

ECHANTILLONNAGE LIEUX-MOMENTS CHEZ LES HOMMES AYANT DES RAPPORTS SEXUELS AVEC LES HOMMES DANS LE CONTEXTE DU VIH : L'IMPORTANCE DE PRENDRE EN COMPTE LES POIDS DE SONDAGE ET LA FREQUENTATION DES LIEUX

Cécile SOMMEN¹ & Leïla SABONI² & Claire SAUVAGE³ & Antonio ALEXANDRE⁴ & Florence LOT⁵ & Francis BARIN⁶ & Annie VELTER⁷

¹ *Santé publique France, 12 rue du Val d'Osne, F-94415 Saint-Maurice, cecile.sommen@santepubliquefrance.fr*

² *Santé publique France, 12 rue du Val d'Osne, F-94415 Saint-Maurice, leila.saboni@santepubliquefrance.fr*

³ *Santé publique France, 12 rue du Val d'Osne, F-94415 Saint-Maurice, claire.sauvage@santepubliquefrance.fr*

⁴ *Equipe nationale d'intervention en prévention et santé pour les entreprises, Paris, France, a.alexandre@enipse.fr*

⁵ *Santé publique France, 12 rue du Val d'Osne, F-94415 Saint-Maurice, florence.lot@santepubliquefrance.fr*

⁶ *Université Tours, 10 Bd Tonnellé, F-37032 Tours, francis.barin@univ-tours.fr*

⁷ *Santé publique France, 12 rue du Val d'Osne, F-94415 Saint-Maurice, annie.velter@santepubliquefrance.fr*

Résumé. Les rapports sexuels entre hommes représentent le mode de transmission du VIH le plus fréquent dans les pays industrialisés. Le suivi des comportements sexuels à risque chez les hommes ayant des rapports sexuels avec des hommes (HSH) est crucial, en particulier pour comprendre la dynamique de l'épidémie à VIH. Une enquête transversale (Prevagay 2015), basée sur l'échantillonnage lieux-moments (TLS), a été réalisée en 2015 auprès des HSH fréquentant des lieux de convivialité gays dans 5 métropoles françaises. La méthode généralisée de partage de poids (MGPP) a été appliquée pour estimer la séroprévalence du VIH, pour la première fois dans cette population en France, en tenant compte de la fréquentation des établissements (FVA). Nos objectifs sont de décrire la mise en œuvre du plan d'échantillonnage, de démontrer l'importance de prendre en compte les poids d'échantillonnage et la FVA en comparant les résultats obtenus avec ou sans l'utilisation des poids d'échantillonnage et de la FVA. Une prévalence globale de 14,3% (IC 95% [12,0 - 16,9]) a été estimée en utilisant la MGPP et une prévalence brute de 16,4% (IC 95% [14,9 - 17,8]). La variance des estimations de la prévalence du VIH dans chaque ville était plus faible lorsque les poids d'échantillonnage et de la FVA n'étaient pas pris en compte. Une association de la FVA et des statuts sérologiques dans la plupart des villes étudiées a été également mise en évidence.

Mots-clés. HSH, prévalence du VIH, échantillonnage lieux-moments (TLS), fréquentation des lieux

Abstract. Sex between men is the most frequent mode of HIV transmission in industrialized countries. Monitoring risk behaviours among men who have sex with men (MSM) is crucial, especially to understand the drivers of the epidemic. A cross-sectional survey (PREVAGAY), based

on time-location sampling, was conducted in 2015 among MSM attending gay venues in 5 metropolitan cities in France. We applied the generalized weight share method (GWSM) to estimate HIV seroprevalence for the first time in this population, taking into account the frequency of venue attendance (FVA). Our objectives were to describe the implementation of the sampling design and to demonstrate the importance of taking into account sampling weights including FVA by comparing results obtained by GWSM and by other methods which use sample weights not including FVA or no weight. We found a global prevalence of 14.3 % (95%CI [12.0 – 16.9]) using GWSM and an unweighted prevalence of 16.4% (95%CI [14.9 – 17.8]). Variance in HIV prevalence estimates in each city was lower when we did not take into account either the sampling weights or the FVA. We also highlighted an association of FVA and serological status in the most of investigated cities.

