

Estimation des incidences par le réseau Sentinelles

Principes généraux

Ce document décrit la méthode d'estimation des incidences utilisée au réseau Sentinelles. Cette méthode est la même quel que soit l'indicateur suivi.

Pour un indicateur donné, les données utilisées sont les déclarations faites par les médecins généralistes (MG) Sentinelles concernant le nombre de cas vus en consultation sur une période donnée. Comme expliqué dans le document « Fonctionnement du réseau Sentinelles » disponible dans la partie « Méthodes » du site Internet, les MG ont la liberté de déclarer leurs cas sur des périodes de surveillance variables ne pouvant pas excéder 12 jours. C'est pourquoi afin d'« harmoniser » les différentes déclarations des médecins et de réorganiser les données brutes en données hebdomadaires, un prétraitement des données est nécessaire. Ce prétraitement consiste à calculer, pour une semaine donnée, la participation de chaque médecin et le nombre de cas vus durant cette semaine.

Après ce prétraitement des données, l'incidence hebdomadaire peut être estimée en deux étapes :

- estimation du nombre moyen de cas vus pour un indicateur donné par MG en consultation durant cette semaine ;
- estimation du nombre total de cas vus pour un indicateur donné par l'ensemble des MG en consultation durant cette semaine, en extrapolant le nombre moyen de cas vus en consultation à l'ensemble des MG français métropolitains.

Les hypothèses permettant cette extrapolation sont les suivantes :

- **H1/** Les MG participant au réseau Sentinelles constituent un échantillon aléatoire de l'ensemble des MG français métropolitains ;
- **H2/** Les MG Sentinelles déclarent une activité représentative de leur activité hebdomadaire.

Afin de la prendre en compte et de corriger les biais liés à la participation différente des MG Sentinelles en fonction des régions ou des départements, les incidences à un niveau géographique Z (national ou régional) sont estimées en sommant les incidences calculées au niveau Z-1 (c'est à dire au niveau géographique inférieur, c'est-à-dire régional pour l'incidence nationale et départemental pour l'incidence régionale) selon la méthode décrite ci-dessus.

Toutes ces étapes sont détaillées dans ce document.

Sommaire

1. Description des données brutes	2
2. Pré-traitement des données brutes : calcul de la participation hebdomadaire et du nombre de cas imputés à chaque semaine	2
3. Estimation de l'incidence hebdomadaire pour une zone z.....	4
4. Estimation de l'incidence au niveau supérieur (englobant K zones)	8
5. Estimation de l'incidence en sous-groupes	9
6. Annexe.....	9

1. Description des données brutes

Les données recueillies par le réseau pour le médecin i sont les suivantes :

- $(t_{i,1}, t_{i,2}, \dots, t_{i,k}, \dots)$ est la suite ordonnée des dates de déclaration,
- $(c_{i,1}, c_{i,2}, \dots, c_{i,k}, \dots)$ est la suite du nombre de cas associé à chaque date de déclaration. Si le médecin fait plusieurs déclarations le même jour, cette suite est agrégée de telle sorte que des cas rapportés indépendamment le même jour ne définissent qu'une seule et même déclaration.
- Δ_{\max} est le nombre maximal de jours autorisés entre 2 déclarations du médecin (12 jours). Ainsi Δ_{ik} la durée de la $k^{\text{ème}}$ période de surveillance du médecin i (en jours) vaut :
 - soit $t_{i,k} - t_{i,k-1}$ dans le cas où $t_{i,k} - t_{i,k-1} \leq \Delta_{\max}$,
 - soit Δ_{ik} est demandé au médecin, avec la contrainte : $1 \leq \Delta_{ik} \leq \Delta_{\max}$. Autrement dit, si un médecin effectue une déclaration plus de Δ_{\max} jours après la dernière en date, on lui demande la date de début de la période de surveillance concernée par cette nouvelle déclaration. Sinon, on suppose que la période de surveillance commence le lendemain de la précédente déclaration.

2. Pré-traitement des données brutes : calcul de la participation hebdomadaire et du nombre de cas imputés à chaque semaine

Désormais on se place chronologiquement du point de vue d'une semaine s , dont l'ensemble des sept dates qui la constituent est noté $S(s)$.

On veut estimer les quantités suivantes pour le médecin i :

- 1) $n_i(s)$ le nombre des cas déclarés par le médecin i qui sont attribués à la semaine s
- 2) $p_i(s)$ la participation du médecin i pour la semaine s .

