UNAIDS/03.36E
Dist: general
Original: English

UNAIDS/WHO
Working Group on HIV/AIDS/STI Surveillance

Estimating the size of populations at risk for HIV

Issues and Methods

updated July 2003

Family Health International
Impact Implementing AIDS Prevention and Care Project
USAID
Joint United Nations Programme on HIV/AIDS UNAIDS UNICEF • UNDP • UNFPA • UNDCP UNESCO • WHO • WORLD BANK
World Health Organization
UNDCP United Nations International Drug Control Programme

UNAIDS/03.36E
Dist: general
Original: English

Estimating the Size of Populations at Risk for HIV

Issues and Methods

updated July 2003

A joint UNAIDS/IMPACT/FHI workshop: Report and Conclusions
Report prepared by Elizabeth Pisani, May 2002, Updated July 2003
Thanks to Sharon Weir, Basia Zaba and Gordon Hay for substantial input into the updated version.

Funded by the United States Agency for International Development (USAID) through FHI's Implementing AIDS Prevention and Care (IMPACT) Project, Cooperative Agreement HRN-A-00-97-00017-00.

The contents of this report do not necessarily reflect the views of FHI or USAID.

ISBN 974-91495-0-5

Table of contents

1 INTRODUCTION

As HIV moves into its third decade, it continues to grab the headlines. The devastating impact of the virus on development has shaken us awake to the need to do more to prevent its spread. Many lessons have been learned about effective prevention but many countries still have difficulty focusing their efforts on the interventions that will have the greatest impact in stemming the spread of the virus.

The difficulty in focusing prevention efforts appropriately is especially pronounced in countries where HIV remains largely concentrated among sub-populations whose behaviour puts them at high risk for contracting and transmitting HIV. Many governments find it politically difficult to invest in services for injecting drug users, men who have sex men and female sex workers and the clients they serve. And yet in most countries in the world outside of sub-Saharan Africa, these sub-populations remain among the most important focal points for effective HIV prevention.

Much of Africa has long-established HIV surveillance systems tracking infection in pregnant women. In Asia, Latin America and Eastern Europe, where HIV is more heavily concentrated in sub-populations with defined risk behaviours, more and more governments are now investing in surveillance systems that track the distribution of the virus and the behaviours that spread it in the hardest-hit populations. However even the best existing surveillance systems have one central weakness. While they can measure the level of risk behaviour, HIV and STI infection in a given sub-population, they cannot give any indication of the absolute size of the sub-population itself.

Recognising this as a limitation, many countries have begun to attempt to estimate the size of populations at high risk for HIV. No standardised methodologies are currently available to guide this process, and many different approaches have been tried. In an effort to synthesize the "state of the art" for population size estimation in one region where HIV is concentrated predominantly in groups with high risk behaviours, the Joint United Nations Programme on HIV/AIDS (UNAIDS) and the United States Agency for International Development (USAID) sponsored a workshop on estimation methods in Asia. The workshop, co-organised by UNAIDS and the USAID-funded Implementing AIDS Prevention and Care Project IMPACT (implemented by Family Health International), was held in Bali, Indonesia, in April 2002. Participants from Bangladesh, Cambodia, China, India, Indonesia, Lao PDR, Myanmar, Nepal and Vietnam attended the workshop, together with participants from international organisations and NGOs including FHI, Population Services International (PSI), UNAIDS, the United Nations Drug Control Programme (UNDCP), USAID, the United States Centers for Disease Control and Prevention (CDC) and the World Health Organisation (WHO). Health Canada also contributed substantially to the workshop presentations.

Much of this document is derived from that workshop. An earlier version of the document was circulated and used in a number of countries, generating debate, commentary and experimentation. This updated version of the document includes material that has surfaced since the original workshop was held. It covers the major methods available for population size estimation, with their strengths and weaknesses, and gives examples. It explores how best to choose the right method for a given country situation and sub-population. The document also provides modified copies of the exercises used during the Asian workshop. It is not intended as a comprehensive guide to population size estimation: more thorough tool kits will be needed and for some sub-populations may already be available.

Many countries have already begun the work of developing estimates of the size of populations with specific risk behaviours. The information in this document is intended to encourage that work, and to provide enough information to stimulate countries to broaden their experimentation with population size estimation.

WHY DO WE NEED TO ESTIMATE THE SIZE OF POPULATIONS AT RISK FOR HIV?

Many countries are in the process of building up robust surveillance systems, tracking the spread of HIV and of other sexually transmitted infections (STIs), as well as of the sexual and injecting behaviours that carry a risk of transmitting HIV. While much remains to be done in strengthening surveillance, many countries are now able to estimate prevalence of HIV and risk behaviours in certain groups with some confidence. They may be able to say, for example, that 36 percent of drug injectors are infected with HIV. They can also report that 60 percent of injectors regularly share needles, and by combining those two pieces of information can conclude that it is highly likely that HIV will continue to spread through the injecting population. However, only few countries have any reliable idea how many people inject drugs in the country. That means it isn't possible to estimate either how many people are already infected through injecting (and so how many injectors may be in need of HIV-related care services), or how many more may be at risk unless effective prevention is introduced.

Broadly, reasons for developing reliable estimates of the size of populations at high risk for HIV can be grouped into two major areas: policy and programming. The area of policy encompasses advocacy, response planning and resource allocation, plus estimations of numbers infected with HIV and projections of the burden of disease. The area of programming encompasses intervention planning, measurement of coverage, and monitoring and evaluation of interventions.

Different methods may be appropriate for different goals. For policy purposes, national estimates are frequently needed, where as local estimates will often do the job if the goal is to design or monitor specific prevention or care programmes.

Population size estimation for better policy

Lobbying for appropriate interventions

The very first use of population size estimation may be to alert policy-makers to the existence and the magnitude of a sub-population that may be at risk for HIV. Where no estimate is available to gauge the size of a problem, or where the basis for estimates is not clear, so that the results can easily be dismissed, it is easy for policy-makers to brush an issue aside. “Drug injectors? Oh, that’s just a handful of street kids, we don’t need to anything for them.” Good estimates of at-risk sub-populations based on sound, well-recorded methods make a potential HIV problem “real”: it becomes more imperative to address the problem with appropriate prevention and care interventions.

This advocacy role may be especially important in decentralised government systems such as those in China, India and Indonesia. Local governments rarely act on national-level information. They are driven, rather, by local votes and local concerns. If it can be demonstrated that significant sub-populations at risk for HIV exist in their own region, action is more likely to follow.

The need for local estimates does not, of course, replace the need for national estimates. Ideally, local estimation exercises could be aggregated to contribute to a national estimate. However this is only possible if certain standards (for example a common definition of the at-risk population) are observed. Currently, because there is so little standardisation in this area, this has not been much considered. In some large countries, size estimation exercises are taking place at the local level with no central coordination, limiting the ability to arrive at robust national estimates. As this area of surveillance develops, the central surveillance authority may want to make recommendations for standard definitions and methods so that local estimates can more easily be used effectively at a national level.

Raising money and allocating it efficiently

Once the extent of the potential problem has been made explicit through sound estimates of the number of people potentially affected, it is far easier to argue persuasively for funding for prevention and care interventions. High HIV prevalence among transvestite sex workers is unlikely, by itself, to move the finance ministry or external donors to provide substantial funding for interventions in this group. They are marginalised, and often not highly visible to policy-makers, who may assume that this population numbers no more than a few hundred people. But if the population of transvestites has been reliably estimated at between 12,500 and 14,700 nation wide, it is far easier to make a case for funding.

Knowing the size of the population at risk means you can calculate the coverage of services provided for them. Demonstrating that only one drug injector in 20 has access to harm reduction programmes — or that 80,000 out of 100,000 drug injectors in the capital city have no access to these services — can be another powerful weapon in fighting for resources appropriate to the scale of the problem.

So population size estimates can help pull in new resources. But good programme planning at the national level also requires that available resources are allocated efficiently. Prevalence figures alone may distort this allocation. For example, surveillance systems may show that HIV prevalence among male sex workers, at 22 percent, is twice as high as the 11 percent recorded among female sex workers. At first glance, this may suggest that twice as much money should be dedicated to programmes for male sex workers as for female sex workers. In fact, such decisions should not be made without good estimates of the relative size of the populations, as well as information about their levels of risk behaviour. At those prevalence rates, if there are 6,000 male sex workers in the country and 47,000 female sex workers, then there will be 5,170 female sex workers and 1,320 male sex workers infected with HIV. In other words, nearly four times as many people in the “less infected” group are living with HIV. If both male and female sex workers have the same average numbers of clients each, then it probably makes sense to dedicate more resources to prevention efforts in commercial sex between men and women than between men and men. Population size information is therefore an important ingredient in deciding how resources should be allocated between groups.

Estimating how many people are living with HIV nation-wide

Perhaps the most politically sensitive of the uses of estimates of the size of at-risk populations is in determining the number of people infected with HIV in a country. In countries where HIV has spread throughout the general population, this is relatively easy. HIV prevalence rates measured in pregnant women are applied, with a few standard adjustments, to the whole population of sexually active age to arrive at a national figure.

In countries outside of sub-Saharan Africa, this is rarely appropriate. Where HIV is concentrated in specific sub-populations, surveillance systems should concentrate on those populations. (Measuring infection in pregnant women may miss the epidemic completely. It will not, for example, record an epidemic driven largely by sex between men, or by needle sharing among predominantly male drug users.) In many parts of Asia, and in some parts of Latin America and Eastern Europe, we already have a good idea of what proportion of major sub-populations are infected with HIV. But to arrive at a national estimate of infection, we have to be able to apply that proportion to a denominator: the absolute number of people engaging in that behaviour. In other words, we need good estimates of the size of the various sub-populations. Without such estimates, important populations are sometimes ignored completely or severely understated when calculating how many people in a country are living with HIV. On the other side of the coin, lack of reliable estimates leaves the door open for wild guesses and hyperbole from organisations interested in maximising the resources allocated to populations with which they work.

National estimates of infection are important for planning appropriate care and prevention services, and, or course, for budgeting for those services. But they are important too in a global context. Many organisations maintain global data-bases on HIV, drug use and other related risks. And many countries have undertaken to report information regularly to these organisations for inclusion in their databases. While meeting global reporting requirements should never be the driving force for any kind of data collection, there is a principle of “enlightened self-interest” at stake. If countries submit estimates of HIV infection that are based on sound surveillance systems and well-documented, credible estimates of the size of various at-risk populations, these estimates will be included in the databases and made available globally to donors and others who may be considering allocating resources in the field of HIV. This may increase international contributions to a national response, and avoid international institutions coming up with their own estimates, independently of national procedures.

Measuring and projecting trends over time

HIV epidemics are not static, in part because the behaviours that drive them are subject to change. Some of this change may be in response to HIV prevention programmes, some may not.

Current HIV prevalence can be used to gether with current estimates of the size of various sub-populations to arrive at an estimate of the existing burden of HIV, as described above. However these pieces of information can also be used together with behavioural information to make projections about the likely course of the epidemic in the absence of effective prevention. Say there are 29,000 female sex workers nationwide. In behavioural surveillance, 12 percent of them report consistently using condoms. That means that some 25,520 sex workers are exposing themselves to HIV. Even if HIV prevalence is currently low, the potential for a very significant future epidemic in this country clearly exists. Decision-makers would do well to encourage effective prevention to raise condom use before the virus takes a hold.

Trends in the size of the at-risk population itself can also help inform policy at a national level. Some countries have, for example, recorded a recent rapid rise in the proportion of drug users reporting the use of non-injected amphetamine type substances. If this rise means drug users are switching from injectible to non-injectible drugs, the implications may be a reduced emphasis on IDUs as a key population for HIV prevention and care. If, however, the number of IDUs is constant and the rising proportion of drug users not injecting reflects not a switch but a rise in the overall number of people taking drugs, then the policy implications are different. Robust estimates of the size of drug-taking populations, together with information about their behaviour, can provide this information.

Population size estimation for better programme planning and management

Countries have to know the relative size of different at-risk populations so that they can plan their overall response and allocate resources efficiently. But knowing the size of specific populations at a local level is also key to good programme planning, implementation and management.

Assessing needs and planning to meet them

It is extremely difficult to plan to provide adequate services for a particular sub-population if you don't know how many members of that sub-population there are. How many STI screening kits are needed in order to provide regular screening and treatment for all sex workers in a city? How many needles must be supplied if we want to make clean needles available to three quarters of the city's drug injectors? How many outreach workers do we need to train if we want to contact 80 percent of gay men at least twice a month, and each outreach worker can contact 30 gay men a week? The answer to all of these questions, and many others which stand at the core of planning, budgeting and evaluating HIV prevention and care interventions, depends on knowing how many individuals there are in a given sub-population.

