée ainsi :

$$\hat{I}(s,a) = \sum_{z=1}^{Z} \frac{\widehat{D}(s,z,a) * C_{IRA}(s,z)}{P(s,\ z)} N_z^{med} \qquad (4)$$

Dans l'idéal, s'il n'y avait pas de données manquantes, $\widehat{D}(s,z,a)$ devrait être la proportion observée de l'ensemble des cas déclarés par les médecins Sentinelles pour le groupe d'âge a dans la zone z durant la semaine s. L'âge est décrit pour 80 à 95% des cas déclarés.

En pratique, afin d'augmenter la fiabilité de l'estimation de la proportion des cas déclarés appartenant au groupe d'âge a, ce taux est calculé au niveau national et appliqué à l'ensemble des zones z. On considère donc que $\widehat{D}(s,z,a) = D(s,a)$, en faisant l'hypothèse que la structure d'âge des cas d'IRA est la même dans chaque région.

Dans la suite on gardera la formulation $\widehat{D}(s,z,a)$ pour conserver une écriture générique à l'estimateur d'incidence, indépendante du choix actuel pour l'estimation de la structure d'âge des cas d'IRA déclarés.

e. Estimation de l'incidence des cas d'IRA vus en consultation de médecine générale attribuables au virus V par classe d'âge

Dans la suite de ce qui a été présenté précédemment, l'incidence des cas d'IRA vus en consultation de médecine générale attribuables au virus V par classe d'âge peut être calculée ainsi :

$$\hat{I}_{v_{\langle type \rangle}}(s,a) = \sum_{z \in Z}^{Z} \frac{\widehat{D}(s,z,a) \ * \ C_{IRA}(s,z) \cdot \hat{V}_{\langle type \rangle}(s,z,a)}{P(s,\ z)} N_z^{med}$$

L'incidence totale est la somme des incidences attribuables pour chaque groupe d'âge

$$\hat{I}_{v_{\langle type \rangle}}(s) = \sum_{a \in A}^{A} \sum_{z \in Z}^{Z} \frac{\widehat{D}(s,z,a) \ * \ C_{IRA}(s,z) \cdot \hat{V}_{\langle type \rangle}(s,z,a)}{P(s,\ z)} N_z^{med}$$

En pratique suite au nombre limité de prélèvements réalisés par semaine par les médecins Sentinelles dans chaque tranche d'âge, on considère que $\hat{V}_{\langle type \rangle}(s,z,a) = \hat{V}_{\langle type \rangle}(s,a)$.

4. Calcul de l'intervalle de confiance à 95%

Trois échantillons sont utilisés pour l'estimation d'incidence :

- les cas rapportés par les médecins ;
- les cas décrits pour la variable âge (sous-ensemble des cas rapportés) ;
- les cas prélevés pour l'analyse virologique dont le résultat est disponible (sous-ensemble imparfait des cas rapportés [1]).

Ces trois échantillons sont utilisés pour l'estimation d'incidence attribuable. Ils contribuent chacun à l'estimation d'une proportion ou d'un total, mais ne sont pas indépendants entre eux, puisque tous constituent des sous-échantillons les uns des autres. La formulation mathématique de l'intervalle de confiance de l'incidence ainsi exprimée (ci-dessus) est donc complexe.

Nous avons donc préféré l'estimation d'un intervalle de confiance de l'incidence attribuable grâce à une procédure de rééchantillonnage multiple (ou « subsampling ») de chaque échantillon, en utilisant un taux de rééchantillonnage fixé (f). Cette procédure a été proposée en 1994 par Dimitris Politis et Joseph Romano [2]. Pour chaque semaine et chaque strate géographique (anciennes régions pour les déclarations – l'incidence nationale du RS étant stratifiée par les anciennes régions et zones téléphoniques pour les prélèvements viraux), on tire au sort sans remise une fraction f de l'échantillon total disponible pour la semaine et on estime l'incidence attribuable à l'aide des données disponibles pour ce sous-échantillon.

Cette procédure est répétée *Niter* fois (*Niter* = 10 000).

On obtient un ensemble de $\hat{\theta}_{v_{\langle type \rangle}}$ estimations d'incidence attribuable

La distribution des incidences attribuables estimées pour l'ensemble des sous-échantillons est utilisée pour construire l'intervalle de confiance par approximation normale après correction pour la fraction de sous-échantillonnage.

$$\hat{I}^{sup}_{v_{\langle type \rangle}}(s) = \hat{I}_{v_{\langle type \rangle}}(s) + u_{1-\alpha} * \sqrt{f} * \sqrt{var(\hat{\theta}_{v_{\langle type \rangle}})}$$

[1] Dans l'idéal chaque cas prélevé est réalisé par un médecin ayant rapporté le volume de cas total observé ce qui n'est pas strictement le cas. Certains médecins ont prélevé mais n'ont pas rapporté ce nombre cas.

[2] Politis DN, Romano JP. Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions. The Annals of Statistics. 1994 Dec;22(4):2031-2050