Consultation on

Network scale-up and other size estimation methods from general population surveys

Co-convened by

UNAIDS & the US Office of the Global AIDS Coordinator

New York City, March 28-30th, 2012

TECHNICAL REPORT AND RECOMMENDATIONS

Contents

Acronyms

AIDS	Acquired immune deficiency syndrome
AIS	AIDS Impact Survey
BSS	Behavioural Surveillance Survey
CDC	Centers for Disease Control and Prevention
DHS	Demographic and Health Survey
FSW	Female sex workers
HIV	Human immunodeficiency virus
IDU	Injecting drug users
MOH	Ministry of Health
MSM	Men who have sex with men
MSW	Male sex workers
NSU	Network scale-up
NSUM	Network scale-up method
OGAC	US Office of the US Global AIDS Coordinator
PNS	Personal network size
PRM	Proxy respondent method
UNAIDS	Joint United Nations Programme on HIV/AIDS
USAID	United States Agency for International Development

Background

Estimates of the size of key populations, including female sex workers (FSW) and their clients, people who inject drugs and men who have sex with men (MSM), are needed to better understand HIV epidemics and plan for programme responses and prevention strategies. The network scale-up (NSU) method is an approach that may be used for population size estimation. Key advantages of this method over other methods include the use of a random sample, results that can be compared over time, validation checks, and importantly, network scale-up does not require the key populations to report on their own highly stigmatised and sometimes illegal behaviours. The method can also estimate the size of several key populations within a single survey. In August, 2008 a consultation was held to determine how feasible the method was for estimating the sizes of key populations at risk for HIV. The consultation found that the method held a lot of promise and since that meeting a number of pilot tests have taken place.

In March, 2012, a consultation was convened by UNAIDS and the US Office of the Global AIDS Coordinator (OGAC) including Centers for Disease Control and Prevention (CDC) and the US Agency for International Development (USAID), to review and discuss applications of these methods in different settings, the issues that limit their use and ongoing work to improve and enhance their use.

The network scale-up method was developed by anthropologists, social network analysts and mathematicians and has been used to estimate the size of populations that may be difficult to count. The critical assumption of this method is that the people who you *know* are, on average, representative of the general population that you belong to. The estimate requires three pieces of information: the number of key population groups known (collected in the survey), the personal network size (PNS) of the respondent (estimated from the survey) and the total number of people in the general population. The PNS of the respondent can be identified through the back estimation of populations of known sizes or through summation.[1]

However, the method has a number of potential biases:

- **Transmission bias:** An individual might *know* someone, but not be aware of all of their behaviours (for example, sex work or drug use). The underlying assumption of this method is that everyone knows everything about their acquaintances.
- **Relative network size:** Members of the key populations may have a personal network size that is different from the general population. For example, if the key population has a smaller personal network size they will be underestimated in the general population
- **Reporting bias:** People may fail to accurately report the populations in question or may be reluctant to do so because of the stigma surrounding the behaviour of the population
- **Barrier effects:** A fourth bias due to non-random mixing or uneven distribution of populations (physical barriers such as geography and social barriers such as sex or race). This may cause variation in the likelihood that respondents know people used in the known populations to estimate the personal network size which can lead to bias. In addition, there could be barriers between the key populations and the respondents that affect the likelihood that a respondent knows someone in the key populations.

Consultation Participants

The consultation was attended by 31 experts including technical experts who have developed and refined the methods, individuals who have used the methods in country applications, social scientists, and the agencies who have supported the use and development of these methods. The list of participants is included in Appendix I.

Consultation Approach

The consultation featured presentations and discussions on application of the methods in eight countries, discussion on specific technical issues and discussions on the key programmatic recommendations. Recommendations for guidance when using these methods and a plan for further research were derived at this meeting. The meeting agenda is included in Appendix II.

Consultation Objectives

The specific objectives of this consultation were to:

- Agree on issues that limit the use of these methods and create a research plan for addressing these issues
- Produce guidance for countries on how to use the methods

A list of the methodological issues and questions on network scale-up will be created and made available to researchers and implementers of the method so that we can continue to improve the method. In addition, a guide will be finalised providing information on how to implement the method including implementation guidance resulting from this consultation. The document will also provide links to questionnaires, manuals, reports and other sources of information in order to support future applications of this method.

1. Country applications of the network scale-up method

A number of countries, across a range of diverse settings, have applied the network scale-up method to estimate the size of key populations at higher risk of HIV infection. Brief descriptions of the country applications presented at the Consultation are included herein.

1.1 Ukraine

Key Features of the NSU applications in Ukraine:

Survey year: 2009 and 2011
Populations: IDU, FSW, MSM
Approach for measuring PNS: Known population (validated with separate summation survey in 2009)
PNS value: 2009 known population = 193 (adjusted = 175), 2009 summation = 157, 2011 known population = 175
Adjustments: For *level of prestige* (bias due to stigma) and *social visibility*

In 2009, there was a special task to produce size estimates of key populations at the regional and national level. Size estimates from the NSU method and the multiplier method were compared. The known population approach with 22 groups of known populations was used to derive the PNS which was validated with summation used on a separate, nationally representative sample.

For estimating the PNS, outliers were removed (5% due to extreme high or low values) and nine groups that were estimated inadequately were excluded. The following adjustments were made to the estimates obtained:

- *Level of prestige:* It was found that across all three key populations the lower the level of respect for a particular group, the lower the number *known* within that group thus a correction coefficient was calculated for adjustment.
- *Level of social visibility*: A correction coefficient was calculated to address transmission bias (note this adjustment was not used in the final estimation for 2009).

