



N° 11-522-XIF au catalogue

**La série des symposiums internationaux  
de Statistique Canada - Recueil**

**Symposium 2004 : Méthodes  
innovatrices pour enquêter  
auprès des populations  
difficiles à joindre**

2004



Statistique  
Canada

Statistics  
Canada

Canada

## ESTIMATION DE LA TAILLE DE LA POPULATION D'UDI UTILISANT LES PROGRAMMES D'ÉCHANGE DE SERINGUES

Shenghai Zhang, Ping Yan et Chris Archibald<sup>1</sup>

### RÉSUMÉ

Dans cet article, nous décrivons un plan d'échantillonnage et d'estimation de la taille de la population d'utilisateurs de drogues injectables (UDI) qui recourent aux programmes d'échange de seringues (PES). On y exploite l'information sur le nombre de seringues distribuées dans les centres PES. Le plan de sondage prévoit la création de l'échantillon d'UDI qui se présentent à des PES dans une certaine période, ainsi que l'obtention de données rétrospectives d'autodéclaration sur le nombre d'amis parmi les UDI et sur le nombre de seringues échangées pour chaque UDI échantillonné. Nous dressons un cadre méthodologique qui permet d'estimer la taille de la population d'UDI qui n'ont jamais utilisés les PES dans la période d'observation fixée, de même que la proportion d'UDI dans le groupe des utilisateurs des PES. On estime la taille de la population en prenant le quotient du nombre total d'UDI dans le groupe d'utilisateurs des PES et de la proportion estimée d'UDI dans ce même groupe. La technique pourrait fournir des données pouvant répondre à des questions comme les suivantes : combien de Canadiens sont-ils des UDI? Leur nombre évolue-t-il? Nous pourrions aussi mieux comprendre la dynamique de la population nationale d'UDI.

MOTS CLÉS : Estimation, population d'UDI, programmes d'échange de seringues.

### 1. INTRODUCTION

La consommation de drogues injectables joue un grand rôle dans l'épidémie mondiale de VIH-SIDA. En 2003, cette infection virale par transmission entre UDI a été rapportée par plus de 130 pays. Dans certains pays, la consommation de drogues intraveineuses est aujourd'hui le principal mode de transmission du VIH. On estime en gros que, dans le monde en 2003, il y avait plus de 13 millions d'utilisateurs de drogues injectables et que, dans certaines régions, plus de la moitié d'entre eux avaient le VIH (ONUSIDA, 2004). Les conséquences néfastes sur la santé de la consommation de drogues injectables ne se limitent pas à cette infection virale. L'échange de seringues présente un risque élevé de transmission d'autres infections à diffusion hématogène comme les hépatites B et C. Ajoutons que, par cette consommation de drogues intraveineuses, l'épidémie se propage bien au-delà des rangs des utilisateurs de drogues injectables. Ces consommateurs, leur conjoint et leurs enfants figurent pour au moins 36 % de l'ensemble des cas de SIDA déclarés aux États-Unis pour toute l'année 1999 (CDC, 2001). Au Canada, l'utilisation de drogues injectables est aussi un problème. Il est difficile de recueillir des données précises sur la prévalence et le profil de la consommation de ces drogues sur le territoire canadien, mais il est clair que les UDI sont en grand nombre à travers le pays (Archibald et coll., 2001; Weekes et Cumberland, 2004).

Les décideurs et les chercheurs ont constaté la nécessité de recueillir systématiquement des données sur l'importance et les tendances des besoins de consommation de drogues injectables pour acquérir une compréhension complète de l'épidémie de VIH chez les UDI, ainsi que pour les efforts de prévention consentis dans ce domaine (Dehne, Adelekan, Chatterjee et Weiler, 2002). En ayant une certaine compréhension de la dynamique des UDI, il devient possible non seulement d'évaluer l'incidence probable de la propagation du VIH-SIDA et autres maladies apparentées, mais aussi de permettre de faire connaître aux décideurs d'une situation qui s'aggrave ou autrement de démontrer que les mesures de prévention et autres initiatives donnent des résultats. Il reste que, pour la plupart des villes, des pays et des régions du globe, nous ne disposons d'aucune estimation sûre du nombre d'utilisateurs de drogues injectables (Dehne, Adelekan, Chatterjee et Weiler, 2002) et que, souvent, les chiffres qui décrivent cette importance et ces tendances sont des « estimations au jugé » produites par les responsables des services de santé

---

<sup>1</sup>Centre de prévention et de contrôle des maladies infectieuses, Santé Canada, Ottawa (courrier électronique : szhang@hc-sc.gc.ca).