Tout d'abord pour chaque déclaration :

- on calcule son poids $W_{i,k}(s)$ défini comme le nombre de jours concernés par la déclaration

qui appartiennent à la semaine s :

$$W_{i,k}(s) = \sum_{j=t_{i,k}-\Delta_{i,k}+1}^{t_{i,k}} 1(j \in S(s)), \text{ pour } k \geq 2.$$

- le nombre de cas déclarés est réparti uniformément sur toute la durée concernée par la déclaration. Cela permet d'imputer un nombre de cas à la semaine s comme suit :

$$n_{i,k}(s) = \frac{c_{i,k}}{\Delta_{i,k}} W_{i,k}(s)$$

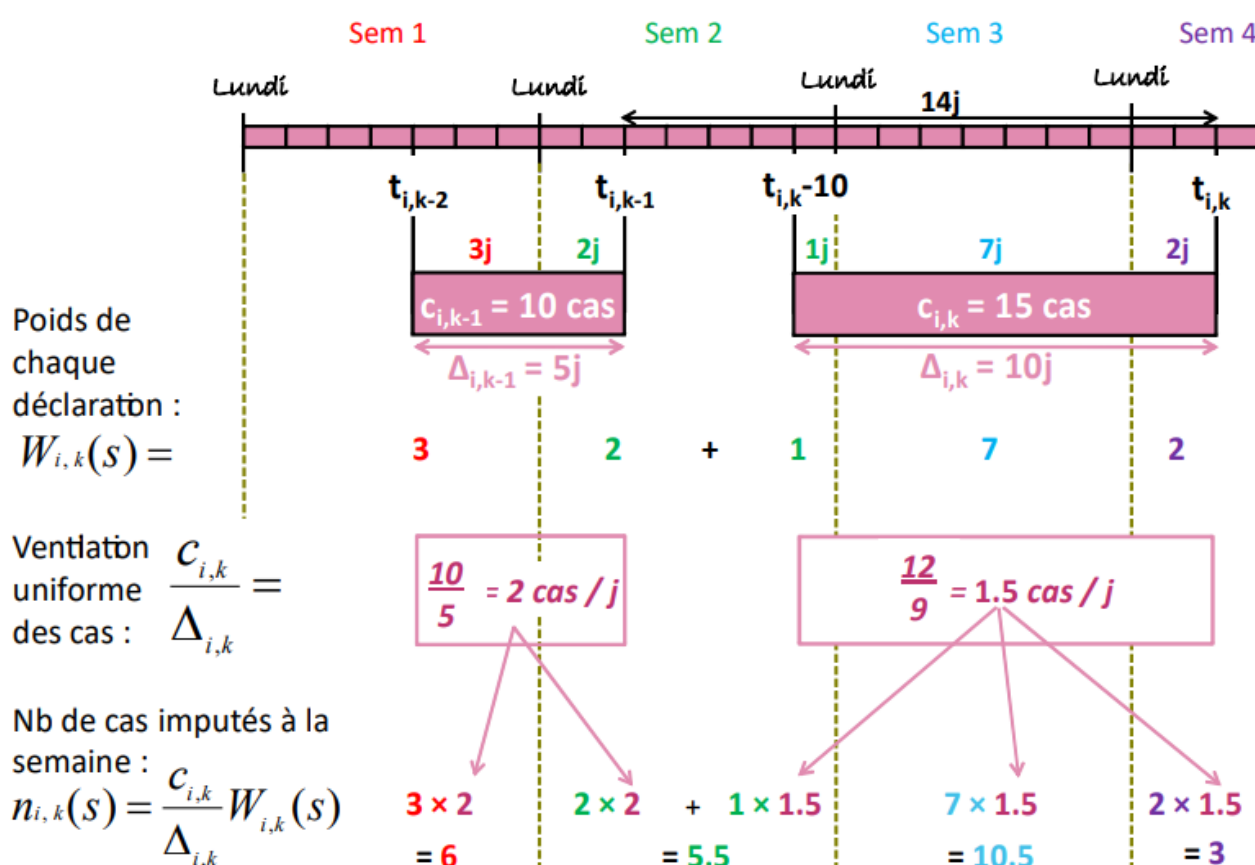


Figure 1 : transformation des déclarations des médecins Sentinelles en données hebdomadaires pour préparer le calcul des incidences, réseau Sentinelles

Ce schéma illustre le pré-traitement des déclarations d'un médecin. Sur ce schéma trois dates de connexion sont représentées : le jeudi de la semaine 1 à $t_{i,k-2}$, le mardi de la semaine 2 $t_{i,k-1}$ et le mardi de la semaine 4 à $t_{i,k}$.