Compared to the multiplier method estimates, after adjustment for the level of prestige, the NSU method resulted in comparable, but slightly larger size estimates for FSW and IDU but much lower size estimates for MSM. For 2009, population size estimates using the multiplier method were used as the national estimates and size estimates obtained from NSU were used to identify the potential range.

In 2011, the NSU method was not formally done, but six questions for known population groups were included to obtain new estimates for comparison. The same correction coefficients for level of prestige were used for adjustment. A new approach was used to address transmission barriers by asking key populations (IDU/FSW/MSM) for each *known* whether that person (the "alter"), knew the key population was an IDU/FSW/MSM, and this information was used to derive a correction coefficient. When compared to estimates derived using the multiplier method, even with the adjustments, the NSU method again produced much lower size estimates for MSM, and somewhat comparable estimates for IDU. Note that while both correction coefficients were applied, they were done independently to produce different adjusted estimates. Both corrections could be used together, but you would first need to identify how to correct for the correlation.

1.2 Moldova

Key Features of the NSU application in Moldova:

Survey year: 2009 and 2011
Populations: IDU, FSW, MSM
Approach for measuring PNS: Known populations

PNS value: 2009 known population, Right bank = 299; 2011 known population, Right bank = 289 (additional differences by north/south and urban/rural)
Adjustments: Tried to use prestige adjustment for response bias, age adjustment for relative network size

Sample:

- 2009 estimates for 5 regions in the Right bank, N=2207
- 2011 estimates for 8 regions in the Right bank, N=3991 and separately, a representative sample of the Left bank, N=600

A "respect" adjustment was calculated, similar to the prestige adjustment used in Ukraine, which did not work therefore other differences were investigated. It was found that younger respondents reported more persons with risk behaviours in their networks than older people, thus an age adjustment was used. There were also geographic differences in the average personal network size with differences observed in the north compared to the south and rural populations having a greater PNS than urban.

For 2009, the results from NSU underestimated the sizes of key populations (even after adjustments) compared to estimates derived using multiplier methods. This led to the development of an *undercount coefficient* (multiplier results/NSUM results) calculated for each key population and extrapolated and applied to remaining regions that had not conducted size estimation using the multiplier method. This approach allowed for sub-national and national estimates of all three key populations for both the Right bank and Left bank, an improvement for areas where there were previously no size estimates at all.

Comparing the Right bank results from 2009 to 2011, the average personal network sizes were quite similar; however, the key population size estimates were all much larger in 2009 (particularly for IDU) compared to 2011. This may be due to differences in the sampling or localities for data collection and is undergoing investigation.

1.3 Kazakhstan

Key Features of the NSU application in Kazakhstan:

Survey year: 2012
Populations: IDU, FSW, MSM in Almaty and Almaty region
Approach for measuring PNS: Known populations and summation
PNS value: 43 in regions around Almaty city and 81 in longer distance regions
Adjustments: create social distance measure (not yet completed)

The NSU pilot project (N=1199) in Kazakhstan covers three conventional regions chosen to address geographical disparities. The results are pending, however, preliminary analysis of the summation approach revealed somewhat similar crude data for urban IDU in the Almaty region compared to other size estimation methods, but less agreement for crude MSM data. Adjustments are planned for age and "social distance" (to address barrier effects and transmission effects) compared to a reference population.

1.4 Japan [2]

Key features of the application of NSU in Japan:

Survey year: 2009
Populations: MSM
Approach for measuring PNS: Known populations (3 populations)
PNS value: 174 males (364 total))
Adjustments: "coming out" adjustment
Survey type: Internet survey

High cost and difficulty accessing MSM are the two main barriers for size estimates of this key population with elevated risk for HIV, thus the NSU method was used to estimate MSM in Japan. A pilot study was conducted to identify the appropriate groups of known size, followed by the 2009 study (N=1500). A "coming out" adjustment (transmission rate) was used to address transmission error, calculated as the average number of persons to whom MSM have "come out" (or have told about their behaviour) divided by the average personal network size. A key assumption of this application was that MSM have the same network size as the general population. Adjusted results from NSU were comparable, but higher, to those found in previous studies using other methods. From this study, the NSU method was deemed a rapid, simple, cost-effective approach for estimating the size of MSM in Japan.

1.5 Thailand

Key features of the NSU application in Thailand:

Survey year: 2010
Populations: IDU, MSM, FSW, MSW and migrant workers
Approach for measuring PNS: Known populations and summation
PNS value: Known population – varied by region range of 53-247, summation – range of 135-697
Adjustments: None

In Thailand, sub-national estimates were derived using both summation and known population approaches for comparison. Sub-nation estimates were aggregated for national estimates. A small qualitative study was conducted to identify the mutually exclusive, culturally relevant, categories for summation (18 categories defined).

The personal network sizes derived using the summation approach were much smaller than those derived using the known population approach, across all region, and the degree of regional variation was greater than expected. For each key population, the estimates derived using the NSU method were compared with other data available and size estimates from different methods. Note that no adjustments were made to the NSU estimates.

To estimate the size of MSM using NSU, two different definitions were tested – MSM and "gay". Both the summation and known population approaches largely underestimated population sizes using "MSM" as opposed to "gay" which highlighted the importance of using culturally relevant definitions. The national consensus estimate for the MSM was derived from BSS (~3%) which was the approximate midpoint between the estimates from the summation and known population approaches using the term "gay".

For migrant workers, both approaches resulted in NSU estimates that were much lower than National Census data. This may have been due, in part, to many migrant workers living in camps and not being included in the random sample from households, which is an important consideration for future applications.