Keywords. MSM ; HIV prevalence; time location sampling ; frequency of venue attendance

1. Introduction

En 2015, Santé publique France a mené une enquête (Prevagay 2015) auprès de HSH fréquentant des lieux de convivialité gays dans 5 villes métropolitaines françaises afin d'estimer la prévalence du VIH dans cette population spécifique. L'objectif principal de cette étude est de démontrer l'importance de l'utilisation du TLS associée à la MGPP. Pour ce faire, nous présentons la méthodologie de l'enquête Prevagay 2015 et décrivons la mise en œuvre du plan d'échantillonnage. Nous estimons la séroprévalence du VIH chez les HSH fréquentant des lieux homosexuels dans cinq métropoles françaises en utilisant différents poids d'échantillonnage : pas de poids, pas de FVA, MGPP avec prise en compte ou pas des valeurs extrêmes pour évaluer l'impact de la prise en compte des poids d'échantillonnage et de la FVA en inférence. La méthode utilisée pour l'inférence est décrite par Léon et al.[1].

2. Méthode

2.1 Design de l'étude

L'enquête Prevagay 2015 a été menée dans cinq villes françaises : Lille, Lyon, Montpellier, Nice et Paris.

Dans chaque ville, la taille attendue de l'échantillon a été déterminée à partir de la prévalence du VIH attendue et de la précision souhaitée de l'estimation de la prévalence du VIH. La prévalence escomptée du VIH était basée sur le statut sérologique auto-déclaré des HSH ayant répondu à l'Enquête Presse Gays et Lesbiennes 2011, réalisée via Internet sur la base du volontariat [2].

Nous avons décidé d'enquêter les lieux de convivialité gay, c'est-à-dire des bars, des discothèques, des saunas et des backrooms, 4 jours par semaine. L'enquête s'est déroulée sur 6 semaines à Paris et 4 semaines dans chacune des autres villes entre septembre et décembre 2015.

Nous avons défini une durée de visite de 4 heures dans les établissements parisiens et de 3 heures dans les établissements des autres villes. Nous avons construit la base d'échantillonnage des lieux-moments. Pour chaque lieu, nous avons choisi deux visites par jour avec un total de 8 visites par semaine.

Une fois les établissements éligibles identifiés, nous avons demandé au personnel de l'association partenaire en lien avec les propriétaires d'établissements de s'assurer de leur participation à l'étude, de collecter des informations sur les heures d'ouverture et le nombre estimé de participants éligibles pour chaque visite, afin de construire la base d'échantillonnage pour chaque site.

Un design TLS en deux étapes a été utilisé. Au cours de la première étape dans chaque ville, nous avons sélectionné des lieux-moments en utilisant un échantillonnage aléatoire simple sans remise,

avec au minimum une visite par établissement. Le nombre de visites échantillonnées pour un établissement était proportionnel au nombre moyen d'HSH présents dans cet établissement lors de toutes les visites au cours de la période d'enquête. Dans la deuxième étape, pour chaque lieu-moment, les HSH ont été sélectionnés en utilisant un échantillonnage aléatoire systématique.

Pendant les visites, le nombre de participants éligibles était estimé par les enquêteurs ainsi que le nombre de refus de participer à l'enquête.

2.2 Données

Les hommes étaient éligibles à l'enquête s'ils avaient au moins 18 ans, avaient eu des relations sexuelles avec des hommes au cours des 12 mois précédents, lisaient et parlaient français et si ils acceptaient d'une part d'effectuer un auto-prélèvement sanguin sur buvard et d'autre part de remplir un auto questionnaire sur tablette électronique. Le test VIH a été réalisé par le Laboratoire National de Référence pour le VIH (Tours, France) à partir de sang séché sur buvard (DBS) avec un immunodosage combiné pour la détection de l'antigène p24 et des anticorps anti-VIH (Genscreen ultra HIV Ag-Biorad). [2]. Les échantillons positifs pour le VIH ont été confirmés par une combinaison de tests d'infection récente, de sérotypage et de Western blot [2, 3].

Dans chaque ville, une liste de tous les lieux de convivialité gay participant à l'enquête, était proposée à chaque participant à la fin du questionnaire en leur posant la question suivante: «Au cours du dernier mois, combien de fois avez-vous fréquenté les lieux suivants?» À partir de cette question précise, nous avons estimé la FVA pendant la période d'enquête pour chaque individu en (1) additionnant le nombre déclaré de visites dans différents lieux le mois précédent, (2) en divisant ce nombre par 30 (nombre moyen de jours dans un mois) et (3) en multipliant par nombre de jours échantillonnés dans la ville du participant. Nous avons supposé que la FVA n'a pas varié au cours de la période d'enquête.

2.3 Poids de sondage

Pour l'inférence, un poids de sondage a été assigné à chaque participant.