plutôt que des chiffres issus d'activités de recherche ou de surveillance systématique. Précisons toutefois que les décideurs et les chercheurs se sont heurtés au problème de collecte de données précises sur le nombre et le comportement de ces consommateurs de drogue, données qui peuvent difficilement s'obtenir dans une enquête type auprès d'une population générale (une enquête type consisterait à étudier les individus provenant d'un échantillon aléatoire de ménages dans une région, une province ou un pays selon l'étendu de l'étude). On peut voir dans les drogues injectables un phénomène socialement inacceptable et, par conséquent, les utilisateurs de ces drogues constituent une « population difficile à joindre ». Ces gens se trouvent parfois même à dissimuler leurs habitudes de vie à ceux qui les entourent : parents, personnes partageant leur logement, partenaires sexuels, etc. Récemment, les chercheurs en santé publique, les statisticiens et les sociologues se sont attaqués à la question de cette limite actuelle de la méthodologie.

## 2. ESTIMATION DE LA TAILLE DE LA POPULATION D'UDI QUI FRÉQUENTENT LES PES

Pour agir optimalement en prévention de la maladie, on a besoin d'estimations régulières et précises de la prévalence à la maladie. Les programmes d'échange de seringues, où on met des seringues stérilisées, gratuitement ou à un coût infime, à la disposition des consommateurs de drogues injectables, offrent un moyen commode d'observer la prévalence d'infections virales à transmission hématogène chez un grand nombre d'utilisateurs de telles drogues qui s'en injectent actuellement. Au Canada, on compte plus de 200 PES, et d'autres verront le jour (Santé Canada, 2004). On dénombre aussi beaucoup de pharmacies qui fournissent des services semblables (Santé Canada, 2004). En 2002 par exemple, 1 109 056 seringues ont été échangées au total dans les PES de Regina au Canada (rapport (sur les toxicomanies) Regina 2002, 2004).

Dans cette section, nous estimerons la taille de la population d'UDI qui fréquentent les PES en nous reportant aux données sur le nombre de seringues distribuées dans une période déterminée. Pour obtenir cette information auprès des centres PES, une simple sélection des centres PES à être interviewés pourrait ne pas convenir. Pour estimer la taille de la population d'UDI qui utilisent les PES, nous proposons un échantillonnage à deux degrés où, d'abord, nous sélectionnons un échantillon de ces centres et ensuite, nous observons durant une semaine d'enquête tous les individus dans les centres sélectionnés, ce qui assure un échantillonnage d'individus dans un nombre limité de ces centres.

Comment choisit-on les centres d'échange de seringues? Le temps et le budget disponibles étant restreints et les caractéristiques de la population d'UDI étant ce qu'elles sont, il faut élaborer une stratégie d'échantillonnage des centres. La sélection pourrait être entièrement aléatoire à partir d'une liste complète de ces centres. Toutefois, si on entend diminuer la variabilité des estimateurs des paramètres basés sur les données de l'enquête, la théorie de l'échantillonnage nous dira qu'il est préférable d'échantillonner les grands centres (couvrant une population nombreuse d'UDI) avec des probabilités supérieures à celles de la sélection de petits centres (couvrant une population peu nombreuse d'UDI). Ainsi, le plan d'échantillonnage des centres PES que nous envisageons est tout à fait général et fréquemment appliqué dans cette situation, soit l'échantillonnage stratifié avec probabilité proportionnelle à la taille. En prenant le nombre de seringues échangées dans chaque centre PES la dernière année, nous pouvons répartir tous les centres à travers le Canada en  $L$  strates avec  $K_h$  centres faisant partie de la strate  $h^e$ . Précisons qu'on peut également acheter des seringues dans les pharmacies (Miller, Tyndall et coll., 2002), bien que les pharmaciens ne soient pas toujours disposés à en vendre aux UDI. Des lieux d'échange de ce genre peuvent constituer une de nos strates. Dans la strate  $h^e$ , nous sélectionnons  $k_h$  centres PES avec les probabilités d'inclusion  $\pi_{h1}, \pi_{h2}, \dots, \pi_{h,k_h}$ , qui sont déterminées par le nombre de seringues échangées dans la dernière année. Les  $\pi_{hi}$  peuvent être déterminés de la façon suivante. Supposons que l'échantillonnage soit sans remise, les probabilités d'inclusion du  $hi^e$  centre PES dans la liste de  $k_h$  centres sont  $\pi_{hi} = k_h Q_{hi} / \sum_{j=1}^{K_h} Q_{hj}$ , où  $Q_{hj}$  est le nombre total de seringues échangées au  $i^e$  centre de la  $h^e$  strate dans la dernière année. Il faut que  $\pi_{hi} < 1$ . Tel est

le cas pour tous les  $i$  quand  $k_h = 1$ . On peut s'attendre à ce qu'il n'y ait pas de valeur  $T_{hj}$  énorme à l'intérieur de chaque strate, ainsi cette condition peut être satisfaite. Autrement, nous pourrions fixer  $\pi_{hi} = 1$  pour les  $i$ , de sorte que  $k_h Q_{hi} / \sum_{j=1}^{K_h} Q_{hj} > 1$  (Sarndal, Swensson et Wretman, 1992).