Overall, there were quite different population size estimates obtained using the summation approach compared to the known population approach; however, it was possible to use these two estimates to inform confidence bounds to compare with estimates derived using other approaches. Across the different key populations, the NSU method was used for the national IDU size estimates, while other estimation approaches were used for consensus estimates for migrant workers, FSW and MSM.

1.6 Chongqing, China

Key features of NSU application in Chongqing:

Survey year: 2010
Populations: MSM, IDU, drug users, current FSW, ever FSW, clients and those with multiple sexual partners in Chongqing, China
Approach for measuring PNS: Known populations
PNS value: 330 (adjusted values 287-290)
Adjustments: Attitude coefficients

In China, the NSU method was applied in three areas with completion in the Chongqing Municipality (N=~3000, 60% female) where approximately 28 million people reside. A pilot study was conducted to identify the known population groups and the results obtained were compared with the actual values and included in the final study if agreement was within a specified limit.

There were various adjustments to the PNS:

A. Crude
B. Adjusted by socio-demographic distribution to account for differences by district and sex(weighted by population)
C. Sensitivity analysis for known population categories, removing 11 of the 19 known populations that were not estimated well
D. Removed outliers

The estimates were adjusted with weighting by "attitude coefficients" defined for each key population. Compared to local consensus estimates, the NSU estimates were much lower, particularly for drug users and clients; however, differences in the definitions used may account for some of this difference). The main challenges using NSU in Chongqing were the random sampling procedure in a large city and quality control in a household survey – usually the most social person would complete the survey and often done with family members around.

1.7 Curitiba, Brazil [3-4]

Key features of the NSU method applied in Brazil:

Survey year: 2010
Populations: heavy drug users in Curitiba, Brazil
Approach for measuring PNS: Known populations and summation
PNS value: Known population = 184, summation = 140
Adjustments: Game of contacts to adjust for transmission rate and relative network size

For the NSU application, both the known population and summation approaches were used to estimate the PNS. Summation yielded a smaller PNS compared to the known population approach. A comparison of the root mean square error and mean absolute error of the two approaches led to the use of the known population approach for the scale-up estimate.

Comparison of five different estimates from four different data sources:

1. **Direct estimate, 2004** – MOH general population survey, 2004
2. **Direct estimate, 2010** – Survey of general population, 2010
3. **NSU estimate, 2010** – Survey of general population, 2010
4. **Generalised scale-up, 2009/2010** – Survey of general population, 2010 and Survey of heavy drug users 2009
5. **Multiplier, 2009** – Administrative records from drug treatment programme, 2009 & Survey of heavy drug users, 2009

The results from comparing the five different population size estimates for heavy drug users in Curitiba, Brazil illustrated that direct estimates and multiplier estimates yielded much lower estimates that the than generalised scale-up and NSU estimates.

The game of contacts was used to address transmission barriers. In this game, the key population (heavy drug users) drew cards with names on them and were asked whether they *knew* individuals with that name, and if so, whether the *known* (or "alter") identified knows that the respondent uses drugs or not, and whether the alter is a heavy drug user or not.

The results from the game of contacts were used to make two adjustments:

1. Information transmission rate: What is the probability that someone connected to a heavy drug user (key population) knows that they are a drug user?
2. Relative network size**:** To identify the relative personal network size of the heavy drug users (key population) compared to the general population.

These adjustments were then used to calculate what was termed the *generalised scale-up estimate* which was double the crude NSU estimate. It is worth noting that adjusting for the transmission rate will almost always increase the NSU estimate and if the key population has a smaller network than the general population, this will further increase the estimate. In this example, the adjustments had more of an impact on the final population size than the data collected in the survey. Thus the adjustment calculations should be carefully considered before applying to the estimates.

The following are findings from the Curitiba study that may or may not generalise to other NSU applications:

- PNS from known population approach higher than with the summation approach
- Known population approach outperformed summation on validation checks
- Scale-up estimates were larger than direct estimates
- Scale-up estimates were larger than multiplier estimates
- Generalised scale-up estimates were larger than scale-up estimates

1.8 Rwanda

Key features of NSU application in Rwanda:

Survey year: 2011
Populations: MSM, SW, IDU and clients
Approach for measuring PNS: Known populations and summation
PNS value: Using basic definition known populations 251, using basic definition summation 168, using meal definition known populations 108, using meal definition summation 71.
Adjustments: Pilot study of game of contacts to adjust for transmission rate and relative network size (not used as numbers were too small and not representative)
Survey: NSU embedded within a Demographic and Health Survey (DHS)

The objectives of the NSU application in Rwanda were as follows:

1. Feasibility of NSU and proxy respondent method in sub-Saharan Africa
2. Feasibility of NSU as part of a Demographic and Health Survey
3. Population size estimates of FSW, clients, MSM and IDU
4. Comparison of these methods to existing size estimates

For the application in Rwanda, both the NSU method and a similar method, the proxy respondent method, were tested. For the NSU method, both the known population and summation approaches were applied. A workshop was held to identify the known population groups and data sources, and to obtain a set of categories that were exhaustive and mutually exclusive for summation. Culturally relevant categories were defined for the summation approach. The terminology used to define the key populations was also reviewed and discussed.

Two different definitions of *known* were tested for both NSU approaches:

1. Basic definition: People who live in Rwanda that you know and they know you by sight and name and that you have had contact with in the last 12 months.

2. Meal definition: People who live in Rwanda that you know and they know you by sight and name and that you have shared a meal or drink with in the last 12 months.