Evaluating our efforts: the question of coverage

In recent years, more resources have become available to help prevent HIV and more is known about what works in helping people to reduce particularly risky behaviours. Those changes, combined with a growing realisation that impact on the epidemic as a whole depends on changing the behaviours of a high proportion of those most at risk, have led to a shift away from "pilot projects" and "boutique interventions" and towards prevention programming on a larger scale. Cheaper, more user-friendly treatments have also increased the demand for widespread provision of care to those affected. More and more, the talk is of "scaling up" and of "achieving coverage". And of course, reaching coverage targets depends on knowing how many people are in need of services.

Again, there is an international sub-text here. In many developing countries, a significant proportion of the funding for HIV-related programmes comes from foreign donors or lenders. Many of them have, together with developing country governments, signed up to international targets for HIV prevention, care and service provision such as those associated with the United Nations Special Session on AIDS (UNGASS) declaration of 2001. And they expect to be able to measure progress towards those targets. Countries able to report substantial progress are likely to benefit from stronger international support. And reporting progress in "scaling up" prevention and care interventions once again depends on knowing the size of the population in need of services.

Population size estimation is not the answer to all our problems

Laid out above are several excellent reasons for increasing efforts to estimate the size of populations at risk for HIV. But they come with a caveat. Population size estimation alone does not solve the problems of HIV planning and programming. Estimates of the size of at-risk populations are meant to be used in conjunction with other sources of information. Many of the purposes given above, especially those surrounding advocacy and resource allocation, can only be fulfilled by combining population estimates with information on HIV prevalence and risk behaviour generated by the surveillance system. The uses of population size estimation associated with measuring coverage depend on good record-keeping for service delivery. And that in turn depends on there being services in place.

In short, the caveat is this: there is little point investing resources in repeatedly estimating the size of at-risk populations unless there is a commitment to provide services for those populations. While many of the methods described below rely on the use of existing information, some also require the collection of primary data. These data collection exercises can be costly and time-consuming. There is a risk that institutions may allocate resources to such exercises as a substitute for core HIV prevention and care efforts rather than as a complement to those efforts, and this should be strongly discouraged.

It is worth bearing in mind some other limitations. HIV and behavioural surveillance systems and at-risk population estimates are likely to be controlled by public health authorities who assume that the information will lead to better HIV prevention and care policies. Other authorities may have different view-points. There is a very real possibility that reliable estimates of the numbers of drug injectors, street-based sex workers, gay men etc., will lead not to a public health response but to a law-enforcement response. The likelihood of this happening of course depends on the national situation, but it should be carefully considered when undertaking — and above all when publicising the results of — population size estimation efforts.

A law-enforcement response that harms the very people we wish to provide services for is the worst case scenario. But there are other possible outcomes. One is that authorities will simply ignore the results of a rigorous and transparent estimation and remain wed to less robust estimates that better suit their political agenda. Public health officials must weigh up the costs of making credible estimates against the likelihood that these estimates will be used constructively.

A final caveat is that population size estimation methods are designed only to arrive at estimated counts of groups at high risk for HIV. Many methods are based on mathematical calculations, and require no contact with individuals. These methods are certainly not designed to identify individuals who may be members of groups at high risk for HIV, or to facilitate access to those groups for programming. The difference between counting people and reaching people must be made abundantly clear at the start of any population size estimation exercise, to avoid creating expectations which cannot be met.

3 GENERAL ISSUES IN POPULATION SIZE ESTIMATION

There are some overall issues which need to be considered in planning population size estimation. These are common to all groups, and to all methodologies. They will be discussed here in general terms. More detailed information about the choice of methods appropriate to different sub-populations and different country settings is given in Chapter 5 on page 33.

Defining the population

The first and often most difficult issue is in defining the population whose size one wishes to estimate. In public health terms, we are really most concerned with people whose behaviour puts them at risk for HIV. That may well not be all members of a defined population: there may be significant sub-populations within that group at no risk at all, and in no need of services. For example, we tend to think of injecting drug users as a group at very high risk for HIV. However, in some countries, physicians and health care workers are among the ranks of injecting drug users. Because they have easy access to sterile injecting equipment and a strong professional motivation to conceal their drug use, they are neither at high risk for HIV infection nor likely to come into contact with any of the locations or data sources that may be used for size estimation. This means they will be excluded from most estimates of drug injecting populations, but since they are not at high risk for HIV, this is not a major concern for HIV programme planners. From an HIV prevention point of view, it would not be worth making a special effort to capture this sub-population.

In other cases, however, it will be important to make estimates separately for different sub-populations within a group that shares a common risk behaviour for HIV, because the various sub-populations may require different HIV prevention and care services. For example, unprotected anal sex between males carries a high risk of HIV transmission, regardless of whether these males identify themselves as transsexuals, transvestites (or even as women), as gay men, or as heterosexual men simply wanting a sexual “change of scene”. Clearly, however, HIV prevention services will have to differ between these groups if they are to be effective. If population size estimations are being undertaken for programme planning and evaluation purposes, they may have to be carried out separately for the different sub-populations involved.

A question of geography

Many behaviours that carry a high risk for HIV are very situation-specific. Different cities may have very different patterns of drug use, with heroin injection being the norm in one city while inhalation dominates in another city. The ethnic majority in the central plains of a country may have a tradition of male sex workers attending temple festivals, while no such tradition exists among the ethnic minorities concentrated in the northern hills. This means it is rarely possible to generalise estimates from one city or region to the rest of the country. National estimation efforts may have to stratify between areas of high, medium and low prevalence of risk behaviours, or may have to aggregate separate estimates from key provinces.

Aggregation is not always as easy as it sounds. One factor that throws off estimates based on the sum of separate regional estimates is migration. Because local level estimates are usually more connected with programme planning and evaluation than with policy, they tend to focus on the total number of people needing services over a given planning cycle, such as a year. If people only sell sex in a city for six months on average before moving to another city where they are considered “new stock” and command higher prices, the annual total will be twice as high as the total at any one time. More complicated still are sex workers who follow the market on a weekly basis, working in the capital on weekdays but flying to resort islands to serve holidaying clients from neighbouring countries at weekends. In other words, national estimates based on a sum of key city estimates risk counting the same women more than once, as they move from city to city. An understanding of contract systems in the sex industry and other predictors of mobility will be necessary to make good estimates in such situations.

Changes over time

Behaviours that carry a risk for HIV are, like all behaviours, subject to change over time. Fashion, the economy, changes in drug routes or tourism patterns, seasonal migration, politics - all can affect the prevalence of risk behaviours and, therefore, the size of the population. This can happen at a local level — an upcoming election for mayor in which law and order is a key issue may lead to high profile police action against transvestite sex workers, and many transvestites may shift to other cities or stop selling sex for a while, upsetting census and enumeration based estimates for this population. Or it can happen at a national level, for example with the introduction of user fees at public STI clinics including those in red light districts, which may have an effect on clinic attendance and therefore record-based estimates of the number of female sex workers.

Accessing populations at high risk for HIV

As has been mentioned, estimation exercises are not intended to provide access to populations for interventions, and indeed some methods require no contact with the population itself. Other methods do, however, require accessing populations at risk for HIV, and this is not always straightforward.

Populations at high risk for HIV are often referred to as “hidden populations” or “hard-to-reach populations”, with mixed justification. Brothel based sex workers, for example, are neither hidden nor hard to reach. Middle class drug injectors are usually both. Regular government social service or health workers can conduct enumerations of brothel-based sex workers. That would be almost impossible in the case of drug injectors, where non-government organisations that provide services to the populations have a much better chance of securing some access. In both cases, methods that require contact with the population (such as enumeration, nomination and multiplication methods using multipliers derived from surveys) will benefit from including members of the population itself in the contact team.

Access through institutions

Populations at risk for HIV are often in contact with institutions which can provide access either to individuals or to records. Such institutions vary by the population at risk, but may include health and treatment services for drug abuse or STIs, other medical services such as emergency wards, the criminal justice system (including courts and prisons), and broader public institutions such as schools.

The problem with individuals contacted through most institutional settings is that they rarely represent all members of a population at risk. For example, drug treatment programs typically attract chronic, long term users at the conclusion of their drug using careers. Jails and criminal justice settings also under-represent newer users, drug users who are not physically dependent and those not involved in criminal activities to support their drug use. Emergency room samples over-represent users of more toxic substances. STI services will tend to over represent sex workers with the riskiest behaviour, especially if treatment, rather than routine screening, predominates.

The critical point with institutional data is to have as much information as possible about what segment of the population it represents. Methadone treatment programmes will only yield information about opioid users, private programmes will only include drug users that can afford to be in treatment, and jail records will reflect those drug users engaged in criminal activities and not able to bribe their way out. With some estimation methodologies, estimates using data from these populations may have to be adjusted to reflect their relationship to a wider population of interest.

Community-based entry points

An additional option to using institutional entry points is to use community-based entry points. These are commonly drawn from mapping exercises which reflect the places and times when members of the population gather. The principle of multiple entry points holds here, too: different community-based entry points will give access to different sub-sections of the community. A mixed but gay-friendly bar in an expensive hotel that is the favoured hang-out of film stars and designers will provide an entry point to a section of the population of men who have sex with men that is very different from the section accessed through a heaving gay night-club in the port area that has a cover charge of a dollar. Again, the rule for robust estimations is to be clear about what fragment of the population is represented in the estimation data.

Where there is no access

It is important, also, to be aware of the likelihood that certain members of the sub-population will be missed entirely. Female drug injectors, for example, are often underrepresented in estimates because they have limited or only indirect contact with the access points that are mapped: they use drugs supplied by their male partners rather than buying themselves, and are less likely to engage in criminal activity (or get arrested for it) than male drug users. Information from the injectors who are accessible may be used to give some idea of the magnitude of those who are not accessible (for example male injectors may be asked whether they buy drugs for anyone else, or whether they regularly inject with their regular female sex partner), but it remains likely that direct access to some sub-populations will not be achieved.

Using available data sources

Many of the estimation techniques described in this document rely in part or entirely on data derived from existing sources. A thorough knowledge of what data are available is vital to choosing the most appropriate estimation method (see page 33). Compiling this knowledge is, once again, easier said than done. Often, relevant data are collected by different institutions or government agencies, and it is sometimes hard for public health officials to get hold of these data. They may not know the data exist, and may not cast their nets widely enough outside familiar health information systems to find out. They may have difficulty persuading agencies such as the police or criminal justice to share the information for the purposes of size estimation, because law enforcement agencies sometimes feel that the presence of sex workers or drug users reflects a failure on their part. Service providers may be reluctant to share information because they fear that it will compromise the confidentiality of their clients.

First, do no harm

The guiding principle of public health applies also to population size estimation. Many techniques do, in fact, require some kind of identification of individuals, and identifying individuals at high risk for HIV is often synonymous with identifying individuals engaged in illegal activities. While this information may be safe in the hands of public health officials, it may not be safe in the hands of other groups. The desire to know how many people are at risk for HIV should never be allowed to take precedence over the rights and welfare of the members of the populations at risk.

The danger of a backlash exists not only at the individual but also at the population level, through the mere publication of information about the existence and size of a sub-population. Even anonymous estimation methods that do not identify individuals in any way can lead to a "spot-light" effect, where the population at risk for HIV is suddenly publicly identified as the 'source' of infection or potential infection. Some countries have moved beyond the stage of targeting sub-populations in a negative way just because of a potential association with HIV, but this is by no means universally true. Where there is a real possibility of widespread "scapegoating", leading to harm to the populations whose size is being estimated, it may be better to drop the whole exercise.

The politics of good estimates

It is generally assumed that governments under-estimate the size of the populations at risk for HIV, and that good estimation exercises will lead to publication of far higher numbers of drug users, sex workers and gay men than were previously acknowledged. This has indeed happened in some countries, and in some cases governments have simply chosen to ignore the results of the estimation work. But in practice, transparent and methodical estimates of the size of populations at risk for HIV have more often than not resulted in far *lower* estimates than those regularly published in the press and HIV-related documents.

This produces headaches of its own. More money than ever before is available for HIV-related programming, and the rhetoric of community driven and "multisectoral" responses means that much of that money is being channelled through non-government organisations. NGOs can be fiercely competitive, and many have an interest in seeing estimates of the size of the populations they are serving kept high. In some countries, NGOs have sought to discredit independent surveillance organisations publishing the results of population size estimation exercises based on transparent and robust methods.

4 METHODS FOR ESTIMATING THE SIZE OF POPULATIONS AT HIGH RISK FOR HIV

"In an effort to deal with the growing devastation wrought by the acquired immunodeficiency syndrome, the public health community is naturally eager to deploy the entire arsenal of epidemiological methods to estimate the scope of the current epidemic. We must, however, remain aware of the shortcomings of each methodology." (Neugebauer and Wittes, 1994, p 1069)

This chapter describes different methods that can be used, either alone or in combination, to arrive at estimates of the size of populations at high risk for HIV. Each method is described together with its strengths and weaknesses. Worked examples are given. The are based in part on actual application of the various methods. While these are derived mostly from the Asian region (because of the participants of the meeting which led to this document were largely from Asia), the methods are equally applicable to other regions. Details of actual country exercises can be found in some of the papers listed in the bibliography. While this section discusses strengths and weaknesses in general terms, Chapter 5 on page 33 gives more details about how to select the appropriate method in a specific situation.