Il est sûr que la taille de la population d'UDI qui est couverte par un centre PES est corrélée avec le nombre de seringues distribuées. Ainsi, un échantillonnage avec probabilité proportionnelle à la taille permet de diminuer la variabilité des estimateurs de totaux (Cochran, 1977). Il y a aussi des avantages opérationnels à un échantillonnage à plusieurs degrés tel que la possibilité d'égaliser la charge de travail entre les régions échantillonnées au premier degré.

La sélection de centres d'échange de seringues se fait en fonction du nombre élevé de clients et d'une représentation de tous les secteurs de juridiction canadienne. À l'heure actuelle, il y a une enquête menée par Santé Canada où on veut décrire la prévalence de l'infection virale VIH et les comportements de risque dans ce domaine chez les consommateurs de drogues injectables au pays. On pourrait ajouter au questionnaire une question sur le nombre de seringues utilisées dans les trois derniers mois. À tous les UDI fréquentant les centres sélectionnés d'échange de seringues dans une semaine déterminée, on demanderait de remplir un bref questionnaire et de fournir un prélèvement de sang qui ferait l'objet d'une analyse de détection du VIH et de l'hépatite C.

Soit  $T_{hi}$  le nombre de seringues échangées dans les trois mois d'observation dans le  $i^e$  centre PES échantillonné de la  $h^e$  strate. Posons que  $N_{hi}$  UDI ont échangé des seringues dans ce centre dans les trois derniers mois et que  $N_{hi_j}$  est le nombre d'UDI qui ont utilisé  $j$  seringues dans cette période et que toutes les seringues proviennent de ce centre. Les valeurs  $N_{hi}$  et  $N_{hi_j}$  visent non seulement les UDI qui ont fréquenté le centre en question, mais aussi les UDI qui en utilisent les seringues dans un échange secondaire. Une enquête (DesJarlais, McKnight, Eigo et Friedmann, 2002) chez notre voisin du sud indique que 90 % des PES encouragent activement les échanges secondaires qui, par définition, sont un échange de seringues qui seront utilisées par des personnes autres que la personne qui fait l'échange. Ainsi, le nombre de clients d'un centre PES ne correspond pas exactement au nombre d'UDI qui échangent les seringues dans ce centre. Avec la méthode que nous venons de décrire, il est possible d'estimer avec plus de précision la taille de la population d'UDI qui fréquentent les PES.

Nous pouvons maintenant établir la relation entre le nombre d'UDI qui utilisent les seringues du centre au nombre total de seringues qui y sont échangées dans les trois mois en question.

$$T_{hi} = \sum_{j=1}^m jN_{hi_j} = N_{hi} \sum_{j=1}^m jR_{hi_j} ,$$

où  $m$  est le maximum de seringues qu'un UDI pourrait utiliser durant les trois mois et où  $R_{hi_j} = N_{hi_j} / N_{hi}$  peut être estimé par l'échantillon obtenu lors de la semaine d'enquête. À noter que  $N_{hi}$  et  $N_{hi_j}$  sont des inconnues. Ce que nous pouvons cependant déterminer, c'est la proportion  $R_{hi_j}$  qui peut être estimée par la proportion  $r_{hi_j} = n_{hi_j} / n_{hi}$  de l'échantillon, où  $n_{hi}$  est le nombre d'utilisateurs de drogues injectables qui sont présents lors des échanges de seringues dans un centre d'échange et qui remplissent un questionnaire durant la semaine d'enquête. Parmi eux, il y a  $n_{hi_j}$  UDI qui ont utilisé  $j$  seringues dans les trois derniers mois. Ainsi,  $N_{hi}$  peut s'estimer par

$$\hat{N}_{hi} = \frac{T_{hi}}{\sum_{j=1}^m jr_{hi_j}} .$$

En fait, le dénominateur  $r_{hi} = \sum_{j=1}^m jr_{hi_j}$  est le nombre moyen de seringues échangées par les UDI qui ont fréquenté le