Results: The known population approach was more accurate than summation for estimating the sizes of known populations and the relative error was lower using the meal definition as opposed to the basic definition (note that the PNS from the basic definition was more than twice the PNS from the meal definition). High levels of stigma and shame were reported for all four key populations which will result in transmission error. The game of contacts was conducted and revealed that a large information transmission bias is likely as only 20% (in a small sample of MSM) had revealed to their acquaintances that they had sex with other men. There is also a response bias as respondents may not want to admit they know people in these populations. The data obtained will need these further adjustments and are likely an underestimate.

Overall, it was feasible to use these methods in sub-Saharan Africa. The key challenges identified were that it was difficult to identify mutually exclusive categories for summation and to identify an appropriated definition for sex workers. There was also a lack of data for which to derive adjustments.

1.9. Conclusions from country presentations

Experiences were presented from eight different countries on the use of network scale-up. A number of key lessons can be drawn from these presentations.

First, all but one of the applications used some sort of adjustment to the results. This is appropriate given the known biases, however it was recognized that in a number of instances the adjustment became more “important” than the collected data. In other words the adjustment was more likely to determine the final population size estimate than the network scale-up method. Given the uncertainty around some of the adjustment methods these values need to be considered carefully.

Second, the method was feasible to implement in a diverse set of circumstances with diverse survey methods. The creativity around the survey methods should be encouraged to find the optimum survey method (although this will vary by setting).

Third, in none of the settings was there a reliable comparative value against which to compare the results of the network scale up estimator. There were other size estimates with their own limitations, minimising the usefulness of the comparisons.

Fourth, challenges continue on how to define the key populations at increased risk of HIV infection and how to ensure that survey respondents understand the definition.

2. Proxy respondent method

The proxy respondent method (PRM) is a new method that is conceptually similar to NSU and may be useful for estimating the size of key populations in a range of settings. In essence, it is a simplified NSU method that does not require estimating the personal network size. The rationale of this method is that if a random sample of respondents (proxy respondents) are selected and asked about the characteristics of people they know, these alters yield a representative sample after adjusting for the sampling frame. The method works by asking the respondents questions about a person they know with a randomly chosen name. For a full description of the method see the Rwanda ESPHS report.

Benefits of the proxy respondent method:

- No need to estimate personal network size
- An interview about specific individuals may be cognitively less demanding
- Individual-level data/analysis may be further stratified
- Multiple traits can be probed in the same alter (age group, marital status) as can lifetime or current high risk behaviour

Limitations

- Subject to the same limitations as NSU
- Much smaller "sample size" of alters
- Rwanda is the only application of this method to date
- Potential for cluster effect if there was more than one alter per proxy respondent
- Ethical considerations if named alter is identifiable because of unique name or other possible reason
- Names chosen can introduce bias including age, race, behavioural and socio-economic biases

Methods used for application in Rwanda

Rwanda represents a unique situation in sub-Saharan Africa where there is a national ID database of all named individuals. Pre-survey, a representative list of first names was generated and 100 names were randomly selected (after removing names that were androgynous, had multiple spellings or nicknames or had a frequency <1% of the most common name). Name cards were created with five names of the same sex on each card. The survey participants, or *proxy respondents,* chose one card each for male and female and were probed if they *knew* anyone with any of the names shown on the card. They were then interviewed about a maximum of two alters per card (the ones they knew best).

Results from Rwanda application

In Rwanda, 4669 proxy respondents identified 5732 alters. Those who were married and over 25 years old identified more alters. One-third of the sample did not identify any alter. A random sample of alters would be expected from a random sample of proxy respondents, thus demographic data for alters and proxy respondents were compared to DHS data and were found to not be in agreement. It is likely that bias was introduced as a result of one-third of the proxy respondents not identifying any alters. A surprising finding of the PRM was the identification that women buy sex from male sex workers in Rwanda, a concept that was recognised locally.

The insights gained from this study suggest that even with nearly 5,000 people, the sample might have been too small thus it may be necessary to add PRM onto a larger study such as DHS or AIS. Strategies are needed to increase the number of alters per proxy respondent in order to reduce the bias.

Comparison of PRM with NSUM

In Rwanda, the PRM was administered after the NSUM which allowed for comparison. Overall, there was a fairly consistent likelihood of respondents *knowing* a high risk individual in each method. However, most NSU respondents who *knew* key populations in their personal network did not

identify a high risk alter in the proxy respondent method because they were not probed about that specific person. However, a small proportion (2.6%) of NSU respondents who did not identify a key population in their network did identify a key population alter with the proxy respondent method.

It was discussed that while there are similar biases in both PRM and NSUM, when you instruct people to only tell them about the two people that "you know best" as was done in the PRM application in Rwanda, this can then introduce bias, particularly for stigmatised behaviours as respondents might be less likely to identify alters with stigmatised behaviours in situations where they know multiple people with the same name.

It is recommended there is follow-up for the small proportion who reported *knowing* individuals with high risk behaviours using proxy respondent but not using network scale-up in order to better understand why this occurs.

3. Adjustments to reduce bias and error

3.1 Correcting for under-recall [5]

Problem: There is a tendency to underestimate the prevalence of more common names

Solution: Develop a calibration curve

Assume that respondents report unbiased counts from a recalled network which is a subset of an actual network. Develop a calibration curve that interpolates between the recalled network and an actual network in which there is a linear relationship for small groups (assume fairly accurate recall) and then a square root relationship for large groups (assume less accurate recall). The point at which you switch from the linear relationship to the square root is determined by the data.

The calibration curve acts to adjust the recalled network to the actual network size. There is little adjustment for small groups (for example, less common names) which have a tendency to be recalled and a much greater adjustment for large groups (for example, common names). The caveat of this adjustment is that there are some unique situations where it is a rare situation but it sticks in your mind very well (for example, suicide) and this would not fit this trend. Note that applying the calibration curve will increase the size of the personal network size.