Census and enumeration methods

How they work

Census and enumeration methods boil down, in essence, to counting people. Census methods try to count every individual in a population, for example by visiting every brothel in the country and collecting information on the number of individual sex workers based in each brothel. This has to take place in a very short space of time, since otherwise migration between sites may lead to double counting.

Enumeration methods are very similar, but instead of counting every individual they generally start with a sample frame or list, choose a sample of "units" (such as brothels or shooting galleries) from within that list, and count only the individuals within those chosen units. The number counted is then scaled up according to the size and structure of the sample frame. In other words they may count the number of brothels, visit a third of them to get an average number of workers per brothel, and then multiply the average number of workers per brothel by the total number of brothels counted.

Enumeration methods: an example

The health minister wants to apply for a World Bank loan to provide HIV prevention services to the population. In preparing a proposed budget, her staff needs to know how many sex workers will be in need of regular STI screening services.

Ministry staff get a list of all registered brothels from the department of social welfare, which lists 67 establishments nationwide. Provincial field workers visit every one of the brothels in a two-day period, and find nine have been closed. In the course of their field visits, they identify 12 new brothels. In all of the brothels, they ask the owner to report how many sex workers work there. In a random selection of 20 percent of the brothels, they go from room to room, counting the number of sex workers present and asking those present to report any regular workers who are on menstrual leave or absent for another reason. In no case does the total vary more than eight percent from the total reported by the brothel owner. No brothel-owner over reports the number of workers, and in three provinces there are no discrepancies at all between the number reported and the number counted, including absentees. The provincial field workers record the total number of workers reported, with the possible margin of error, as follows:

Northern: 12 brothels, 762 workers, discrepancies + 2 to + 5 percent (777 - 800)
North-eastern: 6 brothels, 52 workers, no discrepancies
Western lakes: 3 brothels, 102 workers, discrepancies + 4 percent (107)
South-east autonomous region: 2 brothels, 24 workers, no discrepancies
Eastern frontier: 9 brothels, 150 workers, discrepancies 0 to + 3 percent (150 - 155)
Coast: 18 brothels, 1,064 workers, discrepancies + 1 to +7 percent (1075 - 1139)
Federal capital state: 22 brothels, 1,982 workers, discrepancies +5 to +8 percent (2081 - 2141)

The estimate : first approach

This approach assumes that there is a regional variation in the likelihood of reporting correctly. To get the most accurate count, we should apply the variation for each region to the estimate for that region, then sum up.

To get the upper limit estimate, add up the upper limit figures from all 6 provinces:

800 + 52 + 107 + 24 + 155 + 1139 + 2141 = 4,418

To get the lower limit estimate, add up the lower limit figures from all 6 provinces:

777 + 52 + 102 + 24 + 150 + 1075 + 2081 = 4,261

To get a single point estimate, take an average of the two:

(4418 + 4261)/2 = **4,340**

The estimate: second approach

This approach assumes that the observed variation in under-reporting is due to chance, and that it would in fact be safer to apply a single national “average” figure for under- reporting to all the provinces. This is not possible with the information given above, because only the range of underreporting for each province is given, rather than the figures for each of the establishments visited.

Instead, we have to go back to the full data on underreporting (not shown), to derive the national average over all 72 brothels visited. The figure this gives was an average underreporting of 4.2 percent. So the estimate is the sum of all reported numbers, plus 4.2 percent:

762 + 52 + 102 + 24 + 150 + 1,064 + 1,982 = 4,136 + (4,136 x 0.042) = **4,310**

Enumeration methods: an example

An overseas-funded HIV prevention programme has been asked to expand its activities to a new province, which contains the country's second largest city. It is known that injecting drug use is endemic in some areas of the city, and the organisation would like to plan outreach services and harm-reduction activities for drug injectors. They therefore need to estimate the number of injectors in the city.

The estimation team starts with a map of the city, listing all the city's 100 wards. (The boundaries of a ward are drawn according to the electoral rolls, so that each ward is home to roughly 30,000 adults of voting age.) The team first talks to the city police chief about injection drug use. The police chief lists 16 wards where drug-related arrests and seizures are particularly high. Cross -referral with police records confirms these as problem districts, but show high rates of arrest also in another two districts in the police chief's brother's parliamentary constituency. Police records suggest that another 24 wards have lower but still visible levels of drug-related activity, while the remaining 58 wards are virtually drug free. These are mostly in the middle-class and well-off suburbs. The estimation team then visits the city's largest drug treatment programme. Records there show that over 20 percent of attenders come from two of the medium wards and nine of the "drug free" wards, principally those with university campuses. It turns out these drug injectors' families can afford the treatment fees, which means they are relatively well off. Being relatively well off, they are also less prone to arrest than poorer injectors. Interviews with key informants from the college student and the inner-city slum drug worlds confirm these findings.

The survey team therefore draws up a list of city wards by level of injection prevalence. Twenty-nine of the city's wards are classified as high prevalence (16 supplied by the police chief, plus 2 in his brother's constituency, plus 11 identified from the drug treatment centre — 2 previously classified medium and 9 previously classified as low), 22 wards as medium prevalence (the 24 identified from police records minus the 2 reclassified as high prevalence by the treatment centre), and the remaining 49 as low prevalence.

Approach one: sample wards by level of prevalence

In the first approach, a random sample of 10 wards is selected at each level, and a drug user, former drug user or outreach worker from that ward community takes team members around on an enumeration exercise, moving from street to street, talking to known drug users in the street and identifying drug injectors who live there. They find the following:

High prevalence wards:

Mean number of drug injectors per ward: 839.6
Lowest number of injectors in a high prevalence ward: 620 injectors
Highest number of injectors in a high prevalence ward: 1221 injectors
Using standard statistical methods, the survey team calculates the standard deviation among the 29 high prevalence wards as 140.5. This gives a 95 % confidence interval of 559 - 1120, meaning the team can be 95 percent certain that the true mean of drug injectors per high prevalence ward lies within that range.

Medium prevalence wards:

Mean 320.3 Highest: 595 Lowest: 220
Standard deviation: 44.3 95 % confidence interval: 232 - 409

Low prevalence wards:

Mean 71.9 Highest: 115 Lowest: 22
Standard deviation: 28.5 95 % confidence interval: 15 - 129

The estimate:

To get the central estimate, multiply the mean number for the high prevalence wards by the number of high prevalence wards (839.6 x 29), multiply the mean number for the medium prevalence wards by the number of medium prevalence wards (320.3 x 22) and multiply the mean number for the low prevalence wards by the number of low prevalence wards (71.9 x 49). Then add up the results:
24,348.4 + 7,046.6 + 3,523.1 = **34,918**

To get the upper and lower estimates, repeat using the upper and lower bounds of the 95 percent confidence interval respectively
Upper limit: (1120 x 29) + (409 x 22) + (129 x 49) = 47,799
Lower limit: (559 x 29) + (232 x 22) + (15 x 49) = 22,050

Note that since the wards are based on electoral rolls and are roughly equal in population size, no effort has been made to weight the estimate by size of the ward.

The good and the bad

Census and enumeration methods are mathematically straightforward. Where a good sample frame already exists and where the population of interest is well-defined and visible, census methods can actually be less time-consuming than other, more sophisticated methods, and can give better results. They are especially well-suited to counts at the local level.

Census methods are not well suited to hidden populations, or to areas that are geographically diverse and where populations at risk are scattered. In these situations, it is rarely possible to field a sufficiently large staff to do the count over just a day or two.

Enumeration methods share the advantages and disadvantages of census methods, although because they tend to cover a smaller fraction of the population they usually require fewer people. Working with reliable community guides covering relatively small areas, enumeration methods can also reach more "hidden" populations than census methods. In other words, census methods work well when it is possible and feasible to reach everybody in a given sub-population, while enumeration methods work better when it is only possible or feasible to reach a fraction of them.

Population survey methods

How they work

Surveys of the general population, or subsets of the general population, are common in most countries. The most common form of general population survey is the household survey, where a questionnaire is administered to residents of a sample of households drawn from a sample frame that is representative at a national or regional level. In industrialised countries these surveys are sometimes conducted by telephone, but in developing countries they usually involve face-to-face interviews with members of a survey team. Household surveys commonly explore health and nutrition issues, including contraception. Many countries in Africa and the industrialised world and a few in Asia and Latin America have conducted household surveys of sexual risk behaviour, and a few, mostly in industrialised countries, conduct regular household surveys of drug use. These surveys give a prevalence of risk behaviour among the respondents sampled. Because the sample is known to be representative of a larger population (often the whole national population) that prevalence can then be applied to the whole population to give an estimate of the total number of individuals engaging in the behaviour. If there is reason to believe that a significant proportion of the population is not found in households (because they are in barracks, hospitals or living on the streets for example), and if data are available on how those populations relate to the general population, it may be possible to adjust the estimate to reflect this group.

Certain sub-sectors of the general population can be reached though other, nationally representative institutional surveys such as school surveys.

Other types of population surveys are also possible. These include surveys of sub-populations at high risk for HIV transmission, most commonly male, female and transsexual sex workers, groups of men likely to be their clients, injecting drug users, and men who have sex with men. Generally, the places where members of a given sub-population gather are mapped, and a random sample of the population of interest at those locations is invited to participate in a survey. The Asian region leads the world in repeated cross-sectional surveys of risk behaviour for HIV, and many other countries also have well-established behavioural surveillance systems in these groups. Unlike general population surveys, behavioural surveillance in high risk groups cannot, by itself, give an estimation of population size. Whereas general population surveys provide the prevalence of a behaviour among a larger population, the respondents for behavioural surveillance are chosen *because* they engage in that behaviour (e.g.

Population survey methods: an example

From its behavioural surveillance system among female sex workers and clients, a country with a population of 30 million people knows that although 70 percent of sex workers ask their clients to use condoms, only 21 percent of clients do so. The AIDS programme wants to shift the focus of its interventions away from working only with sex workers. The programme wants to work with male clients more to promote condom use. They would like to run a mass media campaign aimed at lower and middle-income men, but the religious authorities object, saying that most clients of sex workers are foreigners, so no domestic response is necessary.

The national health survey is going in to the field in a few months time, and will interview 8,500 men in a representative sample of households about a range of health issues including family planning. The AIDS programme persuades the ministry to add a single question to the survey, asking whether men have bought sex from a female sex worker in the past 12 months.

The results of the survey show that 13.7 percent of men aged 15-54 have bought sex in the past 12 months. The most recent census shows there are 7.3 million men in that age bracket.

To get the estimate, simply multiply the number of men in the age bracket by the proportion of those men who say they have bought sex in the last year:
7,300,000 x 0.137 = **1,000,100**

This estimate seems low compared to the number of sex workers registered with the social affairs department, and with the client turnover they report. Behavioural surveillance data show that sex workers report that a third of their clients are truckers or sailors, and a fifth are from the military. Some 70 percent of rank and file military live in barracks excluded from household surveys. In behavioural surveillance among truckers, men report being away from home an average of three weeks a month, so on average only a quarter of them would have been found in the household at the time of the survey (although all of them should have been counted in the census).

This should lead to two adjustments. First, the household survey results should only be applied to the number of men in households at the time of the survey, not to the whole population. This means finding out from unions or other sources the size of major institutional or mobile populations (prisoners, soldiers, truckers, sailors) and subtracting all or a proportion of them from the census population to which the estimated proportion buying sex is applied.
Military: 300,000 70% in barracks = 210,000
Truckers: 450,000 Away from home 3 out of four weeks = 337,500
Sailors: 210,000 Away from home 2 out of four weeks = 105,000
Male Prisoners: 90,000 = 90,000

In other words, the census population should be reduced by 742,500 before any "general population" estimate derived from the household data is applied to it.

The second adjustment concerns the clients from the groups identified by sex workers are frequent clients. Behavioural surveillance among these groups suggests that some 61 percent have visited sex workers in the last month, and 65 percent in the last year. The similarity between the monthly and the annual estimates suggests they are frequent clients, accounting for a disproportionately high percentage of all visits to sex workers. This is in line with sex worker reports.

Apply the "population prevalence" of commercial sex to the population likely to be found in households at the time of the survey:
(7,300,000 - 742,500) x 0.137 = **898,379**

Add the "population prevalence" of commercial sex among the high risk groups not found in households (prisoners are excluded because they are not buying sex from female sex workers while in prison):
442,710 x 0.65 = 287,762

The adjusted estimate:
898,379 + 287,762 = **1,186,100**

they are injecting drug users) or because they are more likely than members of the general population to engage in a risk behaviour (e.g. they are men with cash and a mobile occupation whose occupational culture favours buying sex from sex workers, such as sailors). Data collected in behavioural surveillance can, however, be used in some of the other methods described below, such as multiplier methods.