PES pour un échange de seringues dans les trois mois. L'estimateur de la taille  $N$  de la population d'UDI qui ont fréquenté des PES au moins une fois dans les trois derniers mois pour toutes les juridictions canadiennes est défini par la moyenne pondérée

$$\hat{N} = \sum_{h=1}^L \sum_{i=1}^{k_h} \pi_{hi}^{-1} \hat{N}_{hi} .$$

Supposons que les UDI utilisant les PES ont les mêmes profils que ceux de la semaine d'enquête. En d'autres termes, posons que l'échantillon d'UDI observés dans la semaine d'enquête s'obtient par échantillonnage aléatoire simple sans remise pour les UDI utilisant les PES durant la période de trois mois. Selon les enquêtes où on a étudié les comportements de risque des UDI en Australie (MacDonald, Wodak et coll., 1997) et au Canada (Peggy, Myers, Calzavara et coll., 2003), les UDI observés dans la semaine d'enquête sont représentatifs des UDI utilisant les PES pendant une période de trois mois. Ainsi, l'hypothèse formulée au sujet de l'échantillon paraît acceptable. On pourrait appliquer les algorithmes d'un échantillonnage stratifié avec probabilité proportionnelle à la taille et sans remise en faisant appel à de nombreuses méthodes, dont les algorithmes de Hanurav-Vijayan (Vijayan, 1968) et de Brewer (Brewer, 1963).

### 3. ESTIMATION DE LA PROPORTION D'UDI DANS LE GROUPE D'UTILISATEURS DES PES

Selon l'hypothèse de la dernière section, la population d'UDI est formée de deux groupes d'individus selon leur statut de participation aux échanges de seringues des PES. Pour estimer la taille de cette population, il faut connaître non seulement la taille de la population qui utilise les PES, mais aussi la proportion de la population qui participe aux PES durant les trois mois. Pour estimer cette proportion, nous prenons l'idée d'un échantillonnage déterminé selon les répondants « respondent-driven sampling » (Salganik et Heckathorn, 2004).

Il est possible de répartir les UDI échantillonnés du  $hi^e$  centre dans la semaine d'enquête en deux groupes. Le premier groupe aura fréquenté les PES au moins une fois dans les trois mois et l'autre ne les aura jamais utilisés dans la même période. Ces groupes sont respectivement désignés par  $\Omega_{NEP}^{hi}$  et  $\Gamma_{NEP}^{hi}$ . Normalement, le groupe  $\Gamma_{NEP}^{hi}$  sera de petite taille. Pour accroître la taille et la couverture, nous pouvons utiliser chaque UDI du groupe  $\Gamma_{NEP}^{hi}$  comme point de départ pour un échantillonnage déterminé par les répondants où chaque membre initial du groupe se trouvera à recruter un certain nombre d'UDI et à nous renseigner sur le nombre d'amis appartenant respectivement au groupe d'utilisateurs et au groupe de non-utilisateurs des PES. Le nouvel échantillon obtenu par les membres du groupe  $\Gamma_{NEP}^{hi}$  comprendra des utilisateurs des PES et des non-utilisateurs. Le groupe  $\Omega_{NEP}^{hi}$  sera celui qui comprend tous les non-utilisateurs des PES échantillonnés. Pour estimer une proportion de population, il nous faut connaître la structure du réseau, c'est-à-dire le nombre moyen de liens d'amitié, ainsi que la probabilité qu'un non-utilisateur des PES se lie d'amitié avec un utilisateur et vice versa.

Le nombre total de liens d'amitié pour les UDI du groupe  $\Omega_{NEP}^{hi}$  est désigné par  $\Phi_{NEP}^{hi}$  et peut s'écrire comme :

$$\Phi_{NEP}^{hi} = n_{NEP}^{hi} \Psi_{NEP}^{hi} ,$$

où  $\Psi_{NEP}^{hi}$  est le nombre moyen de liens d'amitié des UDI du groupe et où  $n_{NEP}^{hi}$  est le nombre de ces UDI. Soit  $\Pi_{NEP,NEP}^{hi}$  le nombre de liens d'amitié que l'ensemble des UDI du groupe  $\Omega_{NEP}^{hi}$  ont avec des UDI non utilisateurs des PES. On peut alors estimer la probabilité  $P_{NEP,NEP}$  qu'un UDI utilisant les PES ait un lien d'amitié avec un UDI non utilisateur dans la période de trois mois par

$$\hat{P}_{NEP,NEP} = \frac{\sum_{h=1}^L \sum_{i=1}^{k_h} \Pi_{NEP,NEP}^{hi}}{\sum_{h=1}^L \sum_{i=1}^{k_h} \Phi_{NEP}^{hi}}.$$

On peut alors estimer le nombre moyen de liens d'amitié des UDI du groupe  $\Omega_{NEP}$  en regroupant les échantillons des divers centres échantillonnés :

$$\hat{\Psi}_{NEP} = \sum_{d=1} d \frac{\sum_{h=1}^L \sum_{i=1}^{k_h} f_{NEP}^{hi}(d)}{\sum_{h=1}^L \sum_{i=1}^{k_h} n_{NEP}^{hi}},$$

où  $f_{NEP}^{hi}(d)$  est le nombre d'UDI du groupe  $\Omega_{NEP}^{hi}$  qui ont un nombre  $d$  de liens d'amitié entre UDI. De même, nous pouvons obtenir  $\hat{P}_{NEP,NEP}$  et  $\hat{\Psi}_{NEP}$ .