3.2 Reducing bias due to barrier effects [5]

Problem: Individuals systematically *know* more or less members of a key population than expected under random mixing.

Solution: Choose known populations according to the *scaled-down condition*

In order to reduce bias, the alter population should be representative of the general population. But the popularity, for example, of a given name often changes over time which can introduce barrier effects between the respondents and the people they know with that name. However, the population of hypothetical known persons can be determined through survey design whereby a balanced and complementary profile of known population groups is selected. The scaled-down condition is satisfied if the combined set of *known* groups, or alters, is a "scaled-down" version of the randomly selected population. For example, a heat map of the popularity profiles of names over time was constructed pre-survey and a set of names for the known population groups was selected that was in agreement with the scaled-down condition, i.e. it was a complementary profile to the random sample.[5] This was demonstrated to reduce bias due to barrier effects.

Tips to reduce bias when selecting known population groups

1. Use populations which minimise barrier and transmission effects, for example first names
2. Use of rare groups will minimise recall bias
3. Select complementary profiles (sampling approximately the same proportion of each alter group).

3.3 Reducing bias due to information transmission error

Problem: Respondents might not be aware of key behaviours of the people they *know* which results in information transmission error.

Solution: Use the game of contacts to among a representative sample of key populations. Ask questions to estimate the probability that their acquaintances know about their behaviour. The underestimate is a function of the rate at which the information is transmitted and the personal network size.

There are many factors contributing to information transmission bias. A household survey might not capture key population member or their contacts if they are not living in houses. Individuals might

not self-define as a sex worker or MSM, or they might not disclose this behaviour to various reasons. There can also be confusion over the definition used, for example, not using culturally appropriate terminology (for example "gay" vs MSM) or not clearly defining what is explicitly referred to as "sex work".

The PLACE study in China[6] identified venues in China where sex work occurred and then interviewed both workers and patrons to identify how different definitions led to different size estimates. This study highlighted how individuals would report receiving money for sex (as recently as the same day) but did not identify as a sex worker – 44% of those who reported selling sex in the last 12 months. These results have been found in other countries as well and for other behaviours, for example IDU who report that they do not share needles, but do so with a regular sharing partner or sexual partner. This error in self-identification will likely also affect how well respondents identify their acquaintances as having the specified behaviour.

3.4 Differential network size

Problem: People show up in the set of alters proportionate to their network size; the personal network size of a key population may not be the same as the personal network size of the general population – *differential popularity.*

Solution: Again using the game of contacts the researcher can collect data from the key population to estimate the personal network size of the key population. An expression can then be set up using the data obtained for the information transmission rate and for differential popularity combined to produce an adjusted prevalence of the key population in the general population. The adjustment for information transmission error and differential network size has been termed the *generalised scale-up approach*. A caveat is that the assumption is that these two adjustments are independent. If these are correlated, this will affect the estimate.

3.5 Reporting bias

Reporting bias is due to social desirability bias. For example, if the respondent does not tell the truth because of the stigma associated with the behaviour. This bias is likely to be highly correlated with the information transmission error. However, in addressing this there will be differences across countries, likely also within countries, and across key populations. As a result, the findings will not necessarily be generalisable or applicable across different settings. For example, in Moldova the distribution after calculating correction coefficients for *level of respect* was inverted when compared to the results from Ukraine, thus adjustments were made by age and urban/rural instead.

To reduce this bias and improve the estimates, the focus should be on improving the data collection efforts and reducing the potential for social desirability bias and not on measuring the stigma as the latter is far more complicated to measure. It may be useful to consider alternatives to face-to-face interviews.

3.6 Variance estimation and non-sampling error

Survey error depends on sampling error and non-sampling error.

There are two ways to estimate variance from a sample:

1. Analytic: Standard formulas derived for you which give you the confidence intervals to derive (not applicable for these methods)
2. Bootstrap: A more flexible approach which indicates how close or far the results could have been. The simplest form is a simple random sample with replacement.
   Note that the bootstrap includes sampling error but not non-sampling error.

With the NSU method, even when the bootstrap is used to account for the sampling error, we do not get credible intervals. This is likely due to different kinds of non-random sampling error (barrier effects, transmission effects, response errors). Use of the bootstrap to estimate the variance from previous data from the US, from the Brazil data for heavy drug users and from Rwanda revealed that the chance that the bootstrap bounds capture the true value ranged between 10%-25% among the studies (instead of 95%) which suggests there is a lot of non-random sampling error.

There are many sources of error including:

- Who you interview (sampling, this may be the largest source of error)
- What they know
- What they tell you (intentionally)
- What they tell you (unintentionally)
- Making inference from the set of alters to the set of egos (in a regular survey you learn about the people you talk to, here you learn about alters who may be very different from the people who you are interviewing; however, the underlying assumption is that they will be similar)

What if you do the game of contacts and use the generalised scale-up estimator or other adjustments? You will need to account for uncertainty in these adjustments.

The proposed procedure to account for this uncertainty takes the general population sample and resamples to create a general population replicate. Then the target population is resampled in the same way is which they were recruited to create a target population replicate. The two replicate samples are then paired up to get a replicate sample estimate to arrive at the estimated variance of the generalised scale-up estimator. Note that there is propagation of uncertainty thus if there is a poor estimate for the adjustment the estimate can be made worse.

Discussion

When is not including the true value a problem?
For many areas where there are currently no estimates at all, this might be okay as it at least provides something, a starting point.

From a funding perspective, if there is this much error, is it necessary?
We already have rough estimates, regional estimates, etc which may be as good (or better?) but without the additional cost. Note that there are two separate issues:

1. The programmatic issues are and what is needed in order to be able to plan programmes and resource allocation
2. The issue of identifying what the best estimates are.