- Lors de la première connexion à $t_{i,k-2}$ le médecin a fait une déclaration pour les jours passés qui n'est pas explicitée ici.
- La connexion suivante est effectuée cinq jours plus tard soit le mardi de la semaine 2 au temps $t_{i,k-1}$: le médecin y déclare l'observation de dix cas depuis sa dernière connexion. Les dix cas observés sur cette période de cinq jours sont donc répartis uniformément sur cette période à raison de deux cas par jours. Parmi ces cinq derniers jours, trois appartiennent à la

semaine 1 et deux à la semaine 2 donc avec la ventilation de deux cas par jour, les dix cas observés sont affectés aux semaines 1 et 2 de la façon suivante : six cas sont affectés à la semaine 1 et quatre cas à la semaine 2.

- Ensuite le médecin ne se reconnecte à $t_{i,k}$ que le mardi deux semaines plus tard (semaine 4). Comme sa dernière connexion à $t_{i,k-1}$ date de plus de douze jours, on lui demande d'indiquer quelle période concerne cette déclaration. Il déclare qu'au cours des dix derniers jours c'est à dire entre $t_{i,k} - 10$ et $t_{i,k}$ il a observé quinze cas. Comme précédemment, on peut voir sur le schéma que ces dix derniers jours couvrent les semaines 2, 3 et 4. Ainsi les quinze cas déclarés, avec une répartition uniforme d'un cas et demi par jour, sont affectés aux semaines 2, 3 et 4 selon le nombre de jours de la semaine couverts par cette déclaration.

Pour la répartition des cas d'une semaine à l'autre, on considère que chaque jour pèse le même poids, sans prendre en compte des absences brèves potentielles comme celle des week-ends, ou la fraction de temps réellement travaillé par jour.

Ainsi on peut calculer pour chaque médecin :

- 1) le nombre total de cas attribués à la semaine s pour le médecin i (dernière ligne sur le schéma) :

$$n_i(s) = \sum_k n_{i,k}(s)$$

- 2) $p_i(s)$ la participation du médecin i à la surveillance de la semaine s . C'est la somme des jours appartenant à la semaine s sur toutes ses déclarations, divisée par le nombre de jours de la semaine s :

$$p_i(s) = \frac{1}{\text{card}(s(s))} \sum_k W_{i,k}(s)$$

La participation d'un médecin à la surveillance de la semaine s est comprise entre 0 et 1. Elle vaut 1 si ses déclarations couvrent tous les jours de la semaine (c'est le cas du médecin sur la semaine 3 dans l'exemple) et sa participation vaut 3/7 s'il n'a surveillé que trois jours de la semaine, que ce soit en une seule déclaration (la participation du médecin à la semaine 1 est 3/7 dans l'exemple) ou en plusieurs déclarations (dans l'exemple, la participation à la semaine 2 est 3/7 puisque ses deux déclarations couvrent au total 3 jours sur cette semaine).

3. Estimation de l'incidence hebdomadaire pour une zone z

a. Estimation du nombre moyen de cas par médecin de la zone, appelé $\lambda_z(s)$

Soient les constantes :

- $d_z(s)$ le nombre de médecins qui ont fait une ou plusieurs déclarations couvrant la semaine s dans la zone z ,
- $m_z(s)$ le nombre total de médecins exerçant dans la zone z , au cours de la semaine s .

et soient les variables aléatoires propres à chaque médecin i de la zone z , $i \in \{1, \dots, d_z(s)\}$:

- $N_{i,z}(s)$ le nombre des cas déclarés par le médecin i qui sont attribués à la semaine s
- $P_{i,z}(s)$ la participation du médecin i pour la semaine s .

On suppose que :

- les $N_{i,z}(s)$ et les $P_{i,z}(s)$ sont des variables aléatoires indépendantes et identiquement distribuées pour chaque médecin i de la semaine s , de la zone z .
- $N_{i,z}(s) \mid (P_{i,z}(s) = p_{i,z}(s))$ suit une loi de Poisson de paramètre $\lambda_{i,z}(s)p_{i,z}(s)$.

