EPI ALERT

UNAIDS quarterly update on HIV epidemiology / 2Q 2010

Network scale-up: a promising method for national estimates of the sizes of populations at higher risk

Countries can strengthen their HIV responses with better information on the number of people who engage in behaviours that increase their risk for HIV. By estimating the number of people in populations at higher risk of HIV, a country can revise its strategic plans, resource programmes appropriately, improve modelling of its epidemic, and advocate for services for those populations.

UNAIDS
JOINT UNITED NATIONS PROGRAMME ON HIV/AIDS
UNHCR UNICEF WFP UNDP UNFPA UNODC ILO UNESCO WHO WORLD BANK

EPI ALERT

$$\hat{e}_j = \frac{\sum_i m_{ij}}{\sum_i \hat{c}_i} T$$

Limitations to current size-estimation methods

A number of methods exist for estimating the sizes of populations at higher risk, including mapping methods such as census or enumeration, capture–recapture and the multiplier method.[1] These methods have a number of limitations:[2,3] they require stigmatized populations who engage in what may be illegal behaviours to disclose these behaviours to survey interviewers or clinic staff; most studies are geographically limited to one city or even one neighbourhood and are therefore not nationally representative; and most methods collect data on one population at a time, requiring multiple studies for a full description of high-risk behaviours in the country. These limitations make it necessary to develop new and improved methods to measure the size of populations at higher risk.

An alternative method

Ideally we would ask respondents in a nationally representative survey directly about their behaviours. However, social stigma and the illicit nature of the behaviours that put people at higher risk of HIV make this challenging. These behaviours include unprotected anal sex, injecting drugs with unclean equipment, and having unprotected sex with a high number of partners. Alternatively, we could ask respondents in a nationally representative survey about the behaviours of their acquaintances, thus reducing the embarrassment or potential negative consequences of responding to questions about these behaviours. Since the interviewer does not know these acquaintances, a known individual's behaviours are not being disclosed. Each respondent's personal network of acquaintances contributes to the sample. A recently proposed method called network scale-up uses this concept to overcome some of the limitations to size-estimation methods.

Network scale-up was initially proposed after the Mexico City earthquake in 1985. An anthropologist attempted to get an estimate of the number killed in the earthquake by asking survey respondents about the people they knew who died as a result of the earthquake.[4] In the 1990s the method was refined and used in the United States of America to estimate the number of people who were homeless, raped or HIV-positive.[5] Recent efforts have adapted the method to estimate the sizes of populations at higher risk of HIV, which are often hidden or hard to reach.[6]

The general concept behind the method is that if a respondent knows 300 people, 2 of whom are sex workers, then we could estimate that 2 of 300 of the general population are sex workers. We can improve that estimate by calculating the average proportion of sex workers known by a respondent over many respondents.7 If we multiply the average proportion of sex workers by the population of the country, we can estimate the number of sex workers in the population. The underlying assumption of this method is that people's social networks are, on average, representative of the general population.

Equation 1 describes this relationship. If T is the total population in the country, c is the personal network of individual i, and m is the number of individuals in the population at higher risk that i knows in subpopulation j, then we can calculate the estimated population at higher risk size e as follows:

Equation 1

$$\hat{e}_j = \frac{\sum_i m_{ij}}{\sum_i \hat{c}_i} T$$

To collect consistent answers on who each respondent knows and whether they know sex workers, we need to have a standard definition of what it means to know someone. Previous studies have used a standard definition that defines someone as being in your personal network if:

- You know them and they know you by sight and name;
- You have had some contact with them in the past two years;
- You could contact them if you had to;
- They live in the area of reference (usually the country).

This definition can be modified as appropriate for different settings but must be consistent for each respondent to remove any ambiguity in the subpopulation boundaries.

There are three steps to estimating population sizes using network scale-up: (1) estimate the average size of personal networks in a region or country; (2) determine how many people in each personal network are in the population of interest; and (3) calculate the size of the population at higher risk.

Step 1

There are two approaches to measuring personal network size: the summation approach and the known population approach. Depending on which type of data are available in the country, implementers can choose between these two approaches, as described below.