Au premier degré, la probabilité d'inclusion π_k^v d'un lieu-moment k est égal au nombre de visites échantillonné d'un lieu spécifique divisé par le nombre total de visites dans la base de sondage de la ville correspondante. Le poids de sondage du lieu-moment k est égal à l'inverse de sa probabilité d'inclusion : $w_k^v = 1/\pi_k^v$.

Au second degré, la probabilité d'inclusion $\pi_{i|k}$ correspond à la probabilité que chaque participant i fréquente le lieu-moment k . toute chose égale par ailleurs, la probabilité $\pi_{i|k}$ est égale au nombre de HSH tiré au sort dans le lieu-moment k divisé par le nombre estimé d'hommes éligible dans k . Ainsi, le poids de sondage w_i pour le participant i peut être égal à :

$$w_i = w_k^v \times w_{i|k} \text{ avec } w_{i|k} = 1/\pi_{i|k}$$

Cependant, dans le TLS, la fréquentation individuelle des lieux n'est pas identique pour tous les participants et donc w_i est biaisé. Dans ce contexte, nous définissons le poids MGPP \tilde{w}_i , obtenu à partir de la MGPP prenant en compte la FVA en divisant le poids individuel w_i par le nombre de visites dans les lieux participant pendant l'enquête (noté $nFVA_i$ pour le participant i). Ainsi, le poids MGPP est égal à :

$$\tilde{w}_i = w_i/nFVA_i$$

Malgré une recherche formative approfondie, des changements par rapport à la conception initiale peuvent survenir. Par exemple, le nombre de visites initialement planifié dans un établissement peut-être diminué du fait du refus de son propriétaire de permettre aux enquêteurs de poursuivre l'enquête dans son établissement. De même, le nombre attendu de HSH fréquentant un établissement au cours d'une période donnée peut être surestimé par rapport à la fréquentation réelle au moment de l'enquête, entraînant un nombre de participants à l'enquête moins important que prévu initialement. Cela peut conduire à des poids d'échantillonnage extrêmes qui peuvent sur-représenter les individus dans l'estimation d'une statistique d'intérêt (comme ici la prévalence du VIH). Aussi, l'estimation d'une statistique aurait pu être biaisée et sa variance surestimée. En conséquence, il était nécessaire de tronquer les poids les plus importants. Nous avons donc décidé de remplacer (tronquer) les poids dépassant un seuil égal au poids médian plus 4 fois l'écart interquartile des poids dans chaque ville, en gardant la même somme de poids initiaux $\sum_{i=1}^n \tilde{w}_i$ où n est la taille d'échantillon. L'estimation de la taille de la population ne change donc pas. Soit $\tilde{w}_{trim\ i}$ le poids tronqué du participant i :

$$\tilde{w}_{trim\ i} = (\tilde{w}_i \times 1_{\{\tilde{w}_i \leq T\}} + T \times 1_{\{\tilde{w}_i > T\}}) * \frac{\sum_{i=1}^n \tilde{w}_i}{\sum_{i=1}^n (\tilde{w}_i \times 1_{\{\tilde{w}_i \leq T\}} + T \times 1_{\{\tilde{w}_i > T\}})}$$

où $1_{\{a < b\}} = 1$ si $a < b$ et 0 sinon et T est le poids médian plus 4 écarts interquartile des poids de la ville du participant i .

Ces poids de sondage MGPP tronqués sont les poids qui seront définitivement utilisés dans les analyses.

2.4 Prise en compte des poids dans les analyses et analyses

En général, l'objectif principal des enquêtes transversales est d'estimer des statistiques d'intérêt de la population, telles qu'un total (par exemple, le nombre de HSH fréquentant des lieux de convivialité gay), une proportion (par exemple la prévalence du VIH dans la population), ou une moyenne (par exemple l'âge moyen des hommes séropositifs pour le VIH). Nous avons utilisé l'estimateur de Horvitz-Thompson [26] et sa variance [27] qui est largement utilisé dans les enquêtes.

Nous avons déclaré les poids d'échantillonnage, une stratification par ville et une correction en population finie (FPC) à chaque degré. Au premier degré, la FPC est égale au nombre de lieux-moments échantillonnés divisé par le nombre total de lieux-moments. Au second degré, la FPC est égale au nombre de HSH interrogés divisé par le nombre total de HSH pendant les visites.

