

Prise en compte de la fréquentation multiple des lieux d'enquêtes - Illustration sur Coquelicot 2011, une enquête auprès des usagers de drogues

Lucie Léon

Département des Maladies Infectieuses
Institut de Veille Sanitaire (InVS)
Saint-Maurice, France
www.invs.sante.fr

Épiter - 27 septembre 2013

Plan

1. Enquête ANRS-Coquelicot 2011
2. Sondage Indirect
3. Méthode Généralisée du Partage des Poids (MGPP)
4. Étude de simulation
5. Discussion-Conclusion

Coquelicot 2011 - Description

▶ Objectifs

- ▶ Estimer la séroprévalence du VIH et du VHC chez les usagers de drogues (UD)
- ▶ Caractériser la population d'étude d'un point de vue sociodémographique
- ▶ Décrire les pratiques d'usage de drogues.

▶ Population

- ▶ Les UD fréquentant des structures spécialisées et bénéficiant de prestations
- ▶ 5 agglomérations (Lille, Strasbourg, Paris, Bordeaux, Marseille) et 2 départements (Seine et Marne, Seine Saint-Denis)
- ▶ en 2011

Coquelicot 2011 - Structures spécialisées

- ▶ **Services dits à bas-seuil d'exigence** : boutiques, programmes d'échanges de seringues, équipes de rue = **CAARUD** (Centre d'Accueil et d'Accompagnement à la Réduction des Risques pour les Usagers de Drogues) depuis 2008.
- ▶ **Consultations** : centres de soins spécialisés pour toxicomanes, services hospitaliers, médecins généralistes gros et moyens prescripteurs de Buprénorphine Haut Dosage.
- ▶ **Hébergement**: post-cures, appartements thérapeutiques, hôtels payés par des associations.

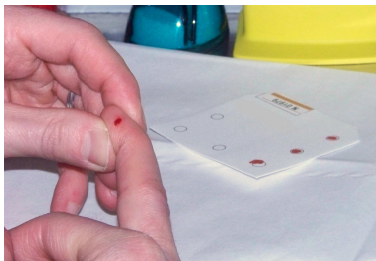
Coquelicot 2011 - Prestations

Un service à la personne

- ▶ hébergement
- ▶ repas
- ▶ échange de matériels
- ▶ consultation médicale
- ▶ ...

Coquelicot 2011 - Caractéristiques

- ▶ Seconde édition (première enquête réalisée en 2004)
[M. Jauffret-Roustide, Y. Le Strat, 2010]
- ▶ Enquête transversale avec un recueil de prélèvements biologiques



- ▶ **Fréquentation multiple et hétérogène des UD dans les structures spécialisées**

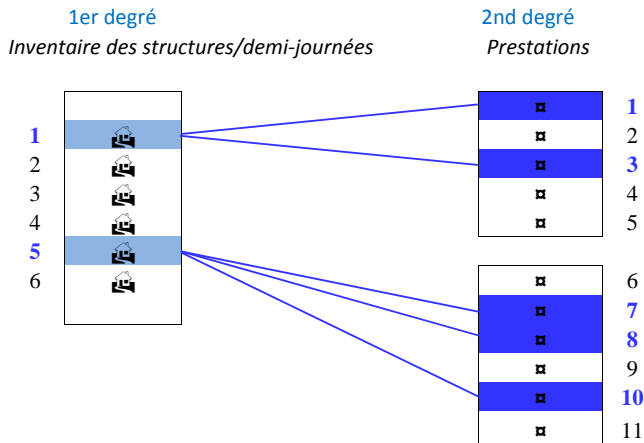
Coquelicot 2011 - Plan de sondage

- ▶ Les UD ont été recrutés dans les lieux d'enquête
- ▶ Plan de sondage à 2 degrés
 - ▶ 1er degré : tirage au sort dans un inventaire listant l'ensemble des demi-journées d'ouverture des structures participantes selon un sondage aléatoire simple (SAS)
 - ▶ 2nd degré : recrutement aléatoire des UD dans les structures/demi-journées tirées au sort en utilisant un pas de sondage adapté.
- ▶ peut être vu comme un échantillonnage de lieux-moments (Time-Location Sampling (TLS))
[M. Marpsat, N. Razafindratsima, 2010]