D'après l'hypothèse précédente, on a de plus : $\forall i \in \{1, \dots, d_z(s)\}, \lambda_{i,z}(s) = \lambda_z(s)$,

d'où $\forall i \in \{1, \dots, d_z(s)\}, N_{i,z} \mid (P_{i,z} = p_{i,z}) \sim P(\lambda_z(s)p_{i,z})$

Notons que si $p_{i,z}(s) = 1$, alors $N_{i,z}(s) \mid (P_{i,z}(s) = 1) \sim P(\lambda_z(s))$, c'est à dire que $\lambda_z(s)$ correspond à l'espérance du nombre de cas vus par un médecin de la zone z s'il surveille les 7 jours de la semaine s .

Calcul de l'estimateur de $\lambda_z(s)$:

Par définition de la loi de Poisson on a $E[N_{i,z}(s) \mid P_{i,z}(s)] = \lambda_z(s)P_{i,z}(s)$

Or par définition de l'espérance conditionnelle $E[N_{i,z}(s)] = E[E[N_{i,z}(s) \mid P_{i,z}(s)]]$

ce qui nous permet d'écrire $E[N_{i,z}(s)] = E[\lambda_z(s)P_{i,z}(s)] = \lambda_z(s)E[P_{i,z}(s)]$

ou encore : $\lambda_z(s) = \frac{E[N_{i,z}(s)]}{E[P_{i,z}(s)]}$.

Puisqu'un estimateur de $E[N_{i,z}(s)]$ est la moyenne arithmétique $\bar{N}_z(s) = \frac{1}{d_z(s)} \sum_{i=1}^{d_z(s)} N_{i,z}(s)$

et qu'un estimateur de $E[P_{i,z}(s)]$ est la moyenne arithmétique $\bar{P}_z(s) = \frac{1}{d_z(s)} \sum_{i=1}^{d_z(s)} P_{i,z}(s)$,

il vient naturellement qu'on peut estimer $\lambda_z(s)$ par :

$$\hat{\lambda}_z(s) = \frac{\bar{N}_z(s)}{\bar{P}_z(s)} = \frac{\frac{1}{d_z(s)} \sum_{i=1}^{d_z(s)} N_{i,z}(s)}{\frac{1}{d_z(s)} \sum_{i=1}^{d_z(s)} P_{i,z}(s)} = \frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{\sum_{i=1}^{d_z(s)} P_{i,z}(s)}$$

En annexe on peut lire la vérification du fait que $\hat{\lambda}_z(s)$ est un estimateur sans biais de $\lambda_z(s)$.

Variance de l'estimateur $\hat{\lambda}_z(s)$ conditionnellement à $\tilde{P}_z(s)$:

La variabilité de $\hat{\lambda}_z(s)$ est considérée conditionnellement à l'observation du vecteur des participations dans la zone z , la semaine s , $\tilde{P}_z(s) = [P_{1,z}(s), \dots, P_{d_z,z}(s)]$.

Par définition de l'estimateur $\hat{\lambda}_z(s)$ on a :

$$\text{var}[\hat{\lambda}_z(s) \mid (\tilde{P}_z(s) = \tilde{p}_z(s))] = \text{var}\left[\frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{d_z} \mid (\tilde{P}_z(s) = \tilde{p}_z(s))\right] = \text{var}\left[\frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{d_z} \mid (\tilde{P}_z(s) = \tilde{p}_z(s))\right]$$

soit

$$var[\hat{\lambda}_z(s) | (\tilde{P}_z(s) = \tilde{p}_z(s))] = \frac{1}{\left(\sum_{i=1}^{d_z(s)} p_{i,z}(s)\right)^2} var\left[\sum_{i=1}^{d_z(s)} N_{i,z}(s) | (\tilde{P}_z(s) = \tilde{p}_z(s))\right]$$

Comme les $N_{i,z}(s)$ suivent une loi de Poisson conditionnellement à $P_{i,z}(s)$ leur somme suit également une loi de Poisson conditionnellement à $\tilde{P}_z(s)$, vecteur observé des $P_{i,z}(s)$:

$$\sum_{i=1}^{d_z(s)} N_{i,z}(s) | (\tilde{P}_z(s) = \tilde{p}_z(s)) \sim P(\lambda_z(s) \sum_{i=1}^{d_z(s)} p_{i,z}(s))$$

Ce qui implique nécessairement $var\left[\sum_{i=1}^{d_z(s)} N_i | (\tilde{P}_z(s) = \tilde{p}_z(s))\right] = \lambda_z(s) \sum_{i=1}^{d_z(s)} p_{i,z}(s)$.