Nous avons estimé la prévalence biologique du VIH dans chaque ville. Nous avons comparé les estimations et leur intervalles de confiance à 95% obtenus à partir des différents poids de sondage : (1) pas de poids, (2) pas de FVA (w_i), (3) MGPP non tronquée (\tilde{w}_i) et (4) MGPP tronquée ($\tilde{w}_{trim\ i}$).

Nous avons également estimé l'effet plan de l'estimation de la prévalence VIH dans chaque ville. L'effet plan est égal à la variance estimée de la prévalence VIH estimée en tenant compte du plan de sondage divisé par la variance estimée de la prévalence VIH estimée à partir d'un sondage aléatoire simple.

3. Résultats

3.1 Profil des répondants

L'étude a recruté 2 646 participants dans les cinq villes (avec un taux de participation de 50%): 478 à Lille, 485 à Lyon, 266 à Montpellier, 328 à Nice et 1089 à Paris. Au total, 247 visites ont eu lieu: 45

à Lille, 42 à Lyon, 45 à Montpellier, 42 à Nice et 73 à Paris. En moyenne, 14 personnes à Paris et 8 dans d'autres villes ont été incluses à chaque visite.

L'âge médian pondéré des participants était de 41 ans. Parmi eux, 64% poursuivaient des études post-bac et 84% se définissaient comme homosexuels. La table 1 décrit le profil des répondants

	Lille (n=478)		Lyon (n=485)		Montpellier (n=266)		Nice (n=328)		Paris (n=1089)		Total (n=2646)	
	UW*	W**	UW*	W**	UW*	W**	UW*	W**	UW*	W**	UW*	W**
Age médian	36	38	35	32	37	42	40	41	43	44	40	41
Avoir suivi des études supérieures	58%	63%	61%	62%	63%	65%	55%	53%	70%	69%	64%	64%
Se définir homosexuel	84%	79%	86%	84%	88%	86%	86%	83%	88%	85%	87%	84%

Table 1 : Profil des répondants * non pondéré ** MGPP tronquée

3.2 Poids de sondage et distribution des FVA

Avant tronquage, les poids de sondage MGPP \tilde{w}_i variaient entre 0.025 et 200, on comptait 85 participants avec des poids supérieurs au poids médian plus 4 écarts interquartiles de leur ville. Le détail par ville est décrit dans la Table 2. La distribution des poids de sondage tronqués est donnée Figure 1. La FVA variait entre 1 to 215 avec une médiane de 6 FVA dans les 5 villes réunies: 4 à Lille, 6 à Lyon et Montpellier, 4 à Nice and 8 à Paris. La distribution des FVA est décrite Figure 1.

Ville	Min \tilde{w}_i	Max \tilde{w}_i	Seuil	nombre $\tilde{w}_i >$ Seuil
Lille	0.025	39.7	15.9	7
Lyon	0.103	90.7	21.5	26
Montpellier	0.029	36	9.17	5
Nice	0.084	200	18.8	14

Table 2: Poids de sondage minimum and maximum avant tronquage, valeur du seuil (médiane plus 4 écarts interquartile dans la ville), et nombre de poids de sondage supérieurs au seuil.