Coquelicot 2011 - Echantillonnage lieux-moments (TLS)

- ▶ Méthode d'échantillonnage utilisée pour des études dans des populations difficiles à joindre en échantillonnant des personnes dans des lieux qu'elles fréquentent.
- ▶ Son principe:
 1. échantillonner des lieux (structures spécialisées, places ...)
et des moments (intervalles de temps, jours, demi-journées...)
 2. échantillonner des personnes fréquentant ces lieux-moments.
- ▶ Une base de sondage est ainsi construite en amont de l'enquête en listant l'ensemble des lieux possibles et des moments possibles (une liste de couples "structures/demi-journées" pour Coquelicot)

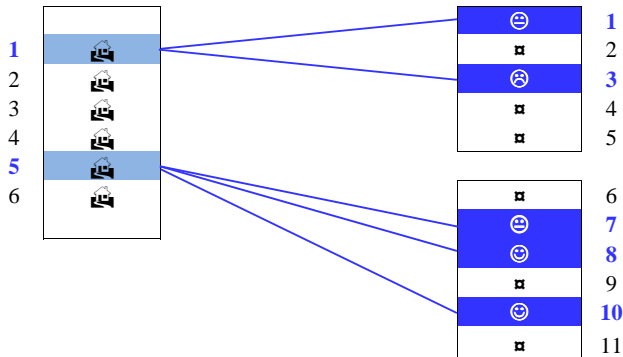
Coquelicot 2011 - Plan de sondage



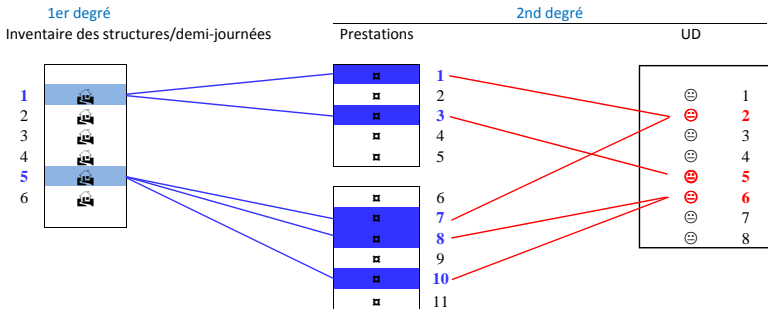
Coquelicot 2011 - Plan de sondage

1er degré
Inventaire des structures/demi-journées

2nd degré
Prestations



Coquelicot 2011 - Plan de sondage



Contexte "classique" vs Coquelicot 2011

Comment calculer la probabilité d'inclusion d'un UD dans Coquelicot compte tenu de la fréquentation multiple?

	Enquêtes "classiques"	Coquelicot 2011
Base de sondage	disponible	pas disponible
Accès	direct	indirect
Unités d'intérêt	identifiants uniques	plusieurs identifiants
Probabilité d'inclusion	calcul facile	calcul difficile
Poids de sondage	calcul facile	calcul difficile

Cette question permet d'introduire la notion de **sondage indirect** et la **Méthode Généralisée du Partage de Poids (MGPP)**
[P. Lavallée, 2002]

Sondage Indirect

- Soient une **population A** de taille N^A où chaque unité est indexée par j ($j = 1, \dots, N^A$) et une **population d'intérêt B** de taille N^B où chaque unité d'intérêt est indexée par i ($i = 1, \dots, N^B$).