The good and the bad

In general, it is easy to construct a sample frame for a general population survey, and easy to calculate exactly how representative the sample is of the general population. This means that results can easily be extrapolated to the entire area covered by a survey - often nationally. Because the methodologies are well established, the results of a general population survey are usually quite easy to defend. What's more, because of theconventional methods and the wide generalisability, general population surveys tend to be politically influential.

General population surveys are, in other words, very robust ways of measuring the prevalence of behaviours which are widespread in the general population. Arriving at a population size estimate is as simple as applying the prevalence measured in the survey to the population represented in the survey. However most behaviours which carry a high risk for HIV in concentrated HIV epidemics are **not** very widespread in the general population. Indeed it is the rarity of these behaviours that allows HIV to remain concentrated, rather than to spread widely through the general population. Because of this, the usefulness of household surveys is somewhat limited.

Firstly, household surveys do not include every single household, they select respondents from a sample of households. Where a behaviour is rare, it may not get picked up in a sample survey, unless the sample is very large, because the small number of individuals who practice the behaviour may not be included as respondents, even where selection is random. For example if only one person in 10,000 injects drugs and a survey covers 2,000 respondents, there is an 80 percent chance that the survey will register zero drug injection. If the sample does happen to include a single drug injector, then the survey will record prevalence of injecting drug use at 0.05 percent - five times its true value in the population. To get an accurate idea of this low prevalence activity, a far larger sample size would be needed. But large samples carry large price tags; generally, it does not make sense to spend US$ 8 - 10 per respondent (the average cost of a household survey) just to get information on a behaviour which may be practiced by fewer than one in every 100 respondents.

Secondly, many of the people engaging in behaviours which carry a high risk for HIV transmission are not in households at all. Sex workers may live in brothels, their clients may be in barracks or company dormitories, drug injectors may be living on the streets or in jail. None of these concentrations of people with risky behaviour will be picked up in a household survey, so the estimates based on the population prevalence may miss a significant number of people who are, in fact, part of the population of interest.

Thirdly, the more stigmatised a behaviour, the less likely people are to tell the truth about it to an interviewer, especially if they are approached in the context of their home and family. Household surveys therefore tend to give significant underestimates of the prevalence of very risky behaviours. The extent to which a behaviour is stigmatised clearly varies enormously by country. In many Asian countries, male consumption of commercial sex is a cultural norm that carries little stigma, and household surveys have been used successfully to estimate the number of male clients of sex workers in several countries, including Thailand and Cambodia. On the other hand it is unlikely in the Asian context that a household survey would yield good data about the much more marginalised practice of sex between men. In many Western countries, the reverse it true. Being gay is more socially acceptable than visiting sex workers. But even then, household surveys lead to underestimates. In Canada, for example, the number of men who have sex with men was estimated from a general population survey reaching 2,500 people by telephone, as well as through three other indirect methods. Estimates arrived at indirectly through multiplier and other methods (see below) were on average twice as high as the estimate based on the survey.

In summary, household surveys can serve to give a bare minimum estimate of stigmatised behaviours, but are unlikely to deliver the full picture in a cost-effective way. If a household survey is happening for another reason, however, it may be worth including questions about the behaviours of interest to get a minimum benchmark nationally, and then conduct special in-depth validations in a smaller area to provide an estimate of the extent to which people are under-reporting the risky behaviour.

Multiplier methods

How they work

Multiplier methods generally rely on having information from two sources that overlap in a known way: the first is usually an institution or service with which the population to be estimated is in contact, and the second is the population at risk itself. Estimates are derived by multiplying the number of people who attend the institution or service over a certain period by the inverse of the proportion of the population who say they attended over the same period.

Multiplier methods can also be used with information from two separate population-based survey samples that intersect in some way, as long as the size of one of the groups is relatively well known. The most common example is sex workers and clients.

In considering whether multiplier methods are appropriate, it is worth casting the net wide for potential sources of information. With a little planning (or luck!) even sources that are not immediately related to the behaviour in question can be used to arrive at estimates of sub-populations at risk for HIV.

The good and the bad

Multiplier methods are relatively straightforward to use. They require good institutional record-keeping, and the right questions inserted into regular behavioural surveillance instruments. They are perhaps the most commonly used of all the population size estimation methods, and are very

Multiplier methods: example 1

The metropolitan authority runs a 100 percent condom campaign, under which all sex workers must be screened for STIs in government clinics every two weeks. Green cards are issued to all sex workers to record screening and treatment.

In a survey of sex workers which used members of the sex work community themselves as field workers, 350 brothel-based and 250 street-based sex workers were interviewed. Altogether, 404 street-based sex workers could show up-to-date green cards. Metropolitan authority clinics report screening 1,750 sex workers each week.

The estimate:
Total number of sex workers interviewed: 600
Number with green cards: 404
Multiplier for clinic attenders: 600/404 = 1.49
Clinic visits: 1,750 per week, multiplied by two because there is a two week screening interval so the number of green cards issued over a two week period = 3,500
Population size estimate: clinic visits time the multiplier
3,500 x 1.49 = **5,215**

commonly found in economic and demographic literature. However very few examples of the use of multiplier methods in estimating populations at risk for HIV have been recorded in the scientific "literature". This is probably precisely because they are simple and user-friendly. There is no mystique, no sophisticated mathematics, no gloss of "high science". These should all be seen as strong positive qualities, and should encourage the use of multiplier methods.

The greatest difficulty in using multiplier methods correctly is finding data for institutions and populations that correspond with one another. To use institutional and survey data together to estimate the size of a population, the members of the population all have to have a chance of being included in both the survey and in the institutional data (for example because they have access to that service). Imagine trying to estimate the number of men who have sex with men from

Multiplier methods: example 2

A country wants to estimate the number of sex workers nationwide. The data available are a nationally-representative household survey, and behavioural surveillance surveys among sex workers and clients. The household survey suggests that 12 percent of the country's 1.45 million adult men reported visiting sex workers in the last year. Behavioural surveillance among male clients of sex workers shows that they go to sex workers on average every two weeks, or 26 times a year, and because they sometimes return to the same woman, they visit an average of 21 different sex workers in a year. Behavioural surveillance among sex workers shows that each sex worker has sex with clients on average 11 times a week, and that she goes home to her village for an average of three weeks during the religious holidays.

The estimate:

Total number of male clients is the number of men in the sexually active age bracket times the proportion that say they buy sex:

12% x 1.45 million = 174,000

Total number of commercial sex acts in a year is the number of men who buy sex, times the average number of commercial sex acts per client per year:

174,000 x 26 = 4,524,000

Average number of sex acts per sex worker per year is the average number of sex acts per sex worker per week, times the number of weeks in a year she works:

11 x (52-3) = 539

Because the absolute number of commercial sex acts between men and women must be the same, we can calculate the number of women working from the total number of commercial sex acts divided by the average number of sex acts per sex worker per year:

Population size estimate:.4,524,000 / 539 = **8,400**

Note that if a substantial sex tourism industry in the country, it may be necessary to adjust this figure upwards.

In behavioural surveillance, 16 percent of sex workers report that their most recent client was a foreigner. On average, then, 16 percent of sex acts will not be reflected in the data collected through household surveys. The number of sex acts must be adjusted upwards.

4,524,000 + (0.15 x 4,524,000) = 5,202,600

This figure is then divided by the average number of sex acts per sex worker per year, as previously:

Revised population size estimate: 5,202,600 / 539 = **9,700**

HIV testing data combined with survey data. In a survey of MSM taken in street cruising areas, 30 percent say they had a voluntary HIV test in the last year. Data from voluntary counselling and testing (VCT) centres show that 700 of the clients seeking HIV tests in the last year gave male-male sex as a risk factor. In principle, those 700 men should represent 30 percent of the total population size, which would by this method be estimated at around 2,330 men (700 x 100/70). But what if VCT is only available to those with health insurance, and gay men with health insurance are wealthier, more likely to be employed in the corporate sector, and less likely than gay men with no health insurance to hang out in street cruising areas? In that case, the men who are represented in the institutional data are not likely to be included in the survey. In other words, the two populations do not correspond, and any population estimate made using these two data sources in conjunction is likely to be quite inaccurate.

One key issue in using multiplier methods successfully is the need to have clear, consistent definitions between different data sources. Firstly, the population definitions must be clear. In principle, it ought to be possible to use the proportion of drug users imprisoned together with the number of drug users in prison to arrive at an estimate of the number of drug users in a city or country. Let us say that 45 percent of respondents in a survey of drug users recruited on the streets and in treatment centres say they have been in prison in the last year in a given city. Over the previous year, the prison service in the city has released 1,234 individuals who had been incarcerated on drug-related offences. A simple estimate would assume

Multiplier Methods: example 3

A country wants to estimate the number of injecting drug users nationwide. There are some sporadic data on injecting habits from injectors in treatment, but the number of injectors in treatment is not recorded, and public health officials know of virtually no other data sources. At a national meeting on estimation, the Narcotics Squad laments the fact that experts from UNDCP have recently released a regional study of law enforcement which estimates that only two percent of drugs coming in to the country are being seized by enforcement agencies.

An alert AIDS programme manager asks for the drug seizure data, which the Narcotics Squad is happy to share. They report that in the last year they seized 1.2 tonnes of cannabis, 47 kilos of cocaine and 39 kilos of heroin.

If only two percent of drugs are seized, this means that there must be
39,000 x 100/2 = 1,950,000 grams of heroin in the country in the year.

Only half of the drug users in treatment inject drugs, and 91 percent of those who do inject say they inject heroin only. Mean injection frequency is three per day. Most injection is in groups with drugs mixed in a syringe, each user gets about 0.05 grams at each shot.

That means in a year, each user is injecting
0.05 x 3 x 365 = 54.75 grams of heroin per year

The total number of heroin injectors must therefore be:
1,950,000 / 54.75 = 35,616

But interviews of those in treatment suggest only 90 percent of injectors use heroin, so we have to add in the people who inject other drugs.

Population size estimate: 35,616 x 100/90 = **39,573** that those 1,234 individuals represented 45 percent of drug users in the city, and that the true number of drug users was therefore in the region of 1,234 / 0.45, or around 2,760. But this estimate relies on the assumption that "drug user in prison" is synonymous with "prisoner incarcerated on drug charges". It assumes, that is, that all drug users in prison were imprisoned for drug-related offences, and that all of those imprisoned for drug-related offences were drug users. In practice, this is unlikely to be true. Many of those classified as "drug users" by this definition may be people who are dealers but not users, while many people who are users and in prison but were imprisoned for theft and other criminal activities (which may or may not be related to generating money to support a drug habit), or who started using drugs while in prison, would not be captured. An estimate based on these unclear definitions would certainly be inaccurate.

Secondly, the time reference period must be clear, and must be the same in both data sources. Say you have a population based survey which asks men if they have visited sex workers in the last year: this would immediately give an estimate of the number of clients of sex workers. As we saw in the example on page 19, if you also have behavioural surveillance data among clients of sex workers which records the average number of visits per client per year, you can calculate the total number of commercial sex acts in a year. If in addition you have data among sex workers which records the average number of clients per sex worker, you can divide the total number of sex acts by the average number of clients per sex worker to arrive at an estimate of the total number of sex workers in a country. If, however, your initial population based survey did not ask about commercial sex in the last year, but asked instead about whether the man had *ever* bought sex (in his lifetime), then none of these estimates could be made. It is worth noting, also, that misreporting of the timing of events is very common, and the longer the reference period, the more common it becomes. People often have a very vague notion of "the last 12 months" and will report events that happened in the not-too-distant past. In measuring levels of risk behaviours this misreporting is not very important - the need to maintain standardised indicators over time and between locations takes precedence over the need for a very accurate time reference for events. However time reference periods becomes more important in size estimation, because you are likely to be comparing reported behaviours or events with records which record the date very accurately. A tip: if data are being collected from individuals just for the purposes of size estimation, and if they are to be compared with accurately recorded facility-based data, it is usually best to choose a memorable event as the limit for the time reference, rather than a more general definition such as "one year". People may not be able to report accurately whether they have been in jail in the last 12 months, but they may find it easier to remember if they were in jail since last Christmas, or since the Soccer World Cup, or some other relevant event.

Thirdly, the age range of the populations to be compared must be similar. If people under 18 were excluded from behavioural surveillance, data from that population cannot be used as a multiplier for data from an STI clinic which has a large number of patients in their mid-teens, unless suitable adjustments are made.

Finally, the catchment area for the services or institutions must be clear, and should ideally be the same as that covered in the sub-population survey from which multipliers are derived. In other words, the populations must correspondeographically, as well as by definition. The number of people visiting a needle exchange in St Petersburg cannot be multiplied by the proportion of those who said they used a needle exchange in a survey in Moscow, since service availability and use may differ very significantly between the two cities.