Basé sur le statut d'un UDI utilisant les PES dans la période de trois mois, nous dirons qu'un UDI est dans l'état  $S_{NEP}$  s'il n'a jamais utilisé les PES durant cette période; il sera dans l'état  $S_{\overline{NEP}}$  dans les autres cas. Supposons que nous prenons un plan d'échantillonnage déterminé selon les répondants pour construire un échantillon. Nous dirons que l'UDI qui appartient à l'échantillon de départ est choisi à l'étape 0. On pourrait choisir un autre UDI en considérant son lien d'amitié avec l'UDI de départ. Nous dirons de ce nouvel UDI qu'il est choisi à l'étape 1, et ainsi de suite. Posons que les probabilités de recrutement d'un autre UDI dans l'état  $S_{NEP}$  dépendent seulement par l'importance des liens d'amitié de cet UDI choisi avec les autres UDI. Posons aussi que, si l'UDI choisi à l'étape 0 est dans l'état  $S_{NEP}$ , alors un UDI dans l'état  $S_{\overline{NEP}}$  sera choisi avec la probabilité  $1 - P_{NEP,NEP}$ . Posons aussi que, si l'UDI choisi à l'étape 0 n'est pas dans l'état  $S_{NEP}$ , alors un UDI dans l'état  $S_{NEP}$  sera choisi à l'étape 1 avec la probabilité  $P_{\overline{NEP},NEP}$ . Si  $X_n$  désigne l'état d'un UDI choisi à la  $n^e$  étape, alors  $\{X_n, n = 0, 1, \dots\}$  est une chaîne de Markov à deux états ayant la matrice suivante de probabilités de passage :

$$\begin{pmatrix} P_{NEP,NEP} & P_{NEP,\overline{NEP}} \\ P_{\overline{NEP},NEP} & P_{\overline{NEP},\overline{NEP}} \end{pmatrix}.$$

Il est évident que c'est là une chaîne de Markov ergodique et irréductible (Ross, 1997). En fait, nous supposons que les UDI du groupe d'utilisateurs des PES présentent une même proportion de liens d'amitié que le groupe de non-utilisateurs de PES et vice versa. À noter aussi que  $P_{NEP,\overline{NEP}} = 1 - P_{NEP,NEP}$  et  $P_{\overline{NEP},NEP} = 1 - P_{\overline{NEP},\overline{NEP}}$ . Dans la pratique, il est bien plus simple d'échantillonner en un seul cycle. Il nous faut une méthode différente pour une approximation des proportions  $\pi_{NEP}$  et  $\pi_{\overline{NEP}}$  d'une chaîne d'UDI dans les états respectifs  $S_{NEP}$  et  $S_{\overline{NEP}}$ . D'après la théorie de Markov, nous avons

$$\pi_{NEP} = \frac{P_{\overline{NEP},NEP}}{1 + P_{\overline{NEP},NEP} - P_{NEP,NEP}} \quad \text{et} \quad \pi_{\overline{NEP}} = \frac{1 - P_{NEP,NEP}}{1 + P_{\overline{NEP},NEP} - P_{NEP,NEP}}.$$

Il est possible d'estimer  $\pi_{NEP}$  et  $\pi_{\overline{NEP}}$  par les estimateurs  $\hat{P}_{NEP,NEP}$  et  $\hat{P}_{\overline{NEP},NEP}$ . D'après le résultat présenté par Salganik et Heckathorn (Salganik et Heckathorn, 2004), nous pouvons estimer de la manière suivante la proportion  $\alpha$  d'UDI qui ont jamais utilisé les PES dans les trois mois :

$$\hat{\alpha} = \frac{\hat{\pi}_{NEP} \hat{\Psi}_{\overline{NEP}}}{\hat{\pi}_{\overline{NEP}} \hat{\Psi}_{NEP} + \hat{\pi}_{NEP} \hat{\Psi}_{\overline{NEP}}}.$$

Ayant estimé la taille  $\hat{N}$  de la population d'UDI utilisant les PES et la proportion  $\hat{\alpha}$  d'UDI utilisant les PES parmi tous les UDI, nous pouvons alors estimer la taille de la population d'UDI sous la forme  $\hat{N} / \hat{\alpha}$ .