Population A
Prestations

	j
▣	1
▣	2
▣	3
▣	4
▣	5
▣	6
▣	7
▣	8
▣	9
▣	10
▣	11
▣	12

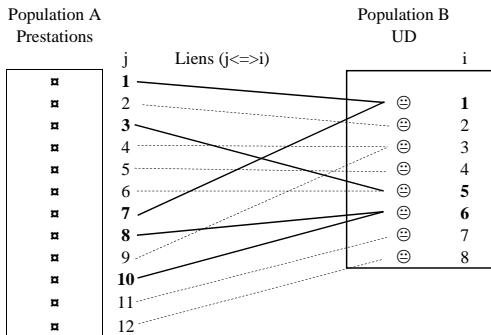
Population B
UD

	i
☹	1
☹	2
☹	3
☹	4
☹	5
☹	6
☹	7
☹	8

Sondage Indirect

Un **sondage indirect** définit un sondage pour lequel :

- ▶ des unités j sont tirées au sort dans une population A afin d'accéder à des unités i appartenant à la population d'intérêt B et
- ▶ les unités i dans la population B sont liées à au moins une unité j appartenant à la population A .



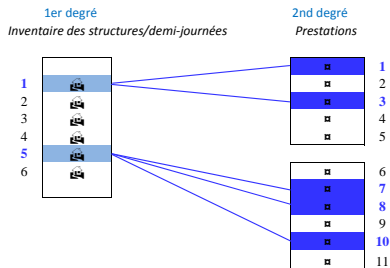
MGPP - Contexte du sondage indirect

- ▶ Difficile d'associer une probabilité d'inclusion aux unités de la population d'intérêt dans un sondage indirect
- ▶ Problème résolu grâce à la MGPP développée par Lavallée à partir de 1995
- ▶ Elle fournit un poids de sondage pour chaque unité enquêtée
- ▶ Elle nécessite 5 étapes.

MGPP - Etape 1

Pour chaque unité $j \in s^A$, on calcule son poids de sondage:

$$w_j = 1/\pi_j.$$



En Pratique:

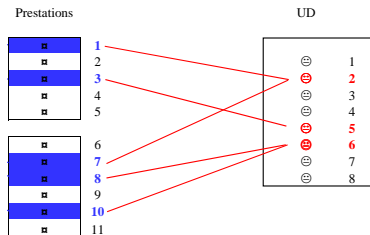
Au 1er degré, la probabilité d'inclusion d'un couple structures/demi-journées est de $1/3$. L'échantillon des prestations est $s^A = \{1, 3, 7, 8, 10\}$.

- ▶ pour les prestations issues du couple 1 : $w_{j|1} = 3 \times (5/2)$
- ▶ pour les prestations issues du couple 5 : $w_{j|5} = 3 \times (6/3) = 6$

MGPP - Etape 2

Pour chaque unité $i \in s^B$ (ici, $s^B = \{2, 5, 6\}$), on définit ses liens l_{ji} non nuls avec les unités j de A :

$$l_{ji} = \begin{cases} 1 & \text{si } i \text{ a un lien avec } j \in s^A \\ 0 & \text{sinon.} \end{cases}$$

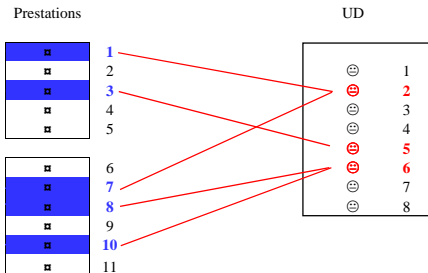


- ▶ pour l'UD 2 : $l_{12} = l_{72} = 1$
- ▶ pour l'UD 5 : $l_{35} = 1$
- ▶ pour l'UD 6 : $l_{86} = l_{106} = 1$

MGPP - Etape 3

Pour chaque unité $i \in s^B$, on calcule son poids de sondage initial w'_i et par définition de l_{ji} , on a :

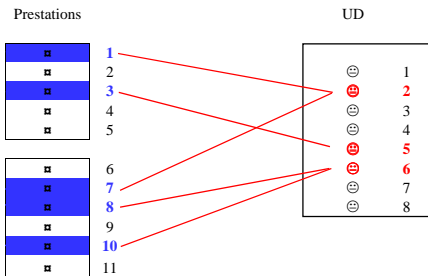
$$w'_i = \sum_{j \in A} l_{ji} w_j = \sum_{j \in s^A} l_{ji} w_j.$$