3.7 Sample size

How big of a sample do I need? Sampling errors get smaller (but diminishing returns) as sample size increases, but non-sampling errors do not.

The mean square error includes variance and bias and the bias term becomes more important the bigger the sample gets. You can think of the estimate as the ratio of these two means. Both are estimated with error and the precision will depend on the amount of variability. Because we do not know enough about the bias we cannot *really* get the mean square error. Note that the summation approach basically assumes there is no error. It is very difficult to quantify error with summation.

Can we use NSU with smaller sample sizes?
While there are diminishing returns, if the trait is extremely rare (for example, IDU) you will need a large enough sample in order to capture some of this.

4. Practical Considerations

4.1 Cost

Among the eight examples presented at the consultation the cost varied widely. This was expected as the survey methods and local costs were very different. In general, cost will vary depending on the sample size, the size of the geographic area, survey mode, technical assistance and salaried staff required and the number of languages. For data sharing, there is also the additional cost to prepare, process, and produce or upload the data. However, it is important to remember that these methods allow the estimation of more than one population at the same time. A large additional cost will be the implementation of surveys to collect information on transmission error from the key populations.

4.2 DHS perspective

The Rwanda application was a stand-alone survey (pilot, testing several methods and approaches), but also tested the feasibility of integrating the survey into the DHS. Adding NSU to a DHS may be cost-effective for size estimation compared to a stand-alone survey; however, it will increase the local cost to the DHS and the amount of technical assistance would increase. Thus the question arises – *Would the additional cost be justified if the results are imprecise?*

Elements to consider:

- Sample (household only)
- Additional questions (complexity)
- Translation
- Preparation of the questionnaires
- Need for external data for adjustments

Additionally, it is worth nothing that if you are working in a country with great diversity it will not be as simple as Rwanda. Translation will be much more complicated if there are many languages and defining culturally relevant populations and names may be much more difficult in other settings. Note that if the method were integrated into the DHS, the objective would be to provide reliable population size estimates. It may not be necessary to add NSUM into the DHS in countries where the key populations are very small and clustered in a capital city. Overall, it was feasible to conduct NSU and proxy respondent methods for size estimation in Rwanda and there were no instances where respondent became upset or stopped the interview.

4.3 Practical consideration for country applications

We need to be mindful about when and where we recommend NSU – *What is needed to be able to plan programmes and resources?* We also need to keep the perspective of the funder in mind. For countries that currently do not have any size estimates at all, this might be a good option to provide some information. If there are already reasonable size estimates or regional estimates then the additional cost might not be fully justified. However, it might be possible to easily add NSU onto other surveys. If you are able to get a direct estimate, then use this approach instead.

5. Key questions and issues identified

A number of issues were raised during the presentations and discussions. The issues were primarily related to potential biases and can be summarised by identifying where in the process of collecting the information the error or bias occurs. NSU collects information on alters that is provided by a respondent to an interviewer. Information is passed from the alter to the respondent and then from the respondent to the interviewer and thus there are two stages where error can occur. Errors related to recall bias, definitional issues, and response bias occur in the first stage. Errors related to information transmission occur in the second stage. Errors at stage one are best addressed through the design of the survey while errors at stage two can only be adjusted during the analysis (through post-survey adjustment).

Methods and approaches

What size estimate methods and approaches are best?

- NSUM vs multiplier/capture-recapture/RDS/etc.
- Summation vs known population approach

How do we know if this method is working?

- What estimates do we believe?
- What numbers do we use?

What advice and recommendations can be given for:

- How to select known population groups?
- How to obtain mutually exclusive (culturally relevant) categories for the summation approach?

Definitions

What should we use for definitions:

- Of key populations of interest?
- Of what it means "to know" someone?

How can we maintain the integrity of the definition?

What are the limitations of our questions (Thailand example of using "gay" vs MSM)

What age-groups should be used?

Personal network size

Does PNS change over time or is it stable?

Once PNS is measured in a location, is this something that we can assume remains constant?

Adjustments

What advice can be given for adjusting crude estimates?

A number of these biases are highly correlated. How can we assure that we are not over-adjusting our estimate?

Considering transmission effects, barrier effects & reporting bias:

- How to incorporate and adjust for the effect of stigma?

  - Effect on transmission rate
  - Effect on personal network size of key population
- What is the ratio of the personal network size of the general population and the key population?
- How do stigma/social desirability relate to not telling the truth and the transmission rate?

Practical considerations

What advice can be given for constructing the sampling frame in order to avoid systematic exclusion or an incomplete sampling frame?

How to derive sub-national estimates?

How to capture uncertainty?

What group of people will be the on-call experts?

Survey mode: Is it possible to conduct NSU surveys using cheaper survey options such as internet based or cell phone based interviews. How will the different mode affect the reliability of the results?