#### 4. CONCLUSIONS

Longtemps, les chercheurs ont essayé d'avoir une mesure précise de la taille de la population d'UDI. Nous avons montré qu'il est possible de le faire par un plan d'échantillonnage utilisant une combinaison de programmes d'intervention corrective, à savoir les programmes d'échange de seringues (PES). En nous reportant à ces programmes et en exploitant les données sur le nombre de seringues échangées dans les centres PES, nous en venons à une nouvelle approche pour estimer la taille de la population d'UDI. Avec les PES, nous pouvons établir un plan d'échantillonnage et d'estimation qui est à moindre coût et d'une plus grande précision par rapport aux méthodes existantes. La méthodologie que nous décrivons ici peut être mise en application plus rapidement et plus facilement que les autres méthodes d'estimation.

Les PES offrent un modèle exhaustif de prévention du VIH et des autres infections à transmission hématogène dans la lutte livrée à la propagation de ces maladies chez les UDI. Il a été prouvé que cette intervention contre les comportements à risque était efficace dans le cas des UDI. Il est possible que le réseau des PES contienne une distribution âge-sexe représentative de la population de la région à laquelle nous nous intéressons. Il peut cependant être difficile de choisir tous les PES pour la collecte de données. Afin d'estimer la taille de la population d'UDI utilisant les PES, le plan de sondage que nous proposons s'en tient au principe fondamental de la théorie de l'échantillonnage selon lequel chaque individu de la population cible devrait avoir des chances non nulles d'être sélectionné dans l'enquête (Korn et Graubard, 1999). À noter que la population cible au premier degré de notre estimation est formée de tous les UDI qui ont utilisé les PES dans la période de trois mois. Ainsi, chaque UDI utilisant les PES a une certaine chance non nulle d'être échantillonné dans une enquête où on veut estimer la taille de la population d'UDI participant aux PES.

Nous prenons l'idée d'un échantillonnage déterminé selon les répondants « respondent-driven sampling » pour estimer la proportion de la population d'UDI qui utilisent les PES. Pour estimer cette proportion, il nous faut connaître les probabilités de passage, c'est-à-dire les probabilités qu'un non-utilisateur des PES ait un lien d'amitié avec un utilisateur des PES et vice versa. Les résultats provenant d'une chaîne de Markov et l'article de Salganik et Heckathorn (2004) sont utilisés pour produire les estimations.

Nous avons produit des estimateurs qui correspondent à un plan d'échantillonnage. On sait bien que le fait de ne pas tenir compte du plan d'échantillonnage peut mener à des estimations trompeuses, notamment, lorsqu'un échantillon est sélectionné à partir d'un plan d'échantillonnage non aléatoire et que l'on ignore ce plan pour traiter l'échantillon comme si c'était un échantillon aléatoire simple.

Nous avons avant tout voulu estimer la taille de la population d'UDI. Il est possible d'intégrer notre stratégie d'échantillonnage et d'estimation à l'étude sur les comportements à risque des UDI. Il faut alors étudier plus en profondeur la variance des estimateurs. Nous avons présenté un certain nombre de résultats analytiques, cependant, ces résultats pourraient être davantage supportés en procédant à une simulation numérique qui n'a pas encore été faite. Ajoutons qu'un biais est possible lorsque nous estimons les probabilités de passage, car la taille de l'échantillon des non-utilisateurs des PES est normalement petite. Il est toutefois possible de prédire l'ordre de grandeur et la direction de ce biais.

Nous espérons que les techniques que nous avons décrites permettront d'estimer la taille de la population d'UDI au Canada.

## ANNEXE

Dans cette annexe, il sera question de la variance de l'estimateur  $\hat{N}$ . Nous regardons d'abord l'estimateur  $\hat{N}_{hi}$ . Soit  $Y_{hi}$  le nombre de seringues échangées par un consommateur de drogues injectables dans un centre au cours des trois mois d'observation. La loi de probabilité de  $Y_{hi} = j$  est désignée par  $f_{hi}(j)$ . La moyenne de  $Y_{hi}$  est  $\bar{Y}_{hi} = \sum_j j f_{hi}(j)$ . La variance de  $Y_{hi}$  conditionnelle à  $N_{hi}$  est

