Research Article

Estimated Number of People who Inject Drugs in Ho Chi Minh City, Vietnam: Findings from a Two-survey Capture–Recapture Population Size Estimation Exercise

Nghia Van Khuu[1], Phuc Duy Nguyen[1], Giang Tong Le[2], Hoa Thi Yen Luong[3], Van Thi Thu Tieu[4], Hau Phuc Tran[1], Thuong Vu Nguyen[1], Meade Morgan[5], Abu S. Abdul-Quader[3,*], (iD)

[1]Department of HIV/AIDS, Pasteur Institute of Ho Chi Minh City, Ho Chi Minh City, Vietnam
[2]Program Development Office, United States Agency for International Development, Hanoi, Vietnam
[3]Division of Global HIV and TB Prevention, Centers for Disease Control and Prevention, Hanoi, Vietnam
[4]Department of HIV Prevention, Ho Chi Minh City Provincial AIDS Committee, Ho Chi Minh City, Vietnam
[5]Division of Global HIV and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA, USA

ARTICLE INFO

ABSTRACT

Background: HIV/AIDS program managers in Ho Chi Minh City (HCMC), Vietnam have always relied on the police reports and the UNAIDS Estimation and Projection Package for population size estimation of People Who Inject Drugs (PWID).

Methods: We used Respondent-driven Sampling (RDS) to implement a two-source capture–recapture study to estimate the population size of PWID in HCMC in 2017. The study was implemented in seven out of 24 districts and included men and women ages 18 years and older who reported injecting illicit drugs in the last 90 days, and who had lived in the city for the past six months. Estimates of the PWID population size for each of the seven districts were calculated accounting for the RDS sampling design. These were then adjusted to account for the district sampling probabilities to give an estimate for HCMC. Chapman two-source capture–recapture estimates of population size, based on simple random sampling assumptions, were also calculated for comparison.

Results: The estimates resulted in a population size for HCMC of 19,155 [95% Confidence Interval (CI): 17,006–25,039] using the RDS approach and 17,947 (95% CI: 15,968–19,928), using the Chapman approach.

Conclusion: The two-survey capture–recapture exercise provided estimates of PWID in HCMC – based on Chapman estimator and RDS approach – are similar. For planning HIV prevention and care service needs among PWID in HCMC, both estimates may need to be taken into consideration together with size estimates from other sources.

1. INTRODUCTION

With an estimated 245,337 people living with HIV, the epidemic in Vietnam remains concentrated among three key populations: People Who Inject Drugs (PWID), men who have sex with men and female sex workers [1]. Ho Chi Minh City (HCMC) is the largest city in the country with 9 million people. Population size estimates of PWID in HCMC range from about 8300, which is the number of PWID managed by the HCMC police department, to about 34,000 which is an estimate based on UNAIDS Estimation and Projection Package (EPP) [2]. The estimate provided by the police department is based on arrests and is considered by public health program authorities to be considerably lower than the actual number. On the other hand, UNAIDS estimates are considered by some policymakers and law enforcement officials too high.

The sizeable difference between these two estimates demonstrates the need for more reliable and accurate population size estimates of PWID in HCMC to help with program planning and allocations of resources for program activities. Population Size Estimation (PSE) of those affected by HIV/AIDS help policymakers and program administrators understand the scope of the HIV epidemic, help plan appropriate interventions, and allocate enough resources. PSE is used as denominators in calculating program coverage and producing national estimates and projections of the HIV epidemic in a country. In the following pages, we describe the findings from a study implemented to estimate the population size of PWID in HCMC, Vietnam. Data were collected in 2017–18.

2. MATERIALS AND METHODS

Using Respondent-driven Sampling (RDS) [3,4], we implemented a two-source capture–recapture study to estimate the population size of PWID in HCMC [5]. Study populations included men and women ages 18 years and older who reported injecting illicit

drugs (not prescribed for treatment purposes) within the previous 90 days, and who had lived in the city for the past 6 months. Implementation involved the following steps: the selection of the districts where data were collected; implementation of the first RDS survey (the captured survey), and implementation of the second RDS survey (the recapture survey).