Table I. Research topics and key issues identified

Research Topic	Issue
Meta analysis of study results	Creation of dataset with lines for each study results
Validation of NSUM	NSUM (PRM) with "traditional" methods (side by side hidden or known populations)
DESIGN ISSUES	
Summation vs Known	Relative strengths; when known gives larger estimates, how summation categories affect estimates statistically, control study that targets overlap
Summation questions	Why do some categories yield higher results? How do the categories work and affect the result? How can we find out more about the overlap? Do the listing exercise with multiple groups across different areas in a country
	Cognitive testing on how respondents interpret summation categories
Sub-national Estimates (UNAIDS priority)	Questionnaire design, sample size, survey experiment to assess estimate by province vs. national
Errors in who you interview	Sampling errors
Errors in inference from alters to respondents	Bias in selection of alters, for example if alters are people you shared a meal with in the last year
Anthropological Methods	(1) Integrate existing ethnographic research in specific settings, risk groups. (2) Ethographic study of participants' attitudes toward size estimation. (3) Ethnographic research on outliers (e.g., inconsistent reporting in PRM). (4) Alias of names (Happy, Joy, Comfort).
	Cognitive research on method of questionning, what people know about populations of known size
Proxy respondent method	Consider alternative methods that require fewer questions
POST SURVEY ADJUSTMENT	
Variation of Network Size	290…, differences in NS between target popn and general popn resulting in underestimates in NSUM (degree ratio)
Errors in what they know	transmission bias
	Possible alternative methods of measuring transmission error
Transmission error and degree ratio	How do these vary over key populations? And different settings? Are findings from Brazil where heavy drug users had smaller PNS generalisable?
BOTH DESIGN AND POST-SURVEY ADJUSTMENT	
Errors in what they tell you (unintentional)	barrier effect, recall bias
Errors in what they tell you (intentional)	stigma
Methods for Migrant Populations	Investigate methods from other fields. Relates to ethnographic research. US study on mobile drug users. NSUM does not directly give estimates to account for mobility. Perhaps some extensions could be developed to allow this.

6. Recommendations to countries

Network scale-up method

Countries using the NSU method should work together with those who have previously used the method and should strive to improve the method by testing one of the issues raised in the *key questions and issues identified* section of this report. The most recent update on the issues tables should be found on a networks scale up website.

Recommendations:

- **Compare results with size estimates obtained using other methods (solicit stakeholders and expert opinion too)**
- **Establish an evidence base that can lead to formal recommendations on the method**
- **Partner with other projects (add to other size estimation, other research projects)**
- **Generate national estimates, stratified by sub-national when possible**

Personal network size

What is the best approach to estimate PNS – summation versus know population?

Results from Brazil and Rwanda illustrated that while both methods worked, the known population approach had less error.

Recommendations:

- **Use both approaches and compare the results**
- **If you can only do one approach, use the known population approach in the first instance; if this approach is not possible, use summation**

Adjustments

More information and research is needed for formal recommendations which will come with more applications and additional insight.

Recommendation: Always report crude and adjusted estimates and define the adjustments made

Proxy respondent method

Further research is needed for this relatively new method in order to gain additional insight.

Recommendation: Use this method with a population that you know you estimate poorly but that you can count for comparison.

Qualitative component

Routinely build qualitative work in to gain more in-depth knowledge of what is being captured with the method and to learn more about incongruent findings. In addition, qualitative research can help us fully understand the programmatic implications of the results.

Recommendation to routinely incorporate a qualitative component whenever possible.

Data availability and data sharing

Sharing data and data reports will help to identify patterns observed across the country studies and will allow the application of what has been learned for the design of new studies. All of the necessary documents, guidance, links and tools should be contained in one place to facilitate ease of use.

Recommendations:

- **Share the data available whenever possible**
- **Incorporate data sharing clauses into future protocols**
- **Data reports readily available and widely circulated for studies where the data cannot be shared**

- **Make all the documentation, data sources and links to additional information available in one place**
- **Create and maintain a single aggregated data source with all results from country studies**
- **When possible, share results with organizations that represent the key populations to get their input during the design of the survey and when interpreting the results.**

7. Conclusions and next steps

The consultation met the objectives of identifying a list of issues regarding the network scale-up method. The issues were not prioritised due to time constraints and the recognition that different partners will have different priorities. In addition, a number of useful recommendations were identified for countries implementing network scale-up which should allow these new surveys to build on lessons learned from previous applications of the method.

Country guidance

The UNAIDS/WHO Surveillance working group will release a document describing how to conduct a network scale-up survey based on the latest information. In summary it will include:

- Procedures for known population approach and summation
- Links to a generic protocol (questionnaire too)
- Guidance on a formative assessment to conduct before using the NSU method (countries where there is high stigma, dealing with illegal practices, etc, quality of data available)
- Guidance on definitions, how to handle problems that arise around definitions
- Intellectual resources (who can the countries go to for assistance)
- Guidance on how to collect these data at different levels -- sub-national/regional/city
- Frequency of such surveys

To completed by August 2012.

Journal Supplement

To increase comparison and allow for a broad audience to access the country results and experiences, the country studies should be published, preferably together as a journal supplement. This process will begin now and continue through 2012 with a completed product expected in 2013.

Follow-up actions

- University of Washington has agreed to compile the data available to make comparisons of what has been learned so far.
- University of Florida will continue to host the NSU website including the list of research topics.