Ainsi on obtient, par définition de $\hat{\lambda}_z(s)$:

$$var[\hat{\lambda}_z(s) | (\tilde{P}_z(s) = \tilde{p}_z(s))] = \frac{1}{\left(\sum_{i=1}^{d_z(s)} p_{i,z}(s)\right)^2} var\left[\sum_{i=1}^{d_z(s)} N_{i,z}(s) | (\tilde{P}_z(s) = \tilde{p}_z(s))\right]$$

ou encore :

$$var[\hat{\lambda}_z(s) | (\tilde{P}_z(s) = \tilde{p}_z(s))] = \frac{1}{\left(\sum_{i=1}^{d_z(s)} p_{i,z}(s)\right)^2} \lambda_z(s) \sum_{i=1}^{d_z(s)} p_{i,z}(s) = \frac{1}{\left(\sum_{i=1}^{d_z(s)} p_{i,z}(s)\right)} \lambda_z(s)$$

Finalement on peut estimer $var[\hat{\lambda}_z(s) | (\tilde{P}_z(s) = \tilde{p}_z(s))]$ par $\frac{1}{\sum_{i=1}^{d_z(s)} p_i} \hat{\lambda}_z(s) = \frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{\left(\sum_{i=1}^{d_z(s)} p_{i,z}(s)\right)^2}$.

Intervalle de confiance de $\lambda_z(s)$:

On sait que $\sum_{i=1}^{d_z(s)} N_{i,z}(s) | (\tilde{P}_z(s) = \tilde{p}_z(s)) \sim P(\lambda_z(s) \sum_{i=1}^{d_z(s)} p_{i,z}(s))$.

Le paramètre $\lambda_z(s) \sum_{i=1}^{d_z(s)} p_{i,z}(s)$ de cette loi correspond au nombre attendu de cas déclarés par les médecins du réseau sur la zone concernée.

A supposer que ce paramètre soit supérieur ou égal à 25, il est raisonnable de faire l'approximation suivante de la loi de Poisson par une loi normale :

$$\sum_{i=1}^{d_z(s)} N_{i,z}(s) | (\tilde{P}_z(s) = \tilde{p}_z(s)) \sim N(\lambda_z(s) \sum_{i=1}^{d_z(s)} p_{i,z}(s), \lambda_z(s) \sum_{i=1}^{d_z(s)} p_{i,z}(s)).$$

Concernant l'estimateur $\hat{\lambda}_z(s) = \frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{\sum_{i=1}^{d_z(s)} p_{i,z}(s)}$ on peut donc faire l'hypothèse que sa distribution

conditionnellement à $\tilde{P}_z(s)$ est la suivante :

$$\hat{\lambda}_z(s) \mid (\tilde{P}_z(s) = \tilde{p}_z(s)) \sim N \left(\lambda_z(s), \frac{1}{\sum_{i=1}^{d_z(s)} p_{i,z}(s)} \lambda_z(s) \right).$$

On retrouve bien le fait que cet estimateur est sans biais et que sa variance est égale à celle calculée plus haut.

Finalement, avec cette approximation normale, l'intervalle de confiance de niveau $(1-\alpha)$ de $\lambda_z(s)$ peut être donné par :

$$\left[\hat{\lambda}_z(s) \pm u_\alpha \sqrt{\text{var}(\hat{\lambda}_z(s) \mid (\tilde{P}_z(s) = \tilde{p}_z(s)))} \right]$$

$$\text{i.e. } \left[\hat{\lambda}_z(s) \pm u_\alpha \sqrt{\frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{(\sum_{i=1}^{d_z(s)} p_{i,z}(s))^2}} \right] \text{ ou encore } \left[\frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{\sum_{i=1}^{d_z(s)} p_{i,z}(s)} \pm u_\alpha \sqrt{\frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{\sum_{i=1}^{d_z(s)} p_{i,z}(s)}} \right]$$

Attention, en pratique lorsqu'on a moins de 25 cas déclarés sur une zone on peut donc s'attendre à ce que l'intervalle de confiance de l'incidence estimée soit très grand du fait que l'hypothèse n'est pas vérifiée et que donc l'approximation normale n'est donc pas tout à fait valable.

b. Estimation de $INC_z(s)$ l'incidence hebdomadaire sur la zone

Calcul de l'estimateur $INC_z(s)$:

Soit $INC_z(s) = \lambda_z(s) \times m_z(s)$, l'incidence attendue sur la zone z , si chacun des $m_z(s)$ médecins exerçant dans la zone z avait surveillé 7 jours pleins.