2.1. Selection of Districts

In HCMC, 24 administrative districts are covering approximately 2100 km^2 with a total population of about 9 million. The UNAIDS EPP, estimated that there are 34,323 PWID in HCMC [1]. According to the public security office, there are 8214 PWID in the city. The public security disaggregates this number of PWID by the district. Using public security office data, we calculated the proportion of PWID in HCMC for each district and then applied those proportions to the EPP estimated number of PWID for the city to derive the number of PWIDs in each district. The HCMC public health authorities expressed interest in getting district-level estimates to compare the estimates derived from this study to the police estimates as well as UNAIDS estimates at the city and the district levels. We then ranked the districts from highest to the lowest estimated number of PWID in HCMC and categorized them into three groups: group one accounted for 50% of the estimated number of PWID in the city, and groups two and three accounted for 30% and 20% of the estimated number of PWID in the city, respectively. Selection of the seven districts assumed that seven districts cover at least 25% of the total estimates. The number of districts selected from each group was allocated proportionally to the estimated number of PWID in each group. Districts were then selected randomly from each group with equal probability. This resulted in the selection of seven districts (Table 1).

2.2. Capture–Recapture Survey using the Coupon

2.2.1. Sample size

The sample size calculation was based on the UNAIDS estimate of 34,000 PWID in HCMC. Logistical and cost considerations allowed for surveys to be done in seven of the 24 districts with an expected PWID population size of 34,000 * 7/24 or roughly 10,000 PWIDs across the seven districts. Assuming that 10–15% of first-round participants would also be second-round participants, we decided to base our calculations on estimating the proportion recaptured as 12.5% with a 95% Confidence Interval (CI) of ±2.5% including a survey design effect of 1.8. This gave an initial sample size of 673 * 1.8 or about 1211 per survey round using the simple asymptotic two-sided confidence interval for a binomial proportion. Based on available resources, we determined that we could sample 175 persons per district per round, giving a final sample size of 1225 total. Assuming that 12.5% of 1225 in the first-round participants also participated in the second round, the Chapman capture–recapture-based estimate [6] for PWID population size would be 9759 (95% CI: 8416–11,103) for the seven sampled districts and 33,460 (95% CI: 28,854–38,067) for the 24 districts in HCMC.

Table 1 | Distribution of people who inject drugs in Ho Chi Minh City, Vietnam by districts

District	EPP provincial estimate distributed across districts with Public Security proportion	Reported number of PWID: Public Security 2014	Percentage contribution of each district to EPP estimate (%)	Cumulative (%)	District selected
Binh Thanh	3823	915	11	11	Randomly selected 3 districts from this category
Cu Chi	2169	519	6	17	
District 3	2160	517	6	24	
District 6	2018	483	6	30	
Binh Tan	1843	441	5	35	
District 11	1834	439	5	40	
Tan Binh	1780	426	5	46	
Go Vap	1684	403	5	50	
District 10	1663	398	5	55	Randomly selected 2 districts from this category
District 1	1634	391	5	60	
District 12	1550	371	5	65	
Tan Phu	1488	356	4	69	
Binh Chanh	1295	310	4	73	
District 9	1258	301	4	76	
Hoc Mon	1149	275	3	80	
District 5	1124	269	3	83	Randomly selected 2 districts from this category
District 8	1074	257	3	86	
Phu Nhuan	1020	244	3	89	
District 7	806	193	2	91	
Thu Duc	794	190	2	94	
District 2	648	155	2	96	
District 4	598	143	2	97	
Nha Be	543	130	2	99	
Can Gio	368	88	1	100	
Total	34,323	8214	100	100	7

2.2.2. Sampling strategy

The size estimation activity was implemented in the seven selected districts and study sites were established in all selected districts for data collection. The selection of study sites was based on discussions with the district authorities and commune health center staff. Every attempt was made to make sure that the sites were accessible to the study participants (easy to find and affordable to travel to), but also discreet enough to not attract any attention. We had one study site in each selected district; the sites operated concurrently and were far enough apart, so those prospective participants were discouraged from participating more than once. Coupons with unique site identifiers were used for recruiting participants. At each study site, only persons with a coupon specific to that study site were eligible to participate. Each coupon contained information on coupon code, study site locations, hours of operation, telephone numbers, and a unique coupon number.