Because of the catchment area issue, multiplier methods based on service use are most commonly used at the local level. They have, however, sometimes been used at the national level. A notable example is a recent national-level estimate of the number of injecting drug users in Pakistan. In this case specific estimates were made for four cities where behavioural surveys and treatment use data were available. Behavioural surveys asked about use of treatment in the last year, the proportion of users who had been in treatment was calculated for that city, and the appropriate multiplier was applied to the numbers in treatment in that city in the past year. The regional variations were examined and with the advice of key informants, multipliers were extrapolated and adjusted to apply to other parts of the country, where treatment data were available but where there was no specific population-based data about treatment use from which multipliers could be calculated. Finally, the four city estimates and the remaining regional estimates were summed up to produce a national estimate for the number of drug users in Pakistan.

Nomination methods

How they work

Nomination methods start at the tip of the iceberg: the small but visible fraction of a larger hidden population. Drug users in treatment or in detention, men who go to bars openly identified as gay meeting places, etc. These individuals are contacted and asked to provide the contacts for other individuals who share their risk behaviour, even if they are not to be found at the same venue. These individuals may then be contacted and asked to provide contacts for other individuals, etc.

This is sometimes used in conjunction with multiplier techniques. People are asked not only to nominate individuals, but to say whether those individuals have been in

Nomination methods: an example

25 gay men are contacted at Starlight, the largest disco in the city, on a Wednesday night. Wednesday is the only night of the week that Starlight is all gay. Each of these men is asked to give the name and contact numbers for five gay friends or acquaintances. 35 duplicates are found in the list. The 90 nominated men are contacted and asked if they went to Starlight last Wednesday. Sixty say they did. They are then asked for the names and contact numbers of another five gay friends or acquaintances, and asked whether these men went to Starlight. 112 names are found to be duplicates with men included either in earlier rounds or in this round; 103 of the newly nominated men were reported to have been in Starlight last Wednesday.

The manager of Starlight reports that 750 entrance tickets were sold at the disco last Wednesday.

The population size estimate:

Total number of men contacted or nominated, after duplicates are removed:

25 initial contacts + (125 of their contacts - 35 duplicates) + (450 of their contacts - 112 duplicates)

25 + 90 + 338 = 453

Total number of men contacted or nominated who were at Starlight last Wednesday:

25 + 60 + 103 = 188

Multiplier for total population = 453/188 = 2.4

Estimated gay population in the city:

Stardust attendance x multiplier = 750 x 2.4 = **1,807** contact with a certain site over a given time frame. The total number of individuals named and the proportion of them said to have attended the site is used as a multiplier for the actual number of people known to have attended the site.

The good and the bad

By using members of a sub-population to identify other members, nomination methods provide a convenient way to access the hardest to reach populations. However, they come with several strong warnings. The hardest to reach populations are usually those whose behaviour is most illegal or most highly stigmatised. It is precisely from these populations that it is most hazardous to take names and identifying information, lest it fall into the wrong hands and lead to arrests or other abuses. But these populations also tend to be very highly networked, which means that many of the referrals given are likely to be duplicates. In order to remove these duplicates and arrive at robust estimates, good identifying information is necessary.

Because nomination techniques are driven by members of the “core” group themselves, it is not possible to know how representative they are of the overall risk population. If there are distinct groups within the sub-population that rarely mix, then this method may lead to some parts of the population being missed entirely. Say, for example, that you want to design an HIV prevention programme for migrant workers in a city with an active commercial sex industry, and want to estimate the number of migrant workers in that city. Nomination techniques may work well if migrant workers from different backgrounds work in similar industries and mix together. However if migrants from Indonesia work in manufacturing while migrants from Bangladesh work in construction and migrants from the Philippines work in the service industry, and the three groups live in different communities and don’t mix, then nomination techniques that only have entry points to one of these communities may miss the other parts of the migrant worker population entirely.

The success of this method is determined to a large extent by the nature of networking in the population. It works best in populations with high levels of interaction between different networks (as is commonly the case in drug injecting populations, because of the focal link provided by dealing networks). In such cases, modelling and experimentation show that a diverse initial sample and at least three waves of referral yield a sample similar in composition to one obtained by probability methods. Some of the limitations of nomination methods can also be overcome by using sophisticated mathematical techniques that estimate the likelihood that an individual will be included in a sample. The skills needed to use these techniques are rarely available in national AIDS programmes.

Of the available methods, nomination methods are probably the least appropriate for widespread use in estimating the size of populations at high risk for HIV.

Capture-recapture methods

How they work

Capture-recapture methods originated among naturalists trying to estimate the size of animal populations. A number of animals are captured, tagged and released. An independent recapture is conducted at a later time in the same area, and the proportion of the captured animals that have already been tagged is calculated. Information on the size of both captures and the proportion captured twice is entered into a mathematical formula to calculate the total size of the population, including the “unseen” portion.

With human populations, direct approaches have sometimes been used. An estimate team visits all cruising sites for male sex workers, for example, counts all the men there, distributes some token, such as a blue invitation card to a male sexual health clinic. A week later, they revisit the sites, take another count of individuals, give out yellow invitation cards but ask whether they have

The formula is follows:

N = (number in first capture x number in second capture)/ number in both captures

95% confidence intervals can be constructed around the resulting number, using the following formula:

$95\% \text{ CI} = N +/- 1.96 \sqrt{Var(N)}$, where Var (N) is calculated as follows:

Var (N) = ((# in first capture x # in second capture)(# in first capture - # recaptures)
(# in second capture - # recaptures)) / (# recaptures)[3]

already received a blue invitation card.

Direct contact is not always necessary in capture-recapture, however. Existing listings of individuals in contact with a certain institution or service may be used, for example. Whatever the method of "capture" used, a number of key criteria must be met.

- Samples must be independent from one another, and not correlated
- Each member of the population should have an equal, non-zero probability of being "captured"
- The individuals identified in both captures must be correctly identified as recaptures, and no-one else should be identified as a recapture
- There should be no major in-migration or out-migration from the population between the initial and the second captures

The good and the bad

Capture-recapture became fashionable in the early 1990s, and there was a rash of publications on its use in estimating the size of populations at risk for HIV around 1994. Since then, it has rather faded from view, for good reason.

Capture-recapture techniques are useful in estimating the size of relatively hidden populations that are, nonetheless, in regular contact with two types of institutions or services. These conditions are sometimes met. In the Thai capital, for example, the Bangkok Metropolitan Authority provides free treatment for opiate users. A high proportion of opiate users pass through this treatment, but arrests are also common and many drug users come into contact with police and the prison services. One attempt at capture-recapture estimates compared individuals in treatment with individuals "captured" in police stations. However even here, the second "capture" was uncertain, since many drug users in detention are detained for offences not directly related to drug use. In this case, researchers performed urine tests on all those in detention to identify opiate users, but this would clearly not be feasible or ethical outside a research setting.

It is exceedingly rare that the four conditions for reliable capture-recapture estimates are met in populations at high risk for HIV infection. The principle of the independence of samples is frequently violated because institutions and services either cross-refer to one another or are mutually exclusive. Many possible entry points provide preferential access to one social class or another (for example middle-class drug users are more likely to be in private treatment programmes while low class drug users are more likely to be in jail), and there may be important geographical variations. A positive correlation will lead to an underestimate of the true population size, because it will increase the number of double-captures and therefore increase the size of the denominator for the population size calculation, as in the example 1.

Capture-recapture methods: Example 1

The true number of drug users in a city is 35,000, but this is not known. A capture-recapture exercise is planned, using lists from a police detention centre and a drug treatment centre. There are 2,700 people detained and 3,506 people in treatment. 1,900 people appear were on both lists, having been both detained and been in treatment in the past year.

The population estimate would then be:
N = (2700*3506)/1900 = **4,982**

This is a huge underestimate of the true number of drug users, and is only about half the amount officially estimated by the Department of Social Affairs. Further investigation reveals that national harm reduction policies state that the police should refer any drug users in detention to drug treatment. In other words, those who have been detained have a very high probability of attending treatment. The two samples are positively correlated, and are not independent.

A negative correlation, on the other hand, will lead to an overestimate of the entire population size, because the denominator (the number of dual-captures) will be smaller than would be expected if the samples were independent, as in the example 2

The assumption that each population member has an equal chance of being included in the sample is also frequently violated. An example already cited is the ability of some sex workers but not others to buy themselves out of detention. Drug treatment programmes that charge the equivalent of two months salary for a three week stay may provide data for a capture recapture. It is clear, however, that only drug users who can afford to pay for treatment have a chance of being included in the sample. In some cases, a significant proportion of the overall population of interest has virtually no chance at all of ending up in either one of the “captures” and will therefore go un-estimated. Sex workers who work by mobile phone and who serve the high end of the market will never be arrested, nor will they ever go to an STI screening and treatment centre in a red light district, even though their behaviour may still mean they are at high risk for HIV, for example.

Capture-recapture methods: Example 2

The true number of sex workers in a city is 35,000 but this is not known. A capture-recapture exercise is planned, using lists from a police detention centre and a STI treatment centre for sex workers There are 2,700 people detained and 3506 people presenting for STI services. 139 people appear on both lists.

The population estimate would then be:
N = (2700*3506)/139 = 68,102

This is a huge overestimate of the true number of sex workers, and is nearly five times the number estimated by the Department of Social Affairs. Qualitative research with sex workers reveals that sex workers who have a high client turnover and good earnings are less likely to go to the detention centre. In the first place, they spend more time with clients and less time on the streets. Secondly, if they do get arrested, they buy their way out before going to the detention centre. They are, however, more likely to go the STI clinic than the less popular sex workers. Their higher client turnover means greater exposure to STIs, and their higher earnings mean they are more willing to pay the nominal fees charged by the clinic for screening and treatment. In other words, women on the detention lists represent a different stratum of sex workers than women on the clinic list. The two samples are negatively correlated, and are not independent.

As for the “integrity of the mark” (which allows all true recaptures to be identified and excludes people who have not actually been captured twice), that too is problematic. The most commonly used “mark” is the name, which appears on registers of those in treatment, in detention etc. In some cultures, the pool of surnames (often associated with clan lineage) is relatively small and it is difficult to distinguish between duplicates and separate individuals who have the same name. In some countries people regularly change names according to life events or because a new name may be associated with a change in fortune. Further, names may be spelt differently, and people may use different parts of their names in different situations. On top of that, where illegal behaviours are concerned, people often give false names or aliases, which are not constant over time.

Where the “mark” is a token given out during direct contact, different difficulties arise. If the token is of little value, people may forget having received it. If it is of significant value and more memorable, they may deny having received it in the hope of receiving another.

The final assumption involves the stability of the population. Capture-recapture techniques are only really reliable in “closed” populations, where there is no major inflow or outflow of people. As we have seen, some of the sub-populations at high risk for HIV are peculiarly mobile. High migration rates undermine the ability to perform good capture-recapture. Equally, it is hard to use these methods when the population itself is changing rapidly, as is often the case in the early phases of an epidemic of drug injection, when large numbers of new injectors are constantly being recruited. Where data sources allow, these problems can be minimised by ensuring that two rounds of “capture” are as close together as possible in time.

Capture-recapture is the most frequently cited of all the estimation techniques. It is seductive in that it appears relatively simple, and has an aura of being “scientific”. However most authors who have reviewed the use of these methods for estimating the size of populations at risk for HIV seem to agree with the following conclusion, reached by Cox and Shipley, in a 1997 paper.

“In practice, estimation of drug-misusing populations using simple, two-sample capture-recapture methods is neither easy nor reliable.”

A comparison between capture-recapture and simple census methods conducted among bar-based sex workers in Bulaweyo, Zimbabwe, found that the more “sophisticated” method was more trouble than it was worth. It took the estimation team one night and 100 field workers to count 6,997 working women in 56 bars, and this was achieved without bothering the women at all. It took two nights, 2,850 interviews with women, 2,850 gifts and a data entry and analysis team to arrive at a capture-recapture estimate of 3,894 sex workers: 55 percent of the number actually counted.

A recent capture-recapture exercise among drug injectors in a Baltic state using distribution of cards did not perform well, either. Drug users were visited in the community and given coloured cards by outreach workers. At a later date, the outreach workers returned to ask the drug users if they had received a coloured card. Some 203 cards were distributed initially, while on the second “capture” 91 out of 154 injectors contacted reported having received a card. This led to as estimate of 344 drug injectors, in a city where the needle exchange alone reported over 5,000 individuals using its services. It is highly probable in this example that two major conditions for reliable capture recapture were violated. Firstly, it is unlikely that the two samples reached by outreach workers were independent of one another, and secondly it is unlikely that all IDUs in the city had an equal chance of being contacted by the outreach workers at either the first or second capture

In short, capture-recapture methods are harder to perform well than their initially simple-looking methodology would sugguest These methods should only be used where the violation of the conditions upon which capture-recapture rests can be retricted to a bare minimum. And they should only be used if it is determined that less “glamorous” but more robust methods such as census and multiplication methods cannot be used in the circumstances.