Summation approach

This approach asks respondents directly how many people they know. Determining the number of people you know is challenging for most people but is made easier if done in manageable subtasks. By categorizing acquaintances (such as family members, work colleagues and neighbours), respondents can estimate the number of acquaintances they have in that category. In general, people can estimate about 20 people from memory without having to create a list, so categories should be narrow. The categories must be exhaustive and mutually exclusive so that all acquaintances can be counted but not counted twice. Finally, the categories must be culturally relevant, tailored for each country in which network scale-up will be used.

Known population method

Another approach to estimating the average personal network size uses the initial network scale-up equation in reverse. For example, by asking how many doctors a respondent knows and then dividing that number by the known number of doctors in the

Example of summation questions

How many people do you know in the following categories:

- Your immediate family
- Your birth family
- The family of your spouse/partner
- Co-workers
- Other people at work
- Close friends
- People known through hobbies/recreation
- People known through faith-based organizations
- People you know from your neighbourhood
- People known through others
- Childhood friends
- People who provide a service

$$\hat{c}_i = T \frac{\sum_j m_{ij}}{\sum_j e_j}$$

Examples of known population questions used in recent surveys

How many people do you know who are:

- Women over 70
- Men over 70
- Women named Christine
- Men named Victor
- People who were married in 2009
- Kindergarten teachers in 2009
- Graduates of higher education institutes in 2009
- Dentists in 2009
- Medical doctors practising in 2009
- Nurses in 2009
- Police officers in 2009
- Teachers in secondary schools and technical schools in 2008

country, we can determine the respondent's personal network size. Instead of estimating the population size e, we estimate a person's network size c. So rewriting Equation 1, we get Equation 2:

Equation 2

$$\hat{c}_i = T \frac{\sum_j m_{ij}}{\sum_j e_j}$$

The researchers need to identify populations for which valid and current statistics are available. About 20–30 such populations should be identified to make the estimated personal network size accurate. By asking the respondents how many people they know in each of these 20–30 populations, each individual's personal network size can be estimated. By averaging this value over numerous respondents, we can calculate a national average personal network size.

For example, we can estimate a personal network size using the known population of medical doctors:

1. Ask respondents how many medical doctors they know who were practising in 2009.
2. Calculate the average number of medical doctors reported as acquaintances (for example, 6).
3. Identify the number of medical doctors who were practising in 2009 in the country based on available statistics from the medical association, ministry of health, etc. (for example, 20 000).
4. Divide the average number of doctors known to the respondents by the total number of medical doctors in the country (m/e = 6/20 000).
5. Multiply this proportion by the total population (for example, 1 000 000 population, so 6/20 000 x 1 000 000 = 300 as the average network size).
6. If we do this for 20–30 known populations, we will improve the accuracy of each individual's estimated personal network size.

It is also possible to test the accuracy of the estimated personal network size. For example, we can test the personal network size derived through the question on how many women the respondents knew named Christine:

1. Using all of the remaining known populations (all except women named Christine), we estimate the average personal network size to be 300.
2. Data from the census inform us that there are 50 000 women named Christine in our country.
3. Based on our survey, each respondent knew on average 14.8 women named Christine.

4. If we divide that by the average network size (14.8/300), we find that approximately 4.93% of women in the country are named Christine.
5. Based on this estimate, we can assume that there are approximately 14.8/300 x 1 000 000 = 49 333 women named Christine in the country.
6. The true value of the personal network size is very similar to the value calculated using the remaining known populations, suggesting that the personal network size of 300 is reliable.

Researchers in the United States of America used the reverse known population method described above to compare the actual population size from available statistics with the predicted size using the remaining known populations. Fig. 1 shows the results from 29 categories of known population sizes. The results suggest that the results were quite similar, except for two outliers. The two outliers were the number of people with diabetes and the number of people who were twins. Further studies of what information people share with acquaintances suggests that health conditions such as diabetes are not always shared with acquaintances. In addition, less significant traits or visually apparent traits such as the status of being a twin are often not shared with acquaintances. This study led to recommendations on how best to choose known populations in order to reduce potential errors.[7]