$\text{var}(Y_{hi}) = \frac{N_{hi}}{N_{hi}-1} \left[ \sum_{j=1}^{N_{hi}} j^2 f_{hi}(j) - \left( \sum_{j=1}^{N_{hi}} j f_{hi}(j) \right)^2 \right]$ . Le  $N_{hi}$  dans  $f_{hi}(j)$  pourrait être respectivement remplacé par  $\hat{N}_{hi}$  et  $r_{hi_j}$ ; le ratio  $\frac{N_{hi}}{N_{hi}-1}$  peut être remplacé par 1, car  $N_{hi}$  est normalement très grand dans notre

cas. Selon ces argument et le fait que  $T_{hi}$  est connu et que  $\bar{Y}_{hi}$  est estimé par  $r_{hi}$ , nous recourons à des méthodes de linéarisation (Thompson, 1997, et Lohr, 1999) pour approximer la variance de l'estimateur  $\hat{N}_{hi}$ . Nous obtenons  $E\left[\left(\hat{N}_{hi} - N_{hi}\right)^2\right] \approx \frac{1}{\hat{Y}_{hi}^2} N_{hi}^2 \text{var}(\bar{y}_{hi})$ , où  $E$  désigne l'espérance mathématique. Enfin, nous pouvons estimer la

variance  $\hat{N}_{hi}$  par  $\hat{\text{var}}(\hat{N}_{hi}) = \frac{1}{\bar{y}_{hi}^2} \hat{N}_{hi}^2 \left(1 - \frac{n_{hi}}{\hat{N}_{hi}}\right) \frac{(\hat{\text{var}}(Y_{hi}))^2}{n_{hi}}$ , où  $\hat{\text{var}}(Y_{hi}) = \frac{\hat{N}_{hi}}{\hat{N}_{hi}-1} \left[ \sum_{j=1}^{\hat{N}_{hi}} j^2 r_{hi_j} - \left( \sum_{j=1}^{\hat{N}_{hi}} j r_{hi_j} \right)^2 \right]$ .

Comme l'échantillonnage est indépendant dans chaque strate et que, dans chacune, il s'agit d'un échantillonnage avec probabilité proportionnelle à la taille, avec taille fixe et sans remise, la variance de l'estimateur  $\hat{N}$  de la taille de la population d'UDI utilisant les PES durant les trois mois peut s'écrire de la façon suivante (Sarndal, Swensson et Wretman, 1992) :

$$\text{var}(\hat{N} \setminus \hat{N}_{hi}, i = 1, \dots, K_h) = \frac{1}{2} \sum_{h=1}^L \sum_{i=1}^{K_h} \sum_{l=1}^{K_h} (\pi_{hi} \pi_{hl} - \pi_{hil}) \left( \frac{\hat{N}_{hi}}{\pi_{hi}} - \frac{\hat{N}_{hl}}{\pi_{hl}} \right)^2,$$

où  $\pi_{hil}$  est la covariance entre les variables indicatrices  $I_{hi}$  et  $I_{hl}$  ( $I_{hi}=1$  si le centre  $hi$  est compris dans l'échantillon et  $I_{hi}=0$  dans les autres cas). Pour un échantillonnage à taille fixe et sans remise, le calcul de  $\pi_{hil}$  peut s'avérer complexe. Ainsi, si l'échantillon est de taille 2 dans la  $h^e$  strate,  $\pi_{hil}$  peut s'exprimer de la façon suivante selon le plan proposé par (Sarndal, Swensson et Wretman, 1992) :

$$\pi_{hil} = \frac{2Q_{hi}Q_{hl}(Q_h - Q_{hi} - Q_{hl})}{Q_h \left( \sum_{j=1}^{K_h} c_{hj} \right) (Q_h - 2Q_{hi}) (Q_h - 2Q_{hl})},$$

où  $Q_h = \sum_{j=1}^{K_h} Q_{hj}$  et  $c_{hj} = Q_{hj} (Q_h - Q_{hj}) / (Q_h (Q_h - 2Q_{hj}))$ . Le plan est tel qu'au premier tirage, le centre  $hi$  a la probabilité  $p_{hi} = c_{hi} / \sum_{j=1}^{K_h} c_{hj}$  d'être sélectionné et que sans remise du premier élément tiré (disons  $h, i_0$ ) le second élément a la probabilité  $p_{hi|h, i_0} = Q_{hi} / (Q_h - Q_{h, i_0})$ . La variance  $\hat{\text{var}}(\hat{N})$  peut se calculer par

$$\text{var}(\hat{N}) = E\left(\text{var}(\hat{N} \setminus \hat{N}_{hi}, i = 1, \dots, k_h)\right) + \text{var}\left(E(\hat{N} \setminus \hat{N}_{hi}, i = 1, \dots, k_h)\right).$$



Si nous considérons que les  $\hat{N}_{hi}$  sont estimés de façon indépendante, nous pouvons obtenir

$$\text{var}\left(E\left(\hat{N} \setminus \hat{N}_{hi}, i = 1, \dots, k_h\right)\right) \approx \sum_{h=1}^L \frac{K_h}{k_h} \sum_{i=1}^{k_h} \text{var}\left(\hat{N}_{hi}\right).$$