Local non-governmental organizations recruited four PWID to serve as seeds. These seeds were selected in terms of diversity regarding the following: age, area of residence, and length of time as PWID. After completing the eligibility assessment as well as the questionnaire, each seed received five referral coupons and was asked to give these coupons to their peers and invite them to participate in the study in their district. The peers who participated in the survey completed the same eligibility assessment and the questionnaire. They were subsequently given five coupons to distribute and invite their peers. The recruitment continued until the study site reached the expected sample size. Participants received 50,000 VND (~$2.5) to cover transportation costs and lost time. Each round of data collection required about 4–5 weeks to meet the sample size.

2.3. First Survey as the Capture Exercise

After the selection of the seven districts, data were collected which included eligibility assessment and individual interviews. RDS with the inclusion of coupons was used to recruit participants in each of the selected districts.

2.3.1. Eligibility

After the eligibility assessment, the study purpose was explained to the participants. After they verbally stated that they understood the purpose and content of the survey and agreed to be included in the study, they were interviewed using a structured questionnaire.

2.3.2. Interview at study sites

Interview questions included basic demographics, network size, and HIV testing history. The interview lasted about 15–20 min. Any individual that showed up at a study site without a valid coupon was not assessed or interviewed.

2.3.3. Fingerprint matching

To avoid potential double counts of a person participating more than once in the same study, we used encoded fingerprint biometric

authentication; the code could not be decoded to personally identify any participant. Fingerprint authentication of this type is expected to yield correct matches 99.9% of the time. If a person's fingerprint did not match a registered fingerprint in the system, the person was assessed for eligibility and data were collected. Only authorized study personnel had access to password-protected computers and files.

2.3.4. Referral of peers

After completion of the interview, each participant was asked to distribute recruitment coupons to their peers and request them to participate in the study.

2.4. Second Survey as the Recapture Exercise

Two weeks after completion of the first survey (capture), a second survey (recapture) like the first one, was conducted in the seven selected districts.

2.4.1. Eligibility

Eligibility criteria were identical to what was used in the first survey.

2.4.2. Interview at study sites

A brief interview, like the one used in the first survey, was conducted at the study sites within the selected districts. Any individual that showed up at a study site without a valid coupon was not assessed or interviewed.

2.4.3. Fingerprint matching

Like the first survey (capture exercise), fingerprint machines were used to avoid potential duplications (one person participates in the study twice) and to determine how many of the eligible participants are new captures and how many were recaptures. The fingerprint provided information on the number of new participants – individuals who were not included in the first survey (new capture), and the number of individuals who were recruited in both surveys (recapture). Codes generated from the fingerprints collected during the first survey (capture exercise) were matched to the codes generated during the second survey (recapture survey). If the codes generated during the second survey did not match the codes generated during the first survey and the participants were not duplicates during the second survey, they were considered new captures. If the codes generated during the second survey matched the codes generated during the first survey and the participants were not duplicates during the second survey, they were considered recaptures. The date of the individual interview was added to the codes generated during the first survey as well as the second. Both the codes and the dates were used to calculate the number of new captures and recapture during the second survey.

2.5. Human Subject Approval

The study protocol received ethical approval from the Institutional Review Board of the Pasteur Institute in HCMC, Vietnam. The protocol was also reviewed per the Centers for Disease Control and Prevention (CDC) human research protection procedures and was determined research, but CDC investigators did not interact with human subjects nor have access to identifiable data or specimens for research purposes.