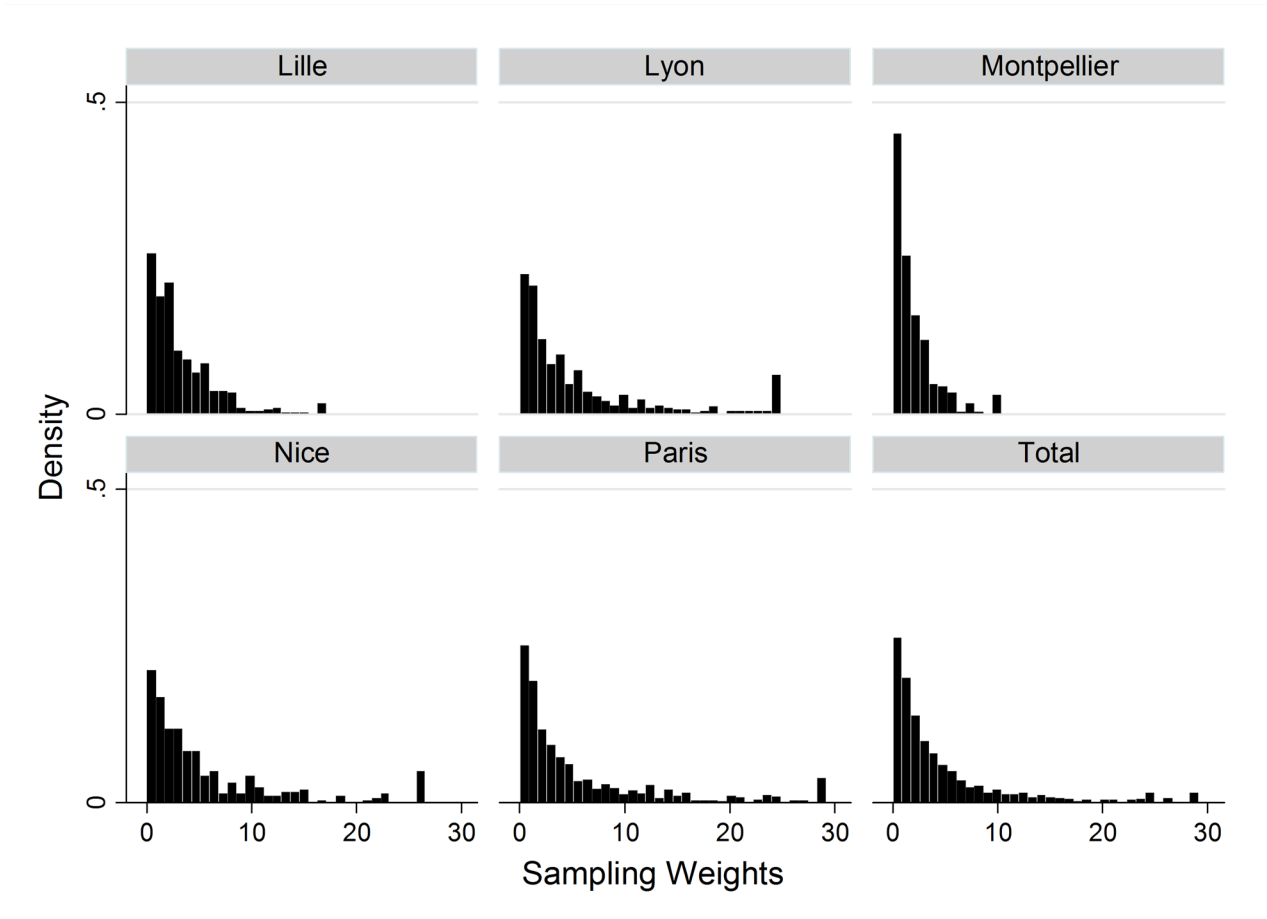


Figure 2 : Distribution des poids de sondage MGPP tronqués

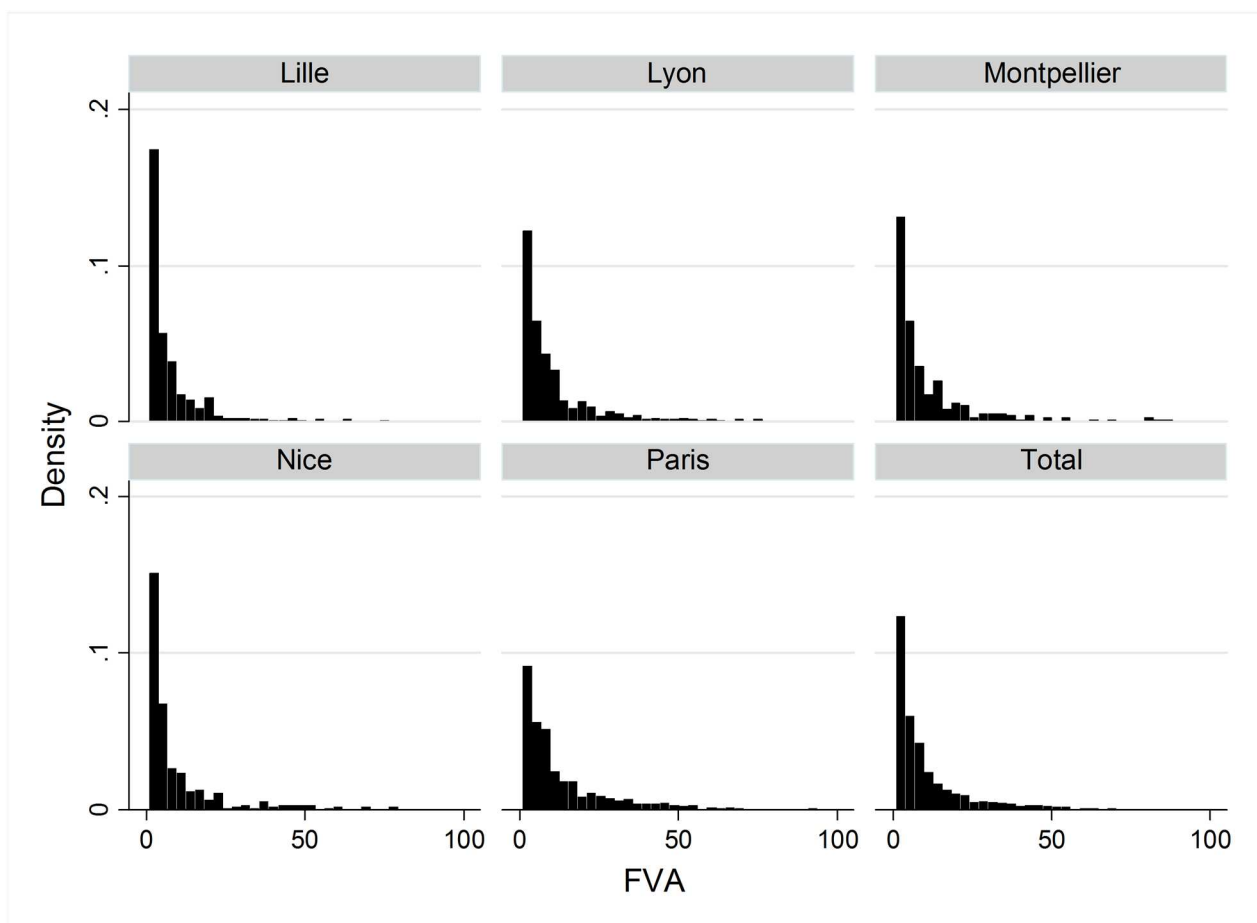


Figure 1: Distribution de la FVA (FVA supérieure à 100 -n=19- a été enlevée de la figure)

3.3 Estimation de la prévalence VIH

La prévalence du VIH parmi les HSH fréquentant des lieux de convivialité gays dans les 5 villes étudiées a été estimée à 14,3% (IC 95% [12,0 - 16,9]). Une régression logistique pondérée sur le statut VIH de tous les participants, ajustée sur la ville, l'âge et le niveau d'éducation, conclut à des différences significatives de prévalence du VIH entre les 5 villes ($p < 0,001$), notamment entre Paris et Lille. Cependant, aucune différence significative n'a été observée entre Paris et les autres villes. Plus précisément, nous avons estimé une prévalence de 7,6% (IC 95% [5,1 - 11,1]) à Lille, 11,4% (IC 95% [6,9 - 18,3]) à Lyon, 16,9% (IC 95% [11,2 - 24,7]) à Montpellier, 17,1% (IC 95% [11,8 - 24,1]) à Nice, et 16,1% (IC 95% [12,5 - 20,4]) à Paris.

Un modèle de régression logistique pondérée, utilisé pour expliquer le statut sérologique VIH dans chaque ville en fonction du nombre de FVA, ajusté pour l'âge et le niveau d'éducation, a montré que la FVA avait un effet significatif sur le statut VIH à Paris, Lille et Nice. Plus la FVA était élevée, plus le risque d'être séropositif était élevé.

Nous avons comparé les estimations de la prévalence du VIH en utilisant différents poids d'échantillonnage (Figure 3). Les intervalles de confiance non pondérés à 95% pour la prévalence du VIH (sans poids) se chevauchaient avec les intervalles de confiance des estimations utilisant les différents poids de sondage $\tilde{w}_{trim\ i}$, \tilde{w}_i et w_i . Les prévalences brutes étaient incluses dans les intervalles de confiance des estimations avec des poids de sondage dans toutes les villes sauf Nice. La variance de la prévalence brute était plus faible que les variances basées sur la MGPP.