- ▶ pour l'UD 2 : $w'_2 = l_{12} w_{1|1} + l_{72} w_{7|5} = w_{1|1} + w_{7|5}$
- ▶ pour l'UD 5 : $w'_5 = l_{35} w_{3|1} = w_{3|1}$
- ▶ pour l'UD 6 : $w'_6 = l_{86} w_{8|5} + l_{106} w_{10|5} = 2w_{8|5}$

MGPP - Etape 4

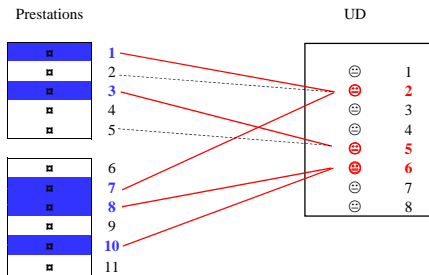
Pour chaque unité $i \in s^B$, on calcule son nombre total de liens avec les unités j de A qu'elles aient été ou non tirées au sort, noté $L_i^B = \sum_{j \in A} l_{ji}$.



- ▶ pour l'UD 2 : $L_2^B = 2$
- ▶ pour l'UD 5 : $L_5^B = 1$
- ▶ pour l'UD 6 : $L_6^B = 2$

MGPP - Etape 4 bis

.... qu'elles aient été ou non tirées au sort. Si on avait,

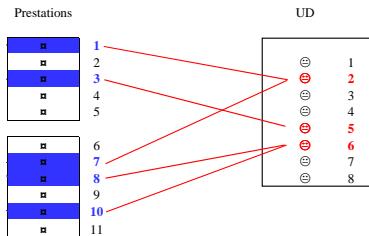


On aurait dans ce cas:

- ▶ pour l'UD 2 : $L_2^B = 3$
- ▶ pour l'UD 5 : $L_5^B = 2$
- ▶ pour l'UD 6 : $L_6^B = 2$

MGPP - Etape 5

Pour chaque unité $i \in s^B$, son poids de sondage final \tilde{w}_i se calcule par: $\tilde{w}_i = \frac{1}{L_i^B} \sum_{j \in s^A} l_{ji} w_j$.



- ▶ pour l'UD 2 : $\tilde{w}_2 = \frac{1}{L_2^B} (w_{1|1} + w_{7|5}) = \frac{1}{2} (w_{1|1} + w_{7|5})$
- ▶ pour l'UD 5 : $\tilde{w}_5 = \frac{1}{L_5^B} w_{3|1} = w_{3|1}$
- ▶ pour l'UD 6 : $\tilde{w}_6 = \frac{1}{L_6^B} 2w_{8|5} = w_{8|5}$

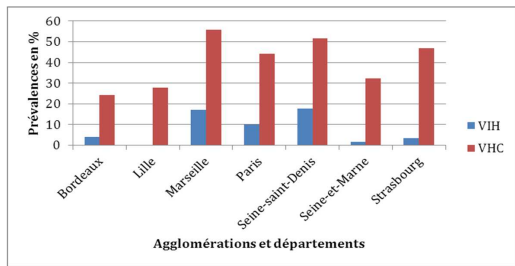
Application de la MGPP dans Coquelicot 2011

- ▶ Entre mai et juillet 2011, 1568 UD dans 121 structures spécialisées; 75% de taux de participation; 92% d'auto-prélèvements de sang.
- ▶ La séroprévalence du VHC est toujours élevée (44%) alors que celle du VIH est de 10%.