Notons que dans le calcul de $INC_z(s)$, l'hypothèse est faite que les $m_z(s)$ médecins exerçant dans la zone z déclareraient, s'ils participaient à la surveillance, des nombres de cas suivant la même loi que les $d_z(s)$ médecins participant au réseau de surveillance.

Un estimateur de $INC_z(s)$ est donc $\hat{INC}_z(s) = m_z(s) \times \hat{\lambda}_z(s) = m_z(s) \times \frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{\sum_{i=1}^{d_z(s)} p_{i,z}(s)}$

En l'absence de données hebdomadaires sur le nombre de médecins exerçant dans chaque zone, on considère $m_z(s)$ constant au fil des semaines, c'est à dire qu'on fait comme si le nombre de médecins en activité dans la zone z était le même tout au long de l'année (en réalité moins de médecins travaillent en saison estivale et pendant les vacances de Noël par exemple donc les médecins sont peut-être surchargés et l'incidence surestimée dans ces cas-là).

Intervalle de confiance de $INC_z(s)$ conditionnellement à $\tilde{P}_z(s)$:

De même, $INC_z(s)$ l'espérance de l'incidence en médecine de ville dans la zone z durant la semaine

s, conditionnellement à l'observation des P_i a 95% de chances de se trouver dans l'intervalle :

$$\left[m_z(s) \frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{\sum_{i=1}^{d_z(s)} p_{i,z}(s)} \pm 1,96 \times m_z(s) \frac{\sqrt{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}}{\sum_{i=1}^{d_z(s)} p_{i,z}(s)} \right]$$

4. Estimation de l'incidence au niveau supérieur (englobant K zones)

Le découpage en zones peut être celui de la France en K régions administratives ou bien celui d'une région administrative en K départements. On définit $INC_G(s)$ l'incidence globale sur le territoire englobant K zones.

Estimateur de $INC_G(s)$ conditionnellement aux $\tilde{P}_z(s)$

Un estimateur de $INC_G(s)$ l'incidence globale est $\hat{INC}_G(s) = \sum_{z=1}^K \hat{INC}_z(s)$ avec $\hat{INC}_z(s) (z = 1, \dots, K)$ les incidences estimées par zone.

Intervalle de confiance de $\hat{INC}_G(s)$ conditionnellement aux $\tilde{P}_z(s)$

On suppose normale la distribution de l'estimateur $\hat{\lambda}_z(s)$ à l'intérieur de chaque zone z , conditionnellement à l'observation de $\tilde{P}_z(s)$:

$$\hat{\lambda}_z(s) \mid (\tilde{P}_z(s) = \tilde{p}_z(s)) \sim N \left(\lambda_z(s), \text{var} \left[\hat{\lambda}_z(s) \mid (\tilde{P}_z(s) = \tilde{p}_z(s)) \right] \right)$$

Or $\hat{INC}_z(s) = \hat{\lambda}_z(s) \times m_z(s)$

D'où $\hat{INC}_z(s) \mid (\tilde{P}_z(s) = \tilde{p}_z(s)) \sim N \left(m_z(s) \times \lambda_z(s), m_z(s)^2 \times \text{var} \left[\hat{\lambda}_z(s) \mid (\tilde{P}_z(s) = \tilde{p}_z(s)) \right] \right)$

La somme $\sum_{z=1}^K Y_z$ de plusieurs variables Y_z qui suivent des lois normales de paramètres (μ_k, σ_k) suit également une loi normale de paramètres : $\left(\sum_{k=1}^{K_z} \mu_k, \sum_{k=1}^{K_z} \sigma_k \right)$.