2.6. Data Analysis

Two population size estimates were based on a capture–recapture approach [5]. The estimated number of PWID in HCMC for each selected district was calculated in two ways. The first was calculated using the raw sample counts with the Chapman estimator [6]. The second incorporated the RDS study design from the second round using the following formula:

$$N = \frac{n1}{p12}$$

where:

N = the estimated PWID population size;

$n1$ = the number of PWIDs recruited and interviewed in the first-round survey;

$p12$ = the proportion of PWIDs recruited and interviewed in the second-round survey who also participated in the first round (i.e., the proportion of second-round participants recaptured).

We calculated population size estimates of PWID for each district separately, then calculated an estimate for HCMC by summing the population size estimates of each district weighted by the inverse of the stratum specific sampling probabilities. The proportions $p12$ and associated confidence intervals for each district were calculated accounting for the RDS survey design using the Gile's Successive Sampling (SS) estimator in the RDS Analyst (RDS-A) software [7]. The personal network size was defined as the number of adults who inject drugs seen in the past 30 days. The value of $n1$ was determined for each of the seven districts as the number of persons captured in survey round one for that district. The lower confidence bounds for the population sizes by district were calculated by dividing $n1$ by the upper bound for $p12$ as calculated by RDS-A, and the upper bounds by dividing by the lower bound for $p12$ from RDS-A. The bounds for the HCMC total estimate were calculated based on the square root of the weighted sum of the variances for each district while retaining the asymmetry between the lower and upper bounds about the district population size point estimates.

For other estimated proportions such as the percentage of PWID who were tested for HIV in the past year, separate estimates were calculated for each survey round by pooling data across districts and treating them as a single RDS survey without regard to residence.

3. FINDINGS

3.1. Demographic and Other Characteristics

During the first survey, 1241 PWID were recruited from the seven selected districts (Table 2). The majority (87.1%) of the participants were men. The mean age was 36 years with 8.0% between the ages of 18 and 24 years and with a similar proportion (8.8%) 50 years or older. Almost half of the participants (47.1%) were between the ages of 35 and 49 years. More than half (56.9%) reported injecting within the last day and 32.1% reported injecting in the past 2 and 6 days. About half (49.1%) of them had been to an HIV testing site within the last 12 months. When asked about the number of adults who inject drugs in their personal network, 70.2% reported to have seen up to four people (who inject) within the past 30 days, 21.1% had seen between five and nine people who inject, and 6.9% had seen 10–19 people who inject.

In the second survey, 1328 PWID were recruited from the seven districts. Like the first round, the majority (83.7%) were men. The mean age was 37.2 years and 6.6% between the ages of 18 and 24 years and 9.5% 50 years or older. Just over half (51.1%) of the participants were between the ages of 35 and 49 years. About two-thirds (69.9%) reported injecting within the last day and 24.1% in the past 2–6 days. Just over one-third (36.4%) had been to an HIV testing site within the last 12 months. When asked about the number of adults they knew who inject drugs, 57.2% reported having seen up to four injectors, 31.1% knew between five and nine injectors, and 9.4% knew between 10 and 19 injectors.

3.2. Population Size Estimates

Of the 1328 participants who participated in the second-round survey, 1240 (93.4%) had fingerprint data collected to indicate whether they participated in the first survey round and were included in the population size analyses. The resulting population size estimates ranges ranged from 437 in Thu Duc district to 1813 in Tan Binh (Table 3). For the HCMC, the population size was estimated to be 19,155 (95% CI: 17,006–25,039).

4. DISCUSSION

Our study conducted in seven randomly selected districts estimated population sizes of PWID aged 18 years or older in HCMC, Vietnam. It assumed that all participants in each survey round were residents of the district where the survey was conducted. The total for the city was calculated by weighting the population size estimates for each district by the inverse of the stratum specific sampling probabilities.