Multiple sample recaptures

The importance of independence is reduced if three or more samples are available. In this case log-linear models can be used to estimate the correlation between the different samples - in other words, to examine whether or not there are any systematic relationships between data sources.

As an example, consider the table above. In this case, three sources of data are available: data from the police on drug injectors arrested in a certain time period, data from treatment programmes on drug injectors in treatment in the same time period, and data from an HIV testing service on drug injectors requesting testing, again, for the same period. All of these data sources provided surname and first name initials, date of birth and gender to identify matches between them.

The data are laid out in an eight-cell table. Cell a) contains those whose names were found in all three sources, cell b) contains those who were in the police and HIV registers but not the treatment registers, etc. The aim is to use all of these bits of data to predict the number of people in cell h) - individuals who did not show up in any register at all. The number not observed by the data sources (h) is generated by fitting a Poisson model to the table.

This multiple-sample method is greatly preferable to a two-sample capture-recapture. In practice, in many developing countries, there are rarely enough sources of information with reliably recorded names to provide a third “capture”. This may be less the case in countries with very highly developed information systems, such as the ex-Communist countries of Eastern Europe.

It is important to note that data should come from different sources wherever possible. On occasion, data from the same source but at different time periods have been used as the basis for a multiple sample capture-recapture. This may be valid if the source is a random selection of individuals in the population at large. But it is unlikely to be valid if the source is a service or location through which only part of the population is accessed. Imagine taking STI treatment records at a sex worker clinic as the basis for a capture recapture. People attending from January to March are considered part of the first “capture”, people attending from April to June are considered part of the second “capture”, and people attending from July to September are considered part of the third “capture”. The results are put into statistical software package, and an estimate is derived. Is this an estimate of the whole sex worker population? No. It does include a fitted estimate of sex workers who had a non-zero probability of attending the clinic but did not happen to attend it during the period of the

Table 1: Multiple sample capture recapture[1]

		Police data (P)			
		Yes		No	
		Treatment Data (T)			
		Yes	No	Yes	No
HIV data (H)	Yes	a) P, T, H	b) P, H	c) T, H	d) H
	No	e) P, T	f) P	g) T	h) ???

[1] Adapted from Gordon Hay, report on ennumeratio in the Baltic Sea region, citing Frischer, M (1992) Estimated prevalence of injecting drug use in Glasgow. Britih Journal of Addiction, 87: 235-243.

estimation exercise. But it does not in any way reflect the sex workers who would never attend the clinic. In other words, it is not an estimate of the number of sex workers, but an estimate of the number of sex workers ever likely to attend the clinic. This is a limitation shared with the truncated Poisson method described below.

The truncated Poisson method

How it works

The truncated Poisson method is an extension of the multiple capture-recapture method. It looks at service use data, and then uses assumptions about the likely statistical distribution of events to estimate the number of people not using the service.

While the method can use any number of observations about the same individual, experimentation has shown that a formula considering only those who visit once and those who visit twice as well as the total number of attenders yields strong results.

There are a number of formulae that can be used to give a truncated Poisson estimate; one that has been used in published estimates of drug injecting populations is:

$$\text{Estimated } N = S / (1 - \exp(2f_2 / f_1))$$

Where f_1 is the number of individuals who attended only once, f_2 is the number of individuals who attended twice, and S is the total number of individuals who attended over the time period.

The assumptions that have to be fulfilled for this method to be valid are the same as those for a two-sample capture-recapture methods using a single data source:-

- Each member of the population should have an equal, non-zero probability of attending the service
- The likelihood of attending the service a second time should be independent of the likelihood of ever attending it at all
- The individuals identified in both captures must be correctly identified as recaptures, and no-one else should be identified as a recapture
- There should be no major inmigration or out-migration from the population between the initial and the second captures

The good and the bad

The strength of this method is that it uses data from a single source, and is very straightforward to compute.

The major weakness is in meeting the first and second assumptions above. In the marginalised populations which are most frequently at risk for HIV and therefore subjects of the estimates described here, there is very, very rarely a service to which all members of the population have universal access, and which all are equally likely to use. In addition, it is unlikely that the probability of a second visit will not be affected by the outcome of the first visit.

This method has been used to estimate the number of drug injectors from needle exchange data in industrialised countries, where barriers to the use of such programmes are low, and where it is believed that a high proportion of drug injectors access the programme at some point. In these cases, the method may have some validity, and since it is cheap and easy to use, it is worth a try.

It is questionable, however, that many developing or transitional countries fulfil the requirements to use this method. Used on needle exchange data, the method essentially assumes that every injector in a city has access to a needle exchange, and have an equal probability of using it over a given time period. There are many reasons why these conditions may not be met. Coverage of needle exchanges is generally rather low, and may be confined to a limited geographical area in the vicinity of the exchange. Many people may have no access to the exchange (and therefore a zero probability of attending) because they cannot afford the transport to get there, or because they are otherwise restricted in their movements, for example in jail. At the other end of the social spectrum are injectors who have easy access to clean syringes and have no desire to be stigmatised as a “common junkie” by using a needle exchange. These people will also have a zero probability of attending the needle exchange. If service is good, people who have ever attended are more likely to come back than people who have never attended, so the likelihood of attending zero, once and twice is not independent, even among those who are potential customers. If the quality of service changes over time, this relationship may be reversed. This change in service utilisation will result in changing estimates of population size, even if the population of injectors is unchanged.

The method assumes that the needle exchange keeps named or coded records of individuals, rather than simply records of total number of client visits.

The time period over which the method is applied greatly affects the outcome. In one example, use of the method over a three month period led to an estimate of 584 injectors, while use over a one year period led to an estimate of 1,041 injectors. Taken as an estimate of the overall population of injectors, this suggests rapid recruitment into the population (violating another of the assumptions necessary for the method to be used with confidence). Taken as a measure of the potential client base for the needle exchange, it reflects a turnover in the user population.

In short, the truncated Poisson method gives an estimate of the total number of drug injectors who have access to the needle exchange and may wish to use it. Where this is the majority of the IDU population, the method will work well (although simple attendance statistics may work nearly as well in this situation).

Compartmental methods

How they work

Compartmental methods look at sub-divisions of populations, splitting a larger population into various “compartments” - for example male sex workers who inject drugs and male sex workers who do not inject drugs. They calculate the sizes of the sub-populations from information about the size of the larger populations, often including information about the rate of transition from one “compartment” to another. The approach is not dissimilar to using Venn diagrams, in

Truncated Poisson method: an example

A city in a transitional country is concerned that drug use is leading to a rise in crime and in blood-borne viruses. Health staff wish to estimate the drug injecting population, and approach the needle exchange for help. The needle exchange provides service records over three months, as follows:

People visiting only once: 473

People visiting twice: 97

Total visits over period: 765

Estimate: 765/ (1-exp ((2x97)/473))

Estimate: 765/(1- exp 0.41) = **1,500**

which different population groups either co-exist or intersect, but every member of the population is in a defined compartment.

The good and the bad

Compartmental methods are rather data-intensive. They are not generally used to make estimates of an overall population, but can be very useful in getting an idea of different sub-populations within a larger population. This is helpful when deciding on programming strategies for various sub-groups. The need for and feasibility of specific HIV prevention services for drug-

Compartmental methods: an example

A cover story in the leading news weekly has alleged that there is "explosive growth" in the male sex industry in the country's main tourist island, and that "the overwhelming majority" of the male sex workers are selling sex to fund their drug habits.

The provincial government initially dismissed the claims, pointing out that in behavioural surveillance only 17 percent of the island's estimated 800 male sex workers said they had injected drugs in the previous year.

However HIV prevention NGOs felt there may be some truth in the claim. They wanted to plan HIV prevention services to tackle the overlap between sex and drugs, and asked for the help of the health department in examining data and suggesting how they should position their services.

The health department re-examined the BSS and the estimates data among IDU and male sex workers and found the following:

17 percent male sex workers said they injected drugs last year.

There were an estimated 2,300 drug injectors, 12 percent of whom said they had sold sex last year.

The first of these data sources would give an estimated 136 injecting sex workers. The second would give an estimated 276 injecting sex workers. It is clear from the difference between the two sources that the populations sampled in the BSS do not reflect all male sex workers. This suggests that a significant proportion of drug injectors who sell sex were not actually identified as "sex workers". Their primary "risk identity" is as drug injectors, which may be why the sex worker sample registered a relatively low level of injection. It is also likely that the IDU sample under-represented men who are primarily identified as sex workers. The overall number of men who both sell sex and inject drugs is probably somewhere between 276 and (276 + 136). More sophisticated compartmental analysis might examine the data on transition between these two behaviours to make a more accurate estimate of the exact number. If for the sake of simplicity we split the difference, we arrive at an estimate of 346 men who both inject drugs and sell sex - 43 percent of all male sex workers. This is not "the overwhelming majority" as the magazine claimed, but it is substantially higher than the BSS data suggested.

Note that because of the difference in the population sizes, there are more "injecting sex workers" with primary risk identities as drug injectors than as sex workers.

This simple example of compartmental analysis would suggest that the NGO would reach more men who both inject drugs and sell sex by using IDU organisations as an entry point than by using male sex worker organisations as an entry point.

injecting sex workers will, for example, depend in part on the size of the injecting and the non-injecting populations.

Multiple methods

There is no reason a country or city should limit itself to a single method for estimating the size of a sub-population at high risk for HIV, nor limit itself to a few data sources. Multiple methods make for checks and balances, and cross validation. Different methods which produce estimates in the same overall range are likely to inspire more confidence than a single estimate produced by a single method.

5 How to choose the best method

Much of the art This chapter discusses those issues briefly, and then turns to the individual sub-populations at highest risk for HIV infection in concentrated HIV epidemics. This section comes with a caveat: it is not possible to give clear guidance about what is appropriate in every situation. General issues are raised, but it is expected that there will be more exceptions than rules in choosing methods suitable for each country and sub-population.

What will you use the estimate for?

Different estimation techniques may produce results that are useful at different levels and for different purposes. You must know before you start why you are doing the estimate and how you plan to use the result, because the end-use of the estimate will guide the choice of methods. If you cannot answer the question: "How will you use this estimate?" then it is better not to bother making the estimate at all.

The first step, then, is to decide how the estimate will be used. If the major purpose of the estimation process is to arrive at a more credible national estimate of HIV infection in order to lobby for more resources for the response, then there is little point carrying out a capture-recapture of sex workers in a single city. It would be better to work on methods which can provide data at a national level, such as the multiplier methods described in Multiplier methods: example 2, on page 20.

If, on the other hand, an estimate is needed to try to plan coverage for a 100 percent condom campaign in a single city, then an enumeration of sex workers in that city, in combination with a behavioural surveillance-based multiplier which calculates the number of clients in the city, would be more appropriate. Enumeration is likely to yield more accurate estimates than multiplier methods, but the former is more time-consuming and expensive than the latter. Even if it were feasible on the national level, it may not be a good use of resources if the goal is simply to generate a broad-brush estimate for advocacy purposes. On the local level, however, and in programme planning situations where relatively precise estimates are needed, enumeration may be both feasible and justified.

The other side of the coin is to think about how the data **could** be used, if they fell into the wrong hands. If the political culture is strongly in favour of law enforcement, or if a burgeoning extremist religious movement has begun vigilante actions against groups they consider "anti-social", then any method which involved identification of individuals (such as nomination and some capture-recapture methods) should be avoided.

It is important, also, that public health authorities bear in mind what each method can and can't deliver. Population size estimation exercises are meant to quantify the problem, not to solve it. Even though some methods share elements that can also be used in building up programme access (enumeration exercises, for example, may use mapping techniques that share much in common with the implementation of surveillance), estimates exercises are intended to count people, not to access them for prevention and care services.

What data do you have available?

Most countries have data sources already available which could contribute to some of the estimation methods described here. Clearly, where such data are available, it makes sense to use them (and therefore to choose the methods that depend upon those data), rather than to set up a parallel data collection system.

Participants at the workshop were asked to compile potential sources of data for use in estimations before coming to the meeting. Separate matrices were provided for sex worker, IDU and MSM populations (examples of these matrices and the accompanying instructions are given in Appendix 2). It was surprising, overall, how many data were available in the region: certainly far more than are currently being used in estimation. But clear patterns were seen. The richest available data sources in almost every country were for IDU populations, while very few country participants were aware of data sources for MSM populations. Sex workers were the most variable, but most countries have at least some data available in this population.

Three types of data were described, overall: service based, site-based and population based. Data specific to the sub-population of interest (for example behavioural surveys of the population, treatment records at drug treatment centres or clinics serving sex workers), and more general data from which the data of interest may be extracted (for example police arrest and prison data, or registers of businesses which identify massage parlours and karaoke bars).