Répartition de la fréquentation des structures

Nombre total de liens	Fréquence	Pourcentage
1	1342	85,59
2	130	8,29
3	40	2,55
4	25	1,59
5	15	0,96
6	11	0,7
7	2	0,13
9	2	0,13
14	1	0,06
Total	1568	100

Séroprévalence du VIH et du VHC



Étude de simulation - Objectifs

Objectif : Évaluer l'impact de la MGPP dans une enquête épidémiologique auprès de personnes bénéficiant de prestations en fonction de scénarios définis selon:

- ▶ la prévalence
(1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%)
- ▶ le nombre de liens (dépendant ou non du statut sérologique des individus) suivant une distribution a priori
- ▶ la qualité des liens

Étude de simulation - Population et plan de sondage

► Population

Des populations ont été simulées pour chaque scénario.
154 scénarios sont en cours et d'autres sont envisagés.

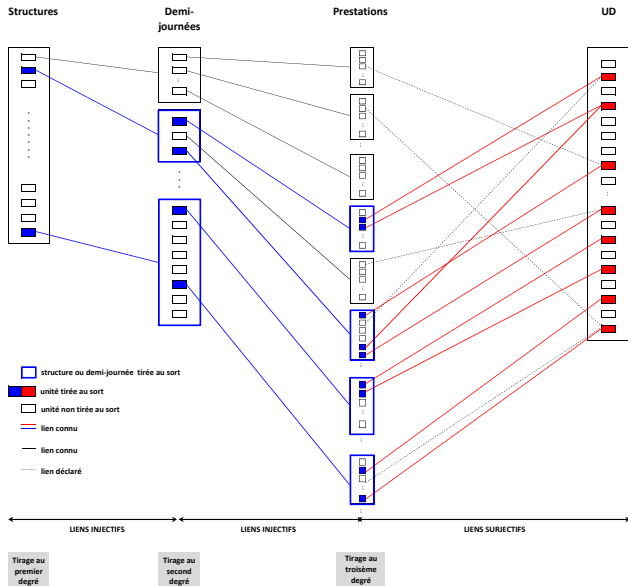
► Plan de sondage

un TLS ou sondage indirect à 3 degrés est réalisé avec un SAS
à chaque degré:

- 1er degré: tirage au sort des structures
- 2nd degré: tirage au sort des demi-journées d'ouverture
- 3ième degré: tirage au sort des prestations (rattachées aux UD)

- 10000 échantillons sont tirés au sort dans chaque scénario

Étude de simulation - Schéma

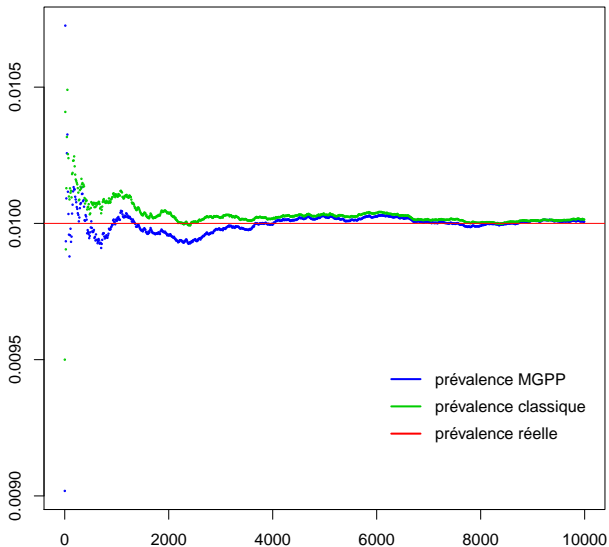


Étude de simulation- Premiers résultats

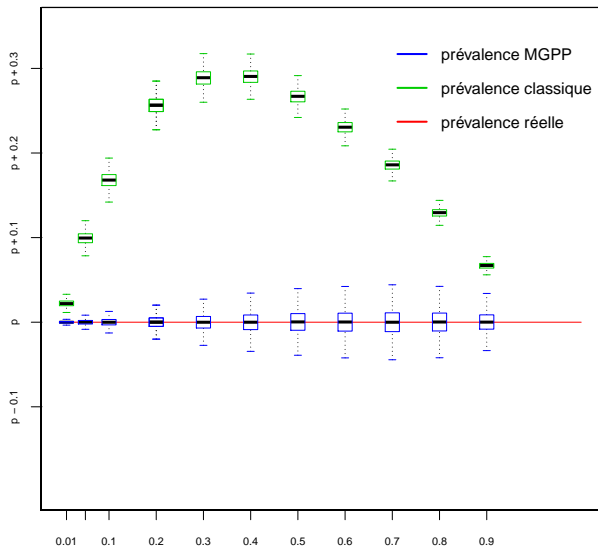
Deux méthodes de calcul des poids de sondage sont comparées: la méthode dite "classique" et la MGPP:

- ▶ en fonction du nombre d'échantillons simulés
- ▶ en fonction du statut sérologique dépendant ou non des fréquentations des UD
- ▶ en fonction du biais relatif

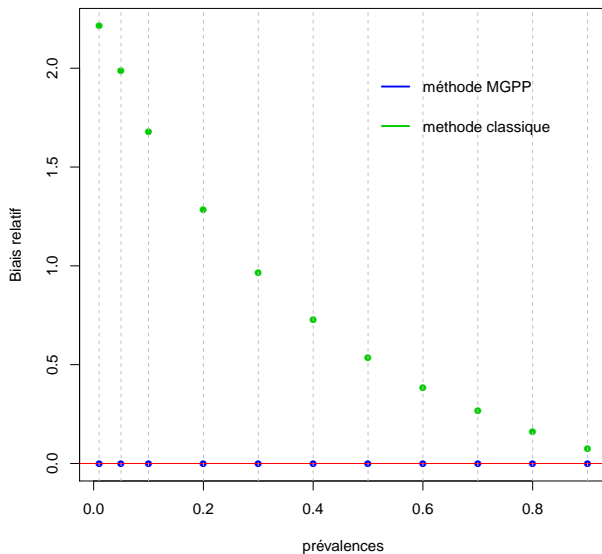
Étude de simulation - Prévalence = 1%



Étude de simulation - Fréquentations et statut sérologique



Étude de simulation - Mesure du biais relatif



Conclusion / Discussion

▶ TLS

- ▶ Sondage indirect à deux degrés voire trois degrés
- ▶ Des avis différents sur le caractère aléatoire de l'échantillonnage lieux-moments. [Meyer, 2009; Xia, 2013]

▶ MGPP

- ▶ Il faut tenir compte de la fréquentation multiple des lieux d'enquête dans le calcul des poids de sondage
- ▶ La MGPP est une solution et permet de fournir un estimateur épidémiologique sans biais
- ▶ La littérature épidémiologique existante soulève le problème des fréquentations multiples mais ne connaît pas la MGPP [Gustafson, 2013]. Sa seule utilisation est en statistiques publiques (Insee, Statistique Canada).
- ▶ Une limite importante de la MGPP est la qualité des liens.

References

- ▶ P. Lavallée , (2002) *Le sondage indirect ou la méthode généralisée du partage des poids*. Editions de l'université de Bruxelles. Editions Ellipses
- ▶ P. Lavallée , (2007) *Indirect Sampling*. Springer Series in Statistics. Springer
- ▶ Ilan H. Meyer and Patrick A. Wilson (2009) *Sampling Lesbian, Gay, and Bisexual Populations*. American Psychological Association, 56, No. 1, 23-31.
- ▶ M. Jauffret-Roustide and Y. Le Strat (2010) *Conception, échantillonnage, organisation du terrain d'une enquête TLS et nouveaux développements: Coquelicot, une enquête auprès d'usagers de drogues*. Methodological innovations. Online
- ▶ M. Marpsat and N. Razafindratsima (2010) *Les méthodes d'enquêtes auprès des populations difficiles à joindre : introduction au numéro spécial*. Methodological innovations. Online
- ▶ Q. Xia and L. V. Torian (2013) *To weight or not to weight in time-locatin sampling: why not do both?*. AIDS Behav, Commentary, Online.
- ▶ P. Gustafson et al, (2013) *Impact of statistical adjustment for frequency of venue attendance in a venue-based survey of Len who have sex with men*. American journal of epidemiology, 177, No. 10, 1157-1164 .