The estimates based on this study are substantially lower than the 34,000 estimated by UNAIDS which is based on several sources such as Vietnam HIV Sentinel Surveillance (HSS) data, prevention program data and other sources such as studies conducted in the country. However, the UNAIDS estimation of the PWID population size has several limitations. HSS is conducted only in selected provinces and the selection of participants is purposive and convenient. Program data that are used as input for the calculation of estimates also do not accurately reflect the risk categories of those receiving various program services.

In addition, several changes have happened which may have contributed to the estimates that are lower than the UNAIDS estimates. During the last few years, the nature of drug use, as well as the number of PWID, had changed significantly. While the law enforcement

Table 2 | Characteristics of survey participants adjusted for respondent-driven sampling, Ho Chi Minh City, Vietnam, 2017–18

Characteristics	Survey 1	Survey 2
Total	1241	1328
Gender		
Male	1065	1135
	87.9% (85.8–90.0)	83.8% (81.2–86.4)
Female	175	193
	12.1% (10.0–14.2)	16.2% (13.6–18.8)
Missing	1	0
Age		
Mean, Median (IQR)	36.4, 35 (31, 40)	37.2, 36 (31, 42)
18–24	94	76
	8.0% (6.0–10.0%)	6.6% (5.0–8.2%)
25–34	449	447
	36.2% (32.8–39.6)	32.9% (30.2–35.5)
35–49	583	678
	47.1% (43.7–50.5)	51.1% (48.1–54.1)
>50	115	127
	8.8% (6.6–10.8)	9.5% (7.7–11.3)
Days since last injected		
Mean, Median (IQR)	3.8, 1 (1–3)	2.3, 1 (1, 2)
0–1	779	981
	56.9% (52.9–60.8)	70.0% (66.5–73.4)
2–6	324	281
	32.1% (28.7–35.6)	24.1% (20.9–27.3)
7–29	92	46
	7.4% (5.6–9.6)	4.2% (2.6–5.7)
>30	44	18
	3.7% (2.0–5.3)	1.7% (0.8–2.7)
Missing	2	2
Received VCT in past 12 months		
Yes	564	466
	49.1% (45.6–52.7)	36.4% (33.1–39.8)
No	672	862
	50.9% (47.3–54.4)	63.6% (60.2–66.9)
Missing	5	0
Number of VCT[a] visits in past 12 months (*n* = 564 for survey 1, *n* = 466 for survey 2)		
Mean, Median (IQR)	1.5, 1 (1, 2)	1.4, 1 (1, 2)
1	391	327
	67.3% (62.1–72.5)	73.2% (68.4–78.0)
2	101	104
	20.6% (16.4–24.9)	21.5 (17.0–25.9)
>3	67	31
	12.1% (8.7–15.5)	5.4% (3.5–7.2)
Missing	5	4
Number of known PWID living in HCMC and seen in the past 30 days		
Mean, Median (IQR)	4.2, 3 (2, 5)	5.3, 4 (3, 5)
1–4	548	469
	70.0% (67.7–72.3)	57.2% (54.1–60.0)
5– 9	386	491
	21.1% (19.2–23.0)	31.1% (28.7–33.6)
10–19	209	259
	6.9% (6.2–7.6)	9.4% (8.4–10.4)
>20	98	109
	2.0% (1.8–2.2)	2.4% (2.1–2.7)

[a]Voluntary counselling and testing for HIV.

personnel are always present on the streets, they also conduct "clean the street and the community" campaigns in the city, which drives the PWIDs out of the city to the surrounding rural communities. In general, there are also reports of increase use of non-injection methamphetamine in HCMC and Vietnam. However, no studies have so far been conducted to verify these reports.