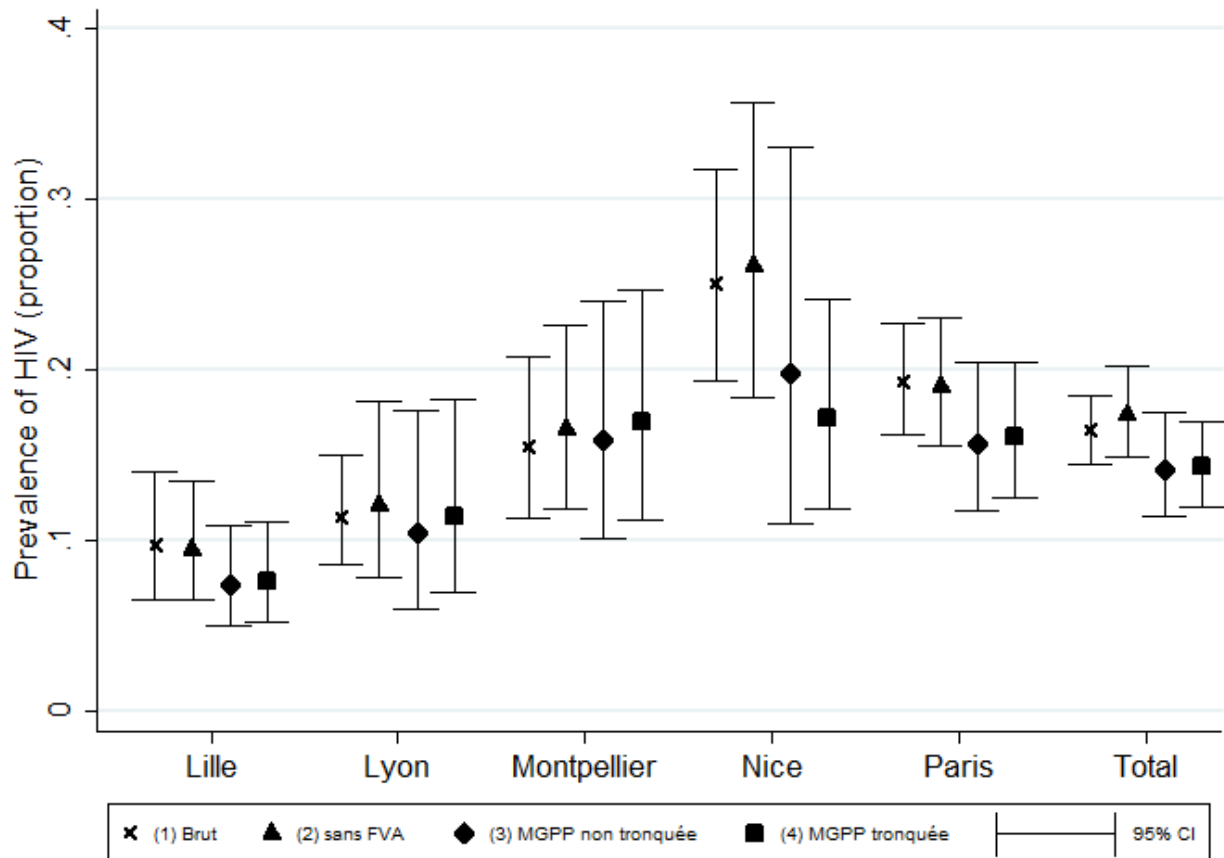


Figure 2: Estimation de la prévalence VIH dans les différentes villes selon les poids de sondage : $\tilde{w}_{trim\ i}$ (MGPP tronquée), \tilde{w}_i (MGPP non tronquée), w_i (sans FVA) and sans poids.

Nous avons calculé l'effet plan de la prévalence estimée du VIH dans chaque ville (Table 3). Les effets plans étaient différents selon les villes, avec un minimum à Lille et un maximum à Lyon. L'effet plan de la prévalence estimée du VIH variait de 1,2 (Lille) à 4,0 (Lyon).

	$\tilde{w}_{i\ trim}$
Lille	1.2
Lyon	4.0
Montpellier	2.4
Nice	1.7
Paris	2.8
Total	3.1

Table 3: Effets plan des estimations de prévalence pour chaque ville

4. Conclusion

Nous avons appliqué la MGPP pour fournir les estimations les plus précises de la prévalence du VIH chez les HSH fréquentant des lieux de convivialité gays. Cette méthode a pris en compte les poids TLS et la FVA des individus. De toutes les études publiées récemment sur les HSH fréquentant des lieux de convivialité gay, à notre connaissance, seuls Gustafson et al. [4] ont produit des estimations en utilisant les poids de sondage et la FVA. Cependant, ils ont utilisé une méthode d'estimation

différente de la MGPP. D'autres études en Australie et aux États-Unis ont utilisé le TLS, mais ont fourni des estimations sans tenir compte de la FVA [5] et parfois sans poids de sondage [6, 7]. La nécessité d'utiliser à la fois des poids d'échantillonnage et la FVA dans l'inférence a été démontrée [1, 4]. Dans notre étude, la variance des estimations de la prévalence du VIH était inférieure aux estimations basées sur la MGPP lorsque nous ne tenions pas compte du plan d'échantillonnage (estimations non pondérées). Les estimations non pondérées, même si elles sont encore couramment utilisées, peuvent conclure à tort qu'il existe des différences significatives dans la prévalence du VIH entre les villes, ajustées en fonction de l'âge et du niveau d'éducation.