Par la technique de la linéarisation de Taylor, nous obtenons une approximation de la variance de l'estimateur  $\hat{N}$  pour la taille de la population d'UDI utilisant les PES durant les trois mois :

$$\hat{\text{var}}(\hat{N}) = \sum_{h=1}^L \left[ \frac{1}{2} \sum_{i=1}^{k_h} \sum_{l=1}^{k_h} (\pi_{hi} \pi_{hl} / \pi_{hil} - 1) \left( \frac{\hat{N}_{hi}}{\pi_{hi}} - \frac{\hat{N}_{hl}}{\pi_{hl}} \right)^2 + \frac{K_h}{k_h} \sum_{i=1}^{k_h} \hat{\text{var}}(\hat{N}_{hi}) \right].$$

## REMERCIEMENTS

Les auteurs remercient la P<sup>re</sup> Mary E. Thompson, de l'Université de Waterloo, et le P<sup>r</sup> J. N. K. Rao, de l'Université Carleton, de leurs suggestions et de leurs observations utiles. Ils sont aussi redevables aux membres du Groupe de modélisation et d'estimation VIH-SIDA du Centre de prévention et de contrôle des maladies infectieuses de Santé Canada pour les éléments de discussion qu'ils ont apportés.

## RÉFÉRENCES

- Addictions Services (2004), "Regina 2002 Report", *A Canadian community epidemiology network on drug use report*, Regina.
- Archibald, C. P., Jayaraman, G. C., Major, C., Patrick, D. M., Houston, S. M. et Sutherland, D. (2001), "Estimating the size of hand-to-reach population: a novel method using HIV testing data compared to other methods", *AIDS*, 15, pp. s41-s48.
- Brewer, K. W. R. (1963), "A model of systematic sampling with unequal probabilities", *Australian Journal of Statistics*, 5, pp. 93-105.
- CDC (2001), "HIV Prevention Strategic Plan Through 2005", Centers for Disease Control and Prevention, É.-U.
- Cochran, W. G. (1997), *Sampling Techniques*, 3<sup>rd</sup> Edition, New York: Wiley.
- Dehne, K. L., Adelekan, M., Chatterjee, A. et Weiler, G. (2002), "The need for a global understanding of epidemiological data to inform human immunodeficiency virus (HIV) prevention among injecting drug users", *LIV, Bulletin on Narcotics*, pp. 117-130.
- DesJarlais, D. C., McKnight, C., Eigo, K. et Friedmann, P. (2000), "2000 United States Syringe Exchange Program Survey", The baron Edmond de Rothschild chemical dependency institute, New York.
- Health Canada (2004), "Harm reduction and injection drug use: an international comparative study of contextual factors influencing the development and implementation of relevant policies and programs", <http://www.hc-sc.gc.ca>
- Korn, E. L. et Graubard, B. I. (1999), *Analysis of Health Surveys*, Wiley series in probability and statistics survey methodology section, New York: John Wiley & Sons.
- Lohr, S. L. (1999), *Sampling: Design and Analysis*, New York: Duxbury Press.

- MacDonald, M., Wodak, A. D., Ali, R. et al. (1997), "HIV prevalence and risk behaviour in needle exchange attendees: a national study", *Med. J. Aust.*, 166, pp. 237-240.
- Miller, C. L., Tyndall, M., Spittal, P., Li, K., Ralepu, A. et Schechter, M. T. (2002), "Risk-taking behaviours among injecting drug users who obtain syringes from pharmacies, fixed sites, and mobile van needle exchanges", *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 79, pp. 257-265.
- Peggy, M., Myers, T., Calzavara, L., Wallace, E., Major, C. et Degani, N. (2003), "Regional variation in HIV prevalence and risk behaviours in Ontario injection drug users", *Canadian Journal of Public Health*, 94, pp. 431-435.
- Roos, S. M. (1997), "*Probability Models*", 6<sup>ème</sup> éd. Londre: Academic Press.
- Salganik, M. J. et Heckathorn, D. D. (2004), "Sampling and Estimation in Hidden Population using Respondent Driven Sampling", *Sociological Methodology*,
- Sarndal, C., Swensson, B. et Wretman, J. (1992), "*Model Assisted Survey Sampling*", Series in Statistics, New York: Springer-Springer.
- Thompson, M. E. (1997), "*Theory of sample surveys*", Londre: Chapman & Hall.
- UNAIDS (2004), "2004 Report on the global AIDS epidemic – Executive Summary", Joint United Nation Programme on HIV/AIDS, Genève, Suisse.
- Vijayan, K. (1968), "An exact  $\pi_{ps}$  sampling scheme: Gneralization of a method of Hanurav", *Journal of the Royal Statistical Society*, 54, pp. 499-513.
- Weekes, J. et Cumberland, K. (2004), "Needle exchange programs {FAQs}", *Canadian Centre on substance abuse*, <http://www.ccsa.ca>