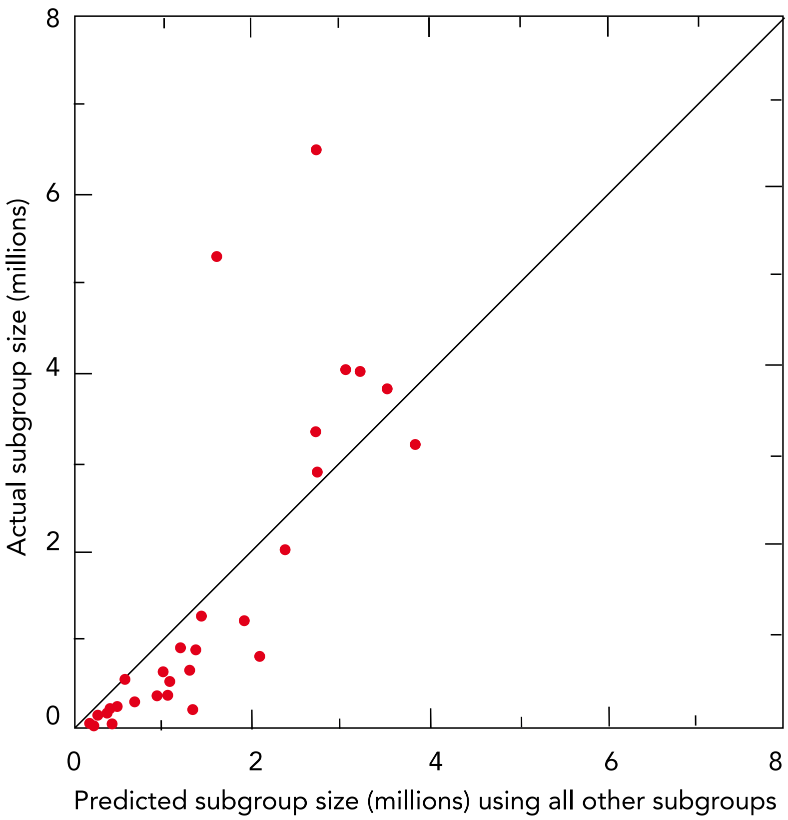
The two approaches for measuring personal network size were compared in a nationally representative survey in the United States of America. The authors found that the mean personal network size using the known population approach was 290.8 (standard deviation 264) and using the summation approach was 290.7 (standard deviation 259). The two methods had similar results. Although there is no gold standard against which to compare the two results, further analysis corroborated the estimates as useful proxies of personal network size.[8]

Step 2

The second step for network scale-up requires asking survey respondents how many people in their personal network are in the populations of interest—in this case, the populations at higher risk of HIV. For example: How many people do you know who inject drugs? How many people do you know who sell sex? How many men do you know who have sex with other men? How many people do you know who purchase sex? These questions must be modified for the specific country context and language.

Step 3

In the third step, the size of the population at higher risk is calculated and adjusted as necessary. This is done by determining the proportion in each population at higher risk and multiplying that by the total adult



population in the community. For example, if, on average, survey respondents knew 1.82 sex workers (m = 1.82) and we divide this average by the estimated average personal network size (c = 300) and multiply this proportion by the total population (t = 1 000 000), we estimate there are approximately 6067 sex workers (1.82/300 x 1 000 000).

The calculation of the adjustment factors based on potential sources of error is described below.

Limitations and biases to networks scale-up

As with all methods for size estimation, the assumptions required for network scale-up are critical. These assumptions include the following:

- Everyone knows everything about everyone they know.
- Everyone in the population has an equal chance of knowing someone in their network.
- People recall and report accurately the number of people they know in the subpopulations.

The violation of these assumptions can be described by three different types of bias. To the extent that these biases can be measured (see below), we can adjust for them.

Transmission error occurs when respondents do not know the behaviours of their acquaintances; for example, do sex workers tell all of their acquaintances about their work? It is possible to estimate transmission error by determining whether the population at higher risk shares their behaviour with their acquaintances. In a separate survey of the population at higher risk (such as a behavioural surveillance survey of sex workers), respondents are asked about the number of their acquaintances (as defined earlier—"who they know") who are aware of their sex work activities. An adjustment factor is calculated from their responses and the population size is corrected accordingly.

The barrier effect occurs when people in one geographical area or socioeconomic class do not have an equal probability of knowing people in another area or class. For example, are people in rural areas less likely than people in urban areas to know a person who injects drugs? The barrier effect is common in national surveys.[9] Most behaviours that put people at increased risk to HIV are clustered among geographical regions and demographic groups. The barrier effect can also affect the known population calculation of the personal network size. For example, men may know more police officer than do women. Even first names are associated with the barrier effect (for example, the names Joseph and Mohammed are common among people of certain faiths, making them specific to different regions of a country where those faiths are prevalent). The barrier effect can be avoided by using a random, nationally representative sample of the general population.