It is important to bear in mind that the fact that data exist does not, of itself, mean that those data can be used for estimation purposes. They may be unusable because it is not possible easily to distinguish which of the data refer to members of the sub-population. An example given earlier is the difficulty of discerning from prison records which of the prisoners are drug users, since many drug users in prison may be charged with other offences. They may be unusable because the data quality is not good enough. Counsellors at VCT clinics may not press clients for correct risk factor information for fear of alienating them, and may therefore record a high proportion of socially acceptable responses which would bias population size estimates using multiplier techniques based on clinic records. Or they may be unusable because the authorities who collect them will not share them with the HIV surveillance or population size estimation team.

On the other hand, the fact that data do not currently exist does not mean that they cannot soon be made available for population size estimation. Most countries also have regular data collection "machine" such as a national census, the national HIV surveillance systems and a national health survey system, which could be used to provide relevant data in the future. These systems are likely to provide data for two main methods: population survey methods and multiplier methods. The inclusion of one or two carefully phrased questions that provide multipliers for other available data sources can greatly increase the possibility for robust estimates.

In other words, before choosing which methods are most appropriate in a country, estimation teams should invest some time and energy in exploring available data but also available data collection systems, in order that estimates can be made with minimum extra investment in original data collection.

Reflections on what may be appropriate in groups with various risk behaviours

There is always a danger in making generalisations about groups. HIV-related behaviours are deeply embedded in specific cultural landscapes. Choices about estimation techniques are also informed by the political landscape. Neither of these landscapes lend themselves to easy generalisation. It is up to every country to examine its own situation and choose what is most appropriate in that situation. The discussions that follow are intended only to provide some framework for the process of making informed choices at the national or local level. Ultimately, readers of this document will have to think for themselves what will work for their country, and will have to take responsibility for recommending the most appropriate method.

Clients of sex workers

Buying sex from female sex workers is the least stigmatised of all the risk behaviours for HIV in many regions with concentrated epidemics. This is particularly true of the Asian region, home to half the world's population. In countries where data are available, between 5 and 25 percent of all adult men say they have bought sex in the last year. This is a big enough proportion of the population that it can easily be captured in a household survey. The proportion of respondents reporting buying sex can be applied to denominators provided by census data. Estimates can be made broken down by age and by marital status if necessary.

Population based surveys reaching a representative sample of households already exist in most countries, and most include male respondents. Adding one or two questions to these existing surveys would be by far the most robust and cost-effective way of obtaining an estimate for the number of clients of female sex workers for many countries in the Asian region, and it may be worth considering in parts of Latin America and other regions, also.

Some governments continue to feel ill-at-ease about including "sensitive" questions such as questions about commercial sex in household surveys. Where it has been tried, however (on every continent including several countries where religious leaders have significant political influence), few if any negative reactions have been recorded and the estimates produced are in range with other data sources.

If a good estimate of the number of sex workers exists, the number of clients can be derived through a multiplier method using behavioural data on partner turnover taken from behavioural surveillance in sex workers and client groups.

Female, male and transvestite sex workers

At a local level, census methods have been shown to give good results for sex workers. This is especially true for sex workers based in establishments such as brothels and massage parlours. It has also been shown to work for bar-based sex workers, and in fact performed far better than capture-recapture in these situations.

Street-based sex workers may also be enumerated, though this will be hard to do in large cities where there is high mobility and overlap between cruising sites. In this case, a simple capture-recapture may be appropriate. Possible sources of information for capture recapture are arrest and clinic records, although as described earlier, these information sources rarely meet the criteria of independence, equal probability and mark integrity required for capture-recapture.

National level estimates may require other methods. Where social affairs or the police list brothels and red light districts, or list sex workers by city or district, these lists can be used together with an enumeration in selected local areas to yield a multiplier for undercounting to arrive at a national estimate.

If general population surveys have given a robust estimate of the number of clients of sex workers, this can be used in a multiplier method in conjunction with information on partner turnover from behavioural surveillance in sex worker and client populations to arrive at an estimate of sex workers nationwide.

Injecting drug users

While household surveys have sometimes been used to attempt to get an estimate of the prevalence of injecting drug use, this is not generally recommended. It is a rare behaviour requiring huge sample sizes, and because it is illegal, significant under-reporting is to be expected. In addition many of those who inject drugs may not be found in households. Certainly, where existing household surveys are being financed and implemented for other purposes, such as general demographic or health surveys, it may be possible to add a question or two on drug use to use as a cross-check with other data sources. However mounting a major household survey simply for the purpose of measuring drug use, and particularly drug injection, is rarely likely to be justified. For this group, other methods are more appropriate.

More sources of information are generally available for IDUs than for any other at-risk population. Records of arrests, of imprisonment and of attendance at treatment programmes are available in many countries. Some countries have registers of addicts. Most countries note that these records are frequently incomplete, and often of poor quality. People who are dead or no longer use drugs may remain on the registers for years. As far as treatment data is concerned, treatment centres often record number of visits rather than number of individuals seen, and it is not always easy to distinguish the number of injectors from that of non-injectors.

In theory, the more reliable of these data sources could be used in capture-recapture methods, which have been used more frequently for this group than for any other at high risk for HIV, though generally in industrialised countries. Capture-recapture may be worth considering for IDU estimates in other countries with good record-keeping practices, for example in Eastern Europe. Elsewhere, however, the use of capture-recapture methods is problematic for IDUs for two reasons. In the first place, there is often strong correlation between the different data sources, and strong biases associated with the data. In many countries, for example, the likelihood of both arrest and of being in treatment are strongly associated with wealth and family position. Secondly, many countries maintain a very strong law-enforcement approach to injection drug use: in many countries in Asia, for example, drug-related offences are punishable by death. As a consequence, any method that relies on comparing names to remove duplicates will either be inaccurate (because people give false names to protect their identity) or may put individuals at risk from law enforcement authorities. This means that nomination methods are also especially inappropriate for estimating the size of IDU populations, and mapping and enumeration are also risky.

Multiplier methods are promising. Because of the difficulties associated with identifying drug users in prison, treatment services are generally preferable as a base for multiplier methods. This is particularly true where treatment services are well-known and records are good. Where many treatment services are available but not all have good data, surveys from which multipliers are derived will have to be quite specific in asking about use of treatment services. Data should be collected about treatment at specific, named services rather than about treatment in general. Since there is pressure to cut down on the length of behavioural survey forms, it may be necessary to lobby hard for the inclusion of these questions. The desire of many national programmes and individual donors to measure service use and evaluate interventions can be used to encourage inclusion of these questions, since the data they generate can often also be used for that purpose.

Multiplier methods based on treatment services tend to be most useful at a local level. Since multipliers may vary significantly from place to place, national-level estimates would do well to aggregate as many local estimates as possible.

Men who have sex with men

This group is very much neglected in many regions, including Asia, Africa and parts of Latin America. This is reflected in the paucity of data sources available. For males who sell sex to men, including for transsexuals, the same basic principles and access points apply as apply for female sex workers. After all, if this population were truly hidden, they would get no clients. However for men who engage in same-sex behaviour without being paid, it is a different story. These men are harder to identify, and in most settings the size of this population is exceptionally hard to estimate.

In most countries, at least some portion of the MSM population is openly "gay", and can be found in bars, cruising spots etc. Theoretically, then, a census of these locations could yield a benchmark number to which a multiplier could be applied, if some way could be found of accessing a wider MSM population that is not openly gay (for example if an anonymous survey could be conducted over the internet). For example, only 15 percent of respondents to an internet MSM survey may report going to bars and cruising spots, so the numbers of individuals counted in those locations could be inflated by 85 percent to yield a population estimate. In practice, however, finding a way to conduct a survey in the wider population at risk in order to generate a multiplier is likely to be extremely difficult, since it is precisely that "hidden" part of the population that cannot easily be reached.

In situations where no other sources of information are available, the only option may be to include questions about same-sex behaviour in an existing household survey system. While the prevalence of homosexual activity is likely to be far higher than the prevalence of injecting drug use (so that the sample size issue should not be so important) it is to be expected that same-sex behaviour would be seriously under-reported in household surveys. For that reason, any estimates based on a general population survey would have to be considered as a bare minimum, and as a starting point for improved estimation techniques.

No method is perfect

In choosing methods, we must be aware that each of them has its limitations. We will never get an accurate count of the size of any specific sub-population. Even if we do get a good count, the dynamics of behaviour and the HIV epidemic itself mean the number will not be valid for long. However with a bit of effort and with a lot more diligent use of existing data sources, we can improve on existing estimates in most cases. With simple methods and existing data, we can ensure that the wild ranges sometimes seen in print (between 25,000 and three million drug injectors nationwide) are a thing of the past.

6 TRYING IT OUT: SOME EXERCISES IN POPULATION SIZE ESTIMATION

This section of the document contains exercises that were used in the working meeting in Bali. Copies of the exercises are provided on a diskette for those who wish easily to reproduce and adapt them for training purposes.

The exercises were devised by meeting organisers, and are based on real situations. They attempt to reflect the limitations of data availability encountered in our work. When using these exercises in training, it is important for participants to understand that there are no "correct" answers. Several approaches are possible for each of the exercises. Participants must take a critical look at the available information, and must make judgements about the relevance and reliability of different pieces of information. Having reviewed the data in this way, they must then decide which estimation method to use.

The important thing in these exercises, as in real estimation procedures, is that participants are able to describe clearly the strengths and limitations of their estimates. They should be able to explain why they chose a certain method, must be able to justify any adjustment procedures, and must be able to give an idea of their level of confidence about the resulting figures.

Exercise 1: Estimating the number of sex workers and clients

Ulutani is a country with an HIV epidemic which has for many years registered low HIV prevalence among sex workers, despite the fact that behavioural surveillance has recorded low levels of condom use in commercial sex.

Peer outreach programmes for sex workers are active in the major cities, but they do not appear to have increased condom use. HIV surveillance systems have over the last two years begun to register a sharp rise in HIV infection.

Ulutani has no central budget for HIV prevention, in part because conservative and religious forces, who are politically powerful, deny that commercial sex is a major problem in the country. The Ministry of Public Health would like to make a national estimate for men and women at risk through commercial sex, in order to lobby for funding and political commitment for HIV prevention in this group.

In addition, the ministry would like to set up a pilot 100 percent condom use programme in the port city of Gulagula, in order to demonstrate that successful prevention is possible.

Use the relevant parts of the data provided to:

- Make a national estimate of the number of sex workers
- Make a national estimate of the number of clients at risk for HIV
- Estimate how many STI screening kits the ministry would need to buy each month if it wanted to achieve 80 percent coverage for screening and treatment among sex workers on a monthly basis in Gulagula chosen for the pilot programme.
- Present these estimates to a cabinet meeting and be prepared to justify your results.

The data available to you are the following:

- National population numbers, from 2000 census data (see spreadsheet Ulutani)
- Population numbers for the district of Gulagula, from 2000 census data (see spreadsheet Gulagula)
- National register of female sex workers and pimps, from 1998 Department of Welfare listing (see spreadsheet Sex worker register)
- Register of female and transvestite sex workers and pimps for Gulagula district, from 1998 Department of Welfare listing (see spreadsheet Sex worker register)
- The results of a national household survey of married men and women aged 15-49, carried out in 1998, which report the following indicators:
    - 70 percent of men had sex before marriage
    - 6 percent of women had sex before marriage
    - 41 percent of men and 15 percent of women have ever used a condom
    - 12 percent of men in urban areas and 7 percent in rural areas have visited a sex worker last year
    - 15 percent of men in urban areas and 10 percent in rural areas had sex with a non-commercial, non-marital partner last year
    - 2 percent of women had sex with a non-commercial, non-marital partner last year

- The results of the latest round of behavioural surveillance (2000) among 1,600 truck drivers and seafarers, carried out in three cities: Tanaskik, Gulagula and Wapati, which reports the following indicators:
    - 65 percent of the sample are married
    - 39 percent of married men and 61 percent of unmarried men had sex with a sex worker last year
    - 20 percent of men had sex with a non-commercial, non-marital partner last year
    - The mean number of cities in which the men had sex last year was 3.6
    - 19 percent of men who had commercial sex used a condom at last sex with a sex worker
    - 7 percent of men who had commercial sex reported using always using a condom in commercial sex
    - The mean number of times men went to a sex workers last year was 10.2
    - There was no significant difference between cities when data are controlled for age and education
- The results of the latest round of behavioural surveillance (2000) among 1500 female sex workers, carried out in three cities: Tanaskik, Gulagula and Wapati, which reports the following indicators:
    - 48 percent of thee sex workers were in brothels, 52 percent were street, nightclub or karaoke based
    - 73 percent had ever been married
    - the mean time working as a sex worker was 21 months
    - the mean time working in this location was 18 months
    - the mean number of days worked per week was 6
    - the mean number of clients per week was 11
    - 41 percent used a condom at last sex with a client, with no significant difference by place of sex work
    - 12 percent used a condom with every client last week, with no significant difference by place of sex work
    - 40 percent reported having sex with a boyfriend last month, half of them accepted payment from this boyfriend at last sex, and 17 percent used a condom
    - 6 percent reported discussing HIV with a government employee or NGO worker
- The results of the national serosurveillance system for 2001 giving population-based HIV prevalence among brothel-based sex workers in all provinces (see spreadsheet: HIV prevalence)
- A single study of HIV in 800 seafarers in the country's largest port, Wapati, which measured prevalence in 1999 at 2 percent
- A copy of a mapping of sex work carried out in Gulagula by an NGO applying for a grant from a World Bank fund for HIV-prevention. The report contains the following statements:
    - Experts estimate the number of sex workers in Ulutani at up to 1.4 million
    - Gulagula is one of the HIV hot-spots for Ulutani
    - There are over 100 brothels in Ulutani; 15 of them are described in detail in this mapping report
    - The average number of women working in these brothels is 117
    - The women work in appalling conditions, and some are forced to take up to eight clients a night. They work all year round, and only have two weeks off for religious holidays
    - Twenty percent of the clients come from the neighbouring country of Denalia