Appendix I: List of Participants

Abu Abdul-Quader
CDC
Atlanta, USA

Apinum Aramrattana
Chiang Mai University
Chiang Mai, Thailand

Bernard Barrere
Measure DHS
Calverton, MD, USA

Noah Bartlett
USAID
Washington DC, USA

Drew Baughman
CDC
Atlanta, USA

Russ Bernard
University of Florida
Gainesville, FL, USA

Jenny Butler
UNFPA
New York, USA

Kelsey Case
Imperial College London
London, UK

Toshi Ezoe
UNAIDS
Geneva, Switzerland

Dennis Feehan
Princeton University
Princeton, NJ, USA

Txema Garcia-Calleja
World Health Organization
Geneva, Switzerland

Jim Harmon
University of Washington
Seattle, USA

Wolfgang Hladik
US CDC
Kampala, Uganda

Manop Kanoto
Khon Khan University
Khon Khan, Thailand

Wen Lin
US CDC
China

Rob Lyerla
OGAC
Washington DC, USA

Mary Mahy
UNAIDS
Geneva, Switzerland

Chris McCarty
University of Florida
Gainesville, FL, USA

Tyler McCormick
University of Washington
Seattle, USA

Patrick Ndimubanzi
US CDC
Kigali, Rwanda

Tonia Poteat
OGAC
Washington DC, USA

Adrian Raftery
University of Washington
Seattle, USA

Keith Sabin
World Health Organization
Hanoi, Vietnam

Matt Salganik
Princeton University
Princeton, NJ, USA

Tetiana Saliuk
HIV Alliance
Ukraine

Otilia Scutelniciuc
UNAIDS
Kazakhstan

Donna Stroup
Data for Solutions
Atlanta, USA

Guo Wei
National Center for AIDS & STD
China

Sharon Weir
Measure Evaluation
Chapel Hill, NC, USA

Tian Zheng
Columbia University
New York, USA

Lev Zohrbayn
UNAIDS
Moscow, Russia

Appendix II: Meeting Agenda

DAY 1 Wednesday, 28 March 2012 (Where we stand and examples)		Presenter
Day 1 Chair person		Abu Abdul-Quader
9:00 – 9:30	**Introductions, purpose of consultation, terminology**	Mary Mahy
9:30 – 10:00	**What have we learned about NSUM**	Russ Bernard
10:00 – 10:30	**What do we still need to know? Key questions for this meeting**	Matt Salganik
10:30 – 11:00	**TEA & COFFEE BREAK**	
11:00 – 11:30	**Ukraine**	Tetiana Saliuk
11:30 – 12:00	**Moldova / Kazakhstan**	Otilia Scutelniciuc
12:00 – 13:30	**LUNCH**	
13:30 – 14:00	**Japan**	Toshi Ezoe
14:00 – 14:30	**Thailand**	Apinun Arramrattana / Manop Kanato
14:30 – 15:00	**China**	Guo Wei / Wen Lin
15:00 – 15:30	**Brazil**	Dennis Feehan / Matt Salganik
15:30 – 16:00	**TEA & COFFEE BREAK**	
16:00 – 16:30	**Rwanda**	Patrick Ndimubanzi
16:30 – 17:30	**DISCUSSION**	

DAY 2 Thursday, 29 March (Alternatives and biases)		Presenter/Facilitator
Day 2 Chair person		Rob Lyerla
9:00 – 9:15	**Review of questions raised on day 1**	Kelsey Case
9:15 – 10:30	**Introduction to proxy respondent, results from Rwanda, Discussion**	Wolfgang Hladik
10:30 – 11:00	**TEA & COFFEE BREAK**	
11:00 – 11:30	**Estimating personal network size: a comparison**	Chris McCarty
11:30 – 12:00	**Refining scale-up to estimate network size**	Tyler McCormick
12:00 – 13:30	**LUNCH**	
13:30 – 14:00	**Some generalizations of the network scale-up method**	Dennis Feehan
14:00 – 14:30	**Alternative data to estimate information trans rate**	Sharon Weir

14:30 – 15:00	**Reporting bias due to stigma**	Tatyana / Otilia /Mary
15:00 – 15:30	**Variance estimation**	Matt Salganik
15:30 – 16:00	**TEA & COFFEE BREAK**	
16:00 – 16:30	**Other technical issues**	
16:30 – 17:30	**DISCUSSION: What studies are needed to resolve these issues**	Donna Stroup

DAY 3 Friday, 30 March (Logistics of using the methods)		**Presenter/Facilitator**
Day 3 Chair person		Noah Bartlett
9:00 – 9:15	**Review of previous day's work**	Kelsey Case
9:15 – 9:45	**Estimated costs of methods**	Mary Mahy
9:45 – 10:15	**Comments from DHS**	Bernard Barrere
10:15 – 10:30	**Other issues: data sharing, journal supplement**	Abu Abdul-Quader
10:30 – 11:00	**TEA & COFFEE BREAK**	
11:00 – 11:30	**Uses of these estimates considering biases**	Rob Lyerla
11:30 – 12:00	**DISCUSSION: Recommendations to countries**	Lev Zohrabyan
12:00 – 13:30	**LUNCH**	
13:30 – 16:00	**Wrap-up discussion and way forward**	Mary Mahy

References

1. McCarty C, Killworth PD, Bernard HR, Johnsen EC, Shelley GA. Comparing two methods for estimating network size. Human Organization. 2001;60(1):28-39.

2. Ezoe S, Morooka T, Noda T, Sabin ML, Koike S. Population Size Estimation of Men Who Have Sex with Men through the Network Scale-Up Method in Japan. PLoS ONE. 2012;7(1):e31184.

3. Salganik MJ, Fazito D, Bertoni N, Abdo AH, Mello MB, Bastos FI. Assessing network scale-up estimates for groups most at risk of HIV/AIDS: evidence from a multiple-method study of heavy drug users in Curitiba, Brazil. American journal of epidemiology. 2011 Nov 15;174(10):1190-6.

4. Salganik MJ, Mello MB, Abdo AH, Bertoni N, Fazito D, Bastos FI. The Game of Contacts: Estimating the Social Visibility of Groups. Social networks. 2011 Jan 1;33(1):70-8.

5. McCormick TH, Salganik MJ, Zheng T. How Many People Do You Know?: Efficiently Estimating Personal Network Size. J Am Stat Assoc. 2010 Mar;105(489):59-70.

6. Huang Y, Henderson GE, Pan S, Cohen MS. HIV/AIDS risk among brothel-based female sex workers in China: assessing the terms, content, and knowledge of sex work. Sexually transmitted diseases. 2004 Nov;31(11):695-700.