Recall and reporting bias occur when the respondent does not remember or does not report accurately on what they know. For example, respondents might be hesitant to report to the interviewer that they know injecting drug users, because of social desirability bias. Respondents systematically overrecall small known populations (for example, people who are physically handicapped) and underrecall large populations (for example, men who are farmers).

Colleagues in the Ukraine have developed a method for adjusting for the potential reporting bias when asking about stigmatizing behaviours. They created a measure of stigma that can be included when using the known population approach to measuring personal network size. For each known population, the respondent was asked: On a scale of 1–5, how much do you respect people in [known population; for example, medical doctors or people named Michael]? The average respect value was calculated for each known population and each population at higher risk. An adjustment factor was created reflecting the chance that someone will underreport their acquaintance with someone in a category with low respect.[10]

Next steps and other resources

The network scale-up method is still being developed, including the methods necessary for adjusting for different biases. As with all size-estimation methods, multiple size estimates should be calculated and compared to determine an estimate and uncertainty range for the population size.

Currently two small-scale studies are taking place in Latin America to design adjustments for transmission error. In addition, the full method will be pilot-tested in a national household survey in sub-Saharan Africa later in 2010. If successful, a thorough guidance document will be created on how to carry out a network scale-up survey and the adjustments that need to be considered.

Other size-estimation methods based on the concepts of network scale-up are also being considered and tested. One method, referred to as proxy respondent, would exclude the need to estimate the personal network size by asking about one randomly selected person in the respondent's network. This could be useful in very large household surveys (W. Hladik, unpublished data, 2010).

In addition, more work is being done by statisticians to understand the sample size required given the very low probability that a respondent will know people with high-risk behaviours. It is currently possible to estimate the confidence bound on the personal network size using known populations, but a full confidence range is not possible for the resulting size estimate.

The developers of the network scale-up method have launched a web site that includes a number of resources: http://nersp.osg.ufl.edu/~ufruss/scale-up.htm. This site contains a list of papers and presentations related to the method.

Contact: Mary Mahy, Epidemic monitoring and Analysis, UNAIDS

References

1 *Estimating the size of populations at risk for HIV: issues and methods.* UNAIDS, FHI, Impact, USAID, WHO, UNDCP, 2003 (http://data.unaids.org/pub/Manual/2003/20030701_gs_estpopulationsize_en.pdf, accessed 22 February 2010).

2 Zhang D et al. Advantages and challenges of using census and multiplier methods to estimate the number of female sex workers in a Chinese city. *AIDS Care*, 2007, 19:17–19.

3 Hook EB, Regal RR. Capture–recapture methods in epidemiology: methods and limitations. *Epidemiology Review*, 1995, 17:243–264. (Corrigenda in American Journal of Epidemiology, 1998, 148:1219.)

4 Bernard HR et al. Estimating the size of an average personal network and of an event subpopulation: some empirical results. *Social Science Research*, 1991, 20:109–121.

5 Killworth PD et al. A social network approach to estimating seroprevalence in the United States. *Social Networks*, 1998, 20:23–50.

6 Stroup et al. Forthcoming.

7 Killworth PD et al. Estimation of seroprevalence, rape, and homelessness in the United States using a social network approach. *Evaluation Review*, 1998, 22:289–308.

8 McCarty C et al. Comparing two methods for estimating network size. *Human Organization*, 2001, 60:28–39.

9 McCormick TH et al. How many people do you know? Efficiently estimating personal network size. *Journal of the American Statistical Association*, 2010, 105:59–70.

10 Paniotto V, Petrenko T, Kupriyanov O. *Estimating the size of populations with high risk for HIV using the network scale-up method.* Kiev, Ukraine, Kiev International Institute of Sociology, 2009.

EPI NEWS

Conference proceedings: New strategies and methods for HIV/AIDS surveillance in low and middle income countries - 2009

On 2-5 March 2009, the IInd global HIV/AIDS surveillance meeting was held in Bangkok, Thailand, supported by the Office of the Global AIDS Coordinator (OGAC) and several specialized USG agencies, the World Health Organization (WHO) and the Joint United Nations Programme on HIV/AIDS (UNAIDS). The meeting brought together approximately 300 surveillance practitioners and experts to review the current state of HIV-related surveillance in low and middle income countries, and to assess advances in surveillance methodologies and implementation since the 2004 Addis Ababa meeting.