Nous avons décidé de réduire les poids d'échantillonnage extrêmes afin d'éviter une variance des estimations artificiellement trop grande. Bien que le tronquage soit souvent utilisé, il n'y a pas de consensus dans la littérature sur la façon de réduire les poids extrêmes [8-11]. Notre choix était basé sur un compromis entre une variation aussi faible que possible des poids et la plus grande réduction de la variance de certaines statistiques clés (prévalence du VIH et prévalence de la charge virale négative).

Nous avons présenté l'effet plan pour aider les chercheurs à calculer la taille des échantillons dans la mise en place de plans de sondage similaires. Comme l'effet plan varie de 1,2 et 4 selon la ville, nous recommandons d'utiliser la valeur maximale. Il suffira alors de multiplier la taille de l'échantillon nécessaire à partir d'une hypothèse d'échantillonnage aléatoire simple par l'effet plan.

Notre étude a également montré l'impact de la non-utilisation de la FVA dans les analyses. En effet, dans les villes où la FVA était positivement associée à l'état sérologique, les estimations de prévalence obtenues sans tenir compte de la FVA étaient différentes de celles obtenues par la MGPP. Dans leur travail de simulation, Léon et al. ont montré que ne pas utiliser la FVA produisait une estimation biaisée [1].

En ce qui concerne la FVA, malgré le fait que la MGPP exige uniquement le nombre total de FVA, nous avons décidé de demander aux participants à quelle fréquence ils visitaient des établissements gay participants spécifiques. Nous aurions pu poser une seule question sur la fréquentation globale des établissements gays, mais il aurait été difficile de faire la différence entre les établissements participants et non-participants. De plus, comme chaque ville était étudiée indépendamment, une question globale aurait pu conduire à une surestimation des fréquentations des HSH qui voyageaient entre les villes. Enfin, nous avons pensé que demander des informations sur chaque établissement fournirait une réponse plus précise qu'une question générale sur tous les établissements.

Finalement, la mise en œuvre du TLS et l'utilisation de la MGPP ont permis, pour la première fois, de réaliser une enquête aléatoire auprès des HSH en France fréquentant des lieux de convivialité gay et de produire des résultats statistiques fiables, notamment concernant l'estimation de la prévalence du VIH dans cette population.

Bibliographie

1. Léon, L., M. Jauffret-Roustide, and Y. Le Strat, *Design-based inference in time-location sampling*. Biostatistics, 2015. **16**(3): p. 565-579.
2. Velter, A., et al., *HIV prevalence and sexual risk behaviors associated with awareness of HIV status among men who have sex with men in Paris, France*. AIDS and Behavior, 2013. **17**(4): p. 1266-1278.
3. Le Vu, S., et al., *Biomarker-based HIV incidence in a community sample of men who have sex with men in Paris, France*. PloS one, 2012. **7**(6): p. e39872.
4. Gustafson, P., et al., *Impact of statistical adjustment for frequency of venue attendance in*

- a venue-based survey of men who have sex with men.* American Journal of Epidemiology, 2013: p. 1157-1164.
5. Mirandola, M., et al., *The Sialon II Project. Report on a Bio-behavioural Survey among MSM in 13 European cities.* 2016. p. 29-42.
 6. Holt, M., et al., *The prevalence and correlates of undiagnosed HIV among Australian gay and bisexual men: results of a national, community-based, bio-behavioural survey.* Journal of the International AIDS Society, 2015. **18**: p. 20526-20533.
 7. Snowden, J.M., et al., *Prevalence and characteristics of users of pre-exposure prophylaxis (PrEP) among men who have sex with men, San Francisco, 2014 in a cross-sectional survey: implications for disparities,* in *Sexually Transmitted Infections.* 2016.
 8. Alexander, C., S. Dahl, and L. Weidman. *Making estimates from the American Community Survey.* in *Annual Meeting of the American Statistical Association (ASA), Anaheim, CA.* 1997.
 9. Elliott, M., *Model averaging methods for weight trimming.* Journal of official statistics, 2008. **24**(4): p. 517-521.
 10. Kish, L., *Weighting for Unequal Pi.* Journal of Official Statistics, 1992. **8**: p. 183-200.
 11. Potter, F., *A Study of Procedures to Identify and Trim Extreme Survey Weights.* American Statistical Association, 1990: p. 225-230.