Donc $\hat{INC}_G(s) = \sum_{z=1}^K \hat{INC}_z(s) \sim \left(\sum_{z=1}^K m_z(s) \times \lambda_z(s), \sum_{z=1}^K m_z(s)^2 \times \text{var} \left[\hat{\lambda}_z(s) \mid (\tilde{P}_z(s) = \tilde{p}_z(s)) \right] \right)$

En utilisant les estimateurs

$$\hat{\lambda}_z(s) \mid (\tilde{P}_z(s) = \tilde{p}_z(s)) = \frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{\sum_{i=1}^{d_z(s)} p_{i,z}(s)} \text{ et } \text{var} \left[\hat{\lambda}_z(s) \mid (\tilde{P}_z(s) = \tilde{p}_z(s)) \right] = \frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{\left(\sum_{i=1}^{d_z(s)} p_{i,z}(s) \right)^2},$$

l'intervalle à 95% de $\hat{INC}_G(s)$ conditionnellement aux $\tilde{P}_z(s)$ peut être estimé par :

$$\left[\sum_{z=1}^K \left(m_z(s) \times \frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{\sum_{i=1}^{d_z(s)} p_{i,z}(s)} \right) \pm 1,96 \sqrt{\sum_{z=1}^K \left(m_z(s)^2 \times \frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{\left(\sum_{i=1}^{d_z(s)} p_{i,z}(s) \right)^2} \right)} \right]$$

5. Estimation de l'incidence en sous-groupes

Lorsque les médecins Sentinelles déclarent les cas vus durant leurs consultations, il leur est également demandé de décrire ces cas à l'aide d'un questionnaire individuel comprenant différentes variables pour chaque patient (âge, sexe, facteurs de risque, ...). Ainsi, en utilisant ces données, il est possible d'estimer la proportion des cas déclarés appartenant à un sous-groupe spécifique, comme par exemple à une classe d'âge donnée.

Toutefois les cas déclarés par les médecins Sentinelles ne sont pas systématiquement décrits pour l'ensemble des variables demandées. Les cas décrits constituent un sous-échantillon de l'ensemble des cas déclarés par les médecins Sentinelles. En faisant l'hypothèse que cet échantillon est représentatif de l'ensemble des cas déclarés, on peut considérer que la proportion observée pour un sous-groupe donné parmi les cas décrits est égale à celle qui aurait été observée parmi l'ensemble des cas déclarés.

Pour estimer les incidences pour un sous-groupe donné, le nombre moyen de cas déclarés par médecin durant une semaine s dans une zone z (ci-dessus appelé $\hat{\lambda}_z(s)$), est multipliée par la proportion des cas déclarés appartenant à ce sous-groupe dans cette zone et durant cette semaine. La méthode du calcul des incidences est ensuite la même que celle précédemment décrite.

En pratique, afin d'augmenter la fiabilité de l'estimation de la proportion des cas déclarés appartenant à un sous-groupe spécifique, en fonction des situations, ce taux peut être calculé pour une semaine s au niveau national et appliqué à l'ensemble des zones z , en faisant l'hypothèse que cette proportion ne change pas en fonction des zones géographiques.

6. Annexe

Vérifions que $\hat{\lambda}_z(s)$ est bien un estimateur sans biais de $\lambda_z(s)$ ce qui signifie que :

$$E[\hat{\lambda}_z(s)] = \lambda_z(s)$$

Notons $\tilde{P}_z(s)$ le vecteur observé des $P_{i,z}(s)$.

Par propriété de l'espérance on a $E[\hat{\lambda}_z(s)] = E[E[\hat{\lambda}_z(s) | (\tilde{P}_z(s) = \tilde{p}_z(s))]]$

et par définition de $\hat{\lambda}_z(s)$, $E[\hat{\lambda}_z(s) | P_{i,z}(s)] = E\left[\frac{\sum_{i=1}^{d_z(s)} N_{i,z}(s)}{\sum_{i=1}^{d_z(s)} P_{i,z}(s)} \mid (\tilde{P}_z(s) = \tilde{p}_z(s))\right]$

Comme les $N_{i,z}$ sont indépendantes et identiquement distribuées de loi $P(\lambda_z(s)p_{i,z})$ on obtient, encore par propriété de l'espérance :

$$E[\hat{\lambda}_z(s) | P_{i,z}(s)] = \frac{1}{\sum_{i=1}^{d_z(s)} p_{i,z}(s)} \lambda_z(s) \sum_{i=1}^{d_z(s)} p_{i,z}(s) = \lambda_z(s)$$

d'où par passage à l'espérance : $E[\hat{\lambda}_z(s)] = E[\lambda_z(s)] = \lambda_z(s)$.