- There is a thriving sex industry outside the brothels: around twice as many sex workers work outside the brothels as in them
- According to a recent WHO study, only 10 percent of brothel based sex workers in Gulagula and far fewer non-brothel based sex workers receive regular screening and treatment for HIV
- Condom use is low, so many of the sex workers are at risk for pregnancy and of passing HIV on to their future infants

Population by age group, urban/rural and sex, Ulutani, National

Age group	Urban			Rural			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
00-04	4,151,806	4,019,919	8,171,725	6,143,895	5,986,756	12,130,651	10,295,701	10,006,675	20,302,376
05-09	4,034,352	3,928,454	7,962,806	6,399,513	6,131,772	12,531,285	10,433,865	10,060,226	20,494,091
10-14	4,017,276	3,934,721	7,951,997	6,443,632	6,058,103	12,501,735	10,460,908	9,992,824	20,453,732
15-19	4,656,022	4,845,775	9,501,797	5,993,326	5,654,394	11,647,720	10,649,348	10,500,169	21,149,517
20-24	4,528,914	4,865,899	9,394,813	4,708,550	5,154,738	9,863,288	9,237,464	10,020,637	19,258,101
25-29	4,321,110	4,399,377	8,720,487	4,809,394	5,111,056	9,920,450	9,130,504	9,510,433	18,640,937
30-34	3,795,282	3,703,739	7,499,021	4,409,020	4,491,679	8,900,699	8,204,302	8,195,418	16,399,720
35-39	3,230,371	3,200,628	6,430,999	4,202,469	4,270,758	8,473,227	7,432,840	7,471,386	14,904,226
40-44	2,767,298	2,554,777	5,322,075	3,666,140	3,479,633	7,145,773	6,433,438	6,034,410	12,467,848
45-49	2,137,790	1,877,742	4,015,532	2,949,462	2,691,011	5,640,473	5,087,252	4,568,753	9,656,005
50-54	1,500,733	1,404,653	2,905,386	2,290,452	2,189,130	4,479,582	3,791,185	3,593,783	7,384,968
55-59	1,145,432	1,089,642	2,235,074	1,737,794	1,705,796	3,443,590	2,883,226	2,795,438	5,678,664
60-64	949,193	1,002,045	1,951,238	1,647,883	1,721,898	3,369,781	2,597,076	2,723,943	5,321,019
65-69	600,348	710,457	1,310,805	1,065,843	1,188,278	2,254,121	1,666,191	1,898,735	3,564,926
70-74	479,243	537,061	1,016,304	888,947	931,786	1,820,733	1,368,190	1,468,847	2,837,037
75 +	442,204	543,932	986,136	815,322	915,527	1,730,849	1,257,526	1,459,459	2,716,985
T.T	2,197	2,235	4,432	3,749	3,666	7,415	5,946	5,901	11,847
Total	42,759,571	42,621,056	85,380,627	58,175,391	57,685,981	115,861,372	100,934,962	100,307,037	201,241,999

Number of sex workers registered with the department of social affairs

NATIONAL	FSW	Pimps
North Hagara	1753	98
Heliacra	10593	1246
Slending	1888	204
Pascora	375	
Rakutaru	3921	
Megataru	2484	180
Singauru	184	20
Malabati	6893	73
Batiuku	1337	97
Bantiki	788	
Gulugulu	890	79
Kitibari	897	98
Kitiubu	410	
Nontrakka	1653	150
Suskalu	201	
Isibira	1513	
Isibantu	1753	216
South Nando	711	65
North Nando	180	15
Xicati	2024	197
Quagarema	412	45
Tulibasa	1922	
Kalybantu	981	

GULUGULU	FSW	Pimps	Transvestites
Missy's		1	7
Lili's	92	3	
Pussy Cat	216	16	
Pink Panther	283	18	6
Hot Lips	88	15	
Devils	15	8	
Showgirls		4	
Sweetness		1	
Chaps		6	
Saturn	196	18	

No data is available for the remaining 9 provinces in the country.
These 9 provinces account between them for 10 percent of the national population.

Number of sex workers tested, and HIV prevalence

NATIONAL	FSW	HIV prevalence
North Hagara	75	17.5
Heliacra	250	1.7
Slending	126	1.4
Pascora	20	0
Rakutaru	150	3.8
Megataru	150	2.4
Singauru	29	23.2
Malabati	245	4.8
Batiuku	120	1.1
Bantiki	0	
Gulugulu	298	8.3
Kitibari	119	7.4
Kitiubu	30	0
Nontrakka	176	0.3
Suskalu	45	15.2
Isibira	133	4.2
Isibantu	123	3.2
South Nando	64	0.8
North Nando	37	0.1
Xicati	268	2.4
Quagarema	42	4.8
Tulibasa	244	2.9
Kalybantu	157	6.2

GULUGULU	FSW	Pimps
Missy's		
Lili's	92	
Pussy Cat	216	
Pink Panther	283	7.9
Hot Lips	88	
Devils	15	15.2
Showgirls		
Sweetness		
Chaps		
Saturn	196	

Possible solutions to Exercise 1

Make a national estimate of the number of sex workers

At least two major approaches to this question are possible with the available data - beginning with the sex worker registers, and beginning with an estimate of the number of clients and deriving sex worker estimates from that. It is worth trying both. If they yield roughly similar estimates, the confidence in the estimate is greatly increased. Other, less robust approaches, are also available.

Approach one: accept published estimates

Why bother with making a new estimate? It has already been done by an NGO, which states that "experts estimate" there are 1.4 million sex workers in Ulutani.

The problem with this approach is that we have no way of knowing what this estimate is based on. We do know that it is published by an NGO that is seeking funding for work with sex workers, and therefore has an interest in having people believe the problem is severe. There is a high likelihood that this estimate overstates the case.

Approach two: start with register of sex workers

Number of sex workers registered in 22 provinces: **43,763**

Completeness of registry

We know the registry is not complete, because in one district where we have more complete information, Gulagula, we can see that the social welfare department lists individual brothels that are not included in the count. (We know this because they give numbers of pimps for these brothels, but not numbers of workers). The number of pimps gives us an idea of the rough size of the brothel: We can make an estimate of the number of women in these brothels by calculating the average number of workers per pimp in the other brothels. There are on average 11.4 sex workers per pimp, and since there are 11 pimps listed in brothels with no FSW lists, that gives on average another 126 sex workers, bringing the total for Gulagula to 1,016. At a minimum, then, we could estimate that social welfare registers miss around seven percent of sex workers in brothels. In this case, we should multiply the national estimates by 1.14 to get an adjusted number.

However we do have another source of information in this case, and that is the mapping of Gulagula brothels by the NGO. While we have a right to be sceptical about their general claims (e.g. "there are over 100 brothels in Gulagula"), there is no reason not to accept the parts of the work that are carefully documented. In this case, the report documents 15 specific brothels, with an average number of workers of 105. These means that the social welfare lists have omitted a third of the brothels entirely. If we use the number of sex workers documented by the NGO (15 x 105) as our benchmark, we see that the social welfare registry lists just over half of the brothel-based workers in Gulagula. This gives us a multiplier of 1.77 for the number of brothel-based sex workers listed by social welfare in Ulutani. In other words, the new estimate for brothel-based sex workers in 22 provinces would be:

43,763 x 1.77 **77,446**

Sex workers outside brothels

The registry does not include sex workers who do not work in brothels. The NGO report claims that in Gulagula twice as many sex workers work outside brothels as in them, but there is no detailed mapping or count.

In BSS in three cities, based on careful mapping of the sex industry, 48 percent of sex workers were in brothels, while 58 percent were outside brothels. So if we want to take into account the non brothel-based sex workers, we need to multiply the numbers in our estimates of brothel-based sex workers by 1.08, and add that to the existing estimate.

(77,446 x 1.08) + 77,446 **161,345**

Missing provinces

Data are available for 22 out of 31 provinces, accounting for 90 percent of the national population. You may consider adjusting the estimate to reflect provinces for which no data are available. One way to do this would simply be to increase the estimate by 10 percent, the proportion of the population which lives in the missing provinces. But these provinces are sparsely populated and have no major urban areas, so it is unlikely that the commercial sex industry exists here in the same proportion as in other, more densely populated parts of the country. An adjustment of a lower proportion, for example five percent, may be more appropriate.

161,345 + (161,345 x 0.05) **169,413**

Estimates should always be rounded to avoid the appearance of greater accuracy than is warranted by the methods. So using multiplier methods based on data from a registry of sex workers together with local mapping exercises, we get

Total estimate for female sex workers in Ulutani: 170,000

Approach 2: start with the client population

This approach involves answering the second question in the exercise before we answer the first.

How many men are clients of sex workers in Ulutani? We can use data from a national household survey, and adjust it as necessary. The key bit of information from the household survey is that 12 percent of men in urban areas and 7 percent in rural areas have visited a sex worker last year. The easiest thing to do would simply be to apply these numbers to men in the urban and rural populations, respectively. Looking at the population aged 15-49 on the Ulutani spreadsheet, that would give us:

(25,436,787 x 0.12) + (30,738,361 x .07)

5,204,100

However the survey was only among married men, and we know from the BSS data that unmarried men are more likely to buy sex than married men. Some 39 percent of married men and 61 percent of unmarried men had sex with a sex worker in the last year, according to the behavioural surveillance in groups of men chosen because their occupations made them highly likely to be clients of sex workers. Clearly we would not expect the *level* of sexual activity to be the same in the general population as in these “higher risk” groups, but it would be reasonable to expect that the *ratio* of commercial sexual activity between married and unmarried men would be similar in all groups of men. So for unmarried men, we would expect to multiply the rates of commercial sex recoded in the general population among married men by 61/39, or 1.56. That would mean that 11 percent of unmarried men in rural areas and 19 percent in urban areas had had commercial sex in the last year.

Unfortunately, we do not have the exact figures for married and unmarried men in the population, although if we took the trouble to call the central bureau of statistics, we could probably find out this information with minimum effort. From the BSS we do, however, have an idea of what proportion of 15-49 year-olds are married: in the sample of 1600 men in mobile occupations it is 65 percent. We might expect men in mobile populations to be less likely to be married than men in more sedentary populations, but the figure is still quite high, so we can confidently expect it to provide a minimum estimate for the proportion of all men in that age group in the general population who are married. To take the easy way out, then, we can simply apportion the population figures we have to married and unmarried in the ratio 65:45. Now we have numbers of men married and unmarried by urban and rural areas. We also have the proportion of married and unmarried men who bought sex last year, by urban and rural areas. All we have to do is apply the proportions who bought sex in each group to the population figures for that group, and we get our estimate of the total number of men in Ulutani who bought sex last year.

(16,533,912 x 0.12) + (8,902,875 x 0.19) + (19,979,935 x 0.07) + (10,758,426 x 0.11) = **6,231,576**

The total number of clients of sex workers in Ulutani is estimated at 6.2 million.

To use this to calculate the total number of sex workers, we need to know how often these men had sex with sex workers. That gives us a total number of commercial sex acts. In the BSS in three cities, men who were clients of sex workers reported an average of 10.2 visits to a sex worker each year. So the total number of sex acts is:

6,231,576 x 10.2 **63,562,073**

Because it takes two to have sex, the number of sex acts reported by clients and sex workers should match. So if we can figure out how many times sex workers have sex with clients in a year, we can divide the total sex acts calculated from client data by that number and get an estimate for the number of sex workers.

From the BSS in sex workers, we know that sex workers report a mean of 11 clients a week. From the NGO report, we know that sex workers work an average of 50 weeks a year. So the average number of clients served by one sex worker in one years is:

11 x 50 **550**

So the total number of sex workers should be the total number of sex acts divided by the number of sex acts per sex worker

63,562,073 / 550 **115,567**

Total estimate for number of female sex workers in Ulutani: 115,000

The second method gives a lower number. Factors that contribute