Table 3 | Population size estimates of people who inject drugs in Ho Chi Minh City, Vietnam

Calculation details	Survey district							PWID total[a]
	Binh Chanh	Nha Be	Quan 3	Quan 6	Tan Binh	Tan Phu	Thu Duc	
Participants in survey round one	175	175	188	178	175	175	175	
Participants in survey round two	187	142	206	187	188	179	151	
Participants in both round one and two	54	38	58	28	21	52	60	
Chapman population size estimates	601 (491–710)	644 (494–795)	662 (544–780)	1159 (810–1509)	1511 (968–2054)	597 (485–709)	438 (369–506)	17,947 (15,968–19,928)
RDS percentage of round two participants also in round one (95% CI)	28.1% (21.4–34.8)	26.2% (19.5–33.0)	30.8% (22.4–39.1)	14.5% (8.6–20.3)	9.7% (4.7–14.6)	27.0% (19.3–34.8)	40.0% (31.0–49.0)	
RDS population size estimate	623 (503–819)	667 (530–899)	611 (481–838)	1227 (879–2079)	1813 (1198–3739)	647 (503–907)	437 (357–565)	19,155 (17,006–25,039)

[a]Calculated as a weighted sum of the district estimates.

The estimates are also higher than those from HCMC police records. Police records are usually based on the arrests. Since the police "clean the street and the community" campaign drives the number of people who use drugs out of the city, the police records generate lower estimates of PWID.

The findings highlight that men are the majority of the PWIDs in HCMC and around half of them are between the ages of 35 and 49 years. It also reflects the aging population of PWID in HCMC and a possible shift in non-injection methamphetamine use among the younger population. Injection frequency is high. More than in round one and more than two-thirds in round two reported having injected within the last day. About half of the participants in round one and more than 60% in round two had not visited an HIV testing site within the last 12 months. This may indicate that either they were not reached by program activities or they already knew their HIV status.

5. LIMITATIONS

Because findings are based on data collected from populations involved in behaviors that are illegal, stigmatized and discriminated, there are several limitations to the findings. First, the standard capture–recapture approach requires meeting several conditions to assure result validity including that each person in the PWID population has an equal probability of being captured, of being recaptured, and that the two samples are independent. The corresponding assumptions are that the dual RDS sampling approach is more complicated and unlikely to be fully met. While some of the assumptions are partially mitigated by estimating the proportion of round two participants who participated in round one while accounting for the RDS design, other assumptions were not. For example, no adjustment was made to account for differing probabilities of participating in the first-round survey nor the likely lack of independence in participation in the two survey rounds. Second, seed selection may have influenced the findings. Some of the seeds in the first survey were recruited from one of the methadone maintenance treatment programs. This was not done in the second survey. That may have influenced some of the random assumptions of the recruitment by the seeds.

6. CONCLUSION

We expect the new PSEs, together with other estimates from the RDS surveys, will be beneficial to program planners in HCMC. The two available size estimates – UNAIDS estimates and police estimates – are wide apart and the program planners had difficulty in accepting one over the other. The estimates from the present study are more up-to-date and based on field investigation. For planning HIV prevention and care service needs among all PWID, both estimates may need to be taken into consideration together with size estimates from other sources. Future studies using better population size estimation methods, such as three-source capture–recapture, may need to be conducted to provide improved estimates for the number of PWID in HCMC.

CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest.

AUTHORS' CONTRIBUTION

FUNDING

REFERENCES

[1] Vietnam Authority of HIV/AIDS Control. Department of HIV/AIDS Prevention and Control. Hanoi, Vietnam: VAAC, Ministry of Health; 2018.

[2] Ho Chi Minh Department of Public Security. Distribution of people who inject drugs and female sex workers in Ho Chi Minh city. Ho Chi Minh City: Ho Chi Minh Department of Public Security; 2016.

[3] Heckathorn DD. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. Soc Probl 2002;49;11–34.

[4] Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. Soc Probl 1997;44;174–99.

[5] UNAIDS/WHO. Guidelines on estimating the size of populations most at risk to HIV. Geneva, Switzerland: UNAIDS—Joint United Nations Programme on HIV/AIDS; 2010.

[6] Chapman DG. Some properties of the hypergeometric distribution with applications to zoological sample censuses. Berkeley: University of California Press; 1951.

[7] Handcock MS, Gile KJ, Mar CM. Estimating hidden population size using respondent-driven sampling data. Electron J Statist 2014;8;1491–521.