The main strategies and methodologies emerging from presentations and discussions at the Bangkok conference were documented in a peer-reviewed journal special issue. This journal special issue aims to advance the science of HIV surveillance by highlighting recent developments in selected areas in HIV surveillance, building on presentations and discussions at the Bangkok meeting. The papers in this special issue cover a range of topics, ranging from surveillance and analytic methods for HIV incidence to surveillance of HIV-related mortality, and from surveillance methods per se to the evaluation of surveillance systems and to the use of surveillance data.

List of papers included in the journal:

- New strategies and methods for HIV surveillance in low and middle income countries Peter D Ghys, Theresa Diaz, Keith Sabin
- Update for sampling most-at-risk and hidden populations for HIV biological and behavioral surveillance Lisa Johnston, Keith Sabin, Dimitri Prybylski
- Applying current methods in size estimation for high risk groups in the context of concentrated epidemics: Lessons learned Tobi Saidel, Virginia Loo, Tetyana Salyuk, Faran Emmanuel , Guy Morineau, Rob Lyerla
- HIV prevalence surveillance among HIV prevalence surveillance among TB patients in resource-constrained countries: a look forward Abraham G Miranda, Philippe Glaziou, Eleanor Gouws, Christian Gunneberg, Irum Zaidi
- Can PMTCT program data replace HIV sentinel surveillance among pregnant women in Africa? Kimberly Anne Marsh, Omotayo Bolu, Stephane Bodika, Khumo Seipone, Suriya Wonkongkathep, Fulgentius Baryarama, Aisha Yansaneh, Chineta Eure-Miller, Jesus M Garcia-Calleja
- Filling a gap: HIV pediatric surveillance Sadhna Patel, Olive Shisana, Evelyn Kim, Avi Hakim, Margaret Brewinski, Dmitry Kissin, Lauren Zapata, Chris Murrill, Okechukwu Nwanyanwu, Thomas Rehle, Mary Lou Lindegren
- Estimating HIV Incidence in Populations Using Tests for Recent Infection: Issues, Challenges and the Way Forward Timothy D Mastro, Andrea A Kim, Timothy Hallett, Thomas Rehle, Alex Welte, Oliver Laeyendecker, Tom Oluoch, Jesus M Garcia-Calleja
- A Proxy Measure for HIV Incidence among Populations at Increased Risk to HIV Mary Mahy, Chhorvann Chhea, Tetiana Saliuk, Olga Varetska, Rob Lyerla
- Mortality surveillance practices for measuring the demographic impact of HIV/AIDS in sub-Saharan Africa Alison L Smith, Angela Smith, Tekebash Araya Gebrekidan, Elisio Mazive, Shebo Nalishebo, Palver Sikanyiti, Loraine West, Keith Sabin
- Improving HIV surveillance systems: Country experiences and a proposal for evaluation framework Virginia Loo, Evelyn Kim, Edgar Monterroso, Thuy Nguyen, Tobi Saidel, Neha Shah, DCS Reddy, Jesus Garcia Calleja
- Maximizing the use of surveillance data: Innovative examples from around the World Mary Mahy, Tasnim Azim, Lolita Ganina, Kayode Ogungbemi, Ruton Hinda, Aliya Bokazhanova

This special issue "Vol 2, No 1 (2010) Special Issue Theme: New Strategies and Methods for HIV/AIDS Surveillance in Low and Middle Income Countries – 2009"[11] is published in the "Journal of HIV/AIDS Surveillance & Epidemiology"[12] a new online journal specialized in HIV surveillance and epidemiology. We hope that this special issue, building on the 2009 IInd global HIV/AIDS surveillance meeting in Bangkok, contributes to increased exchanges and learning about the science and practice of HIV surveillance around the world.

11 jHASE 2010, 2(1):1
12 http://www.ieph.org/hase/jHASE.htm

UNAIDS – 20 AVENUE APPIA – CH-1211 GENEVA 27 – SWITZERLAND – Tel.: (+41) 22 791 36 66 – Fax: (+41) 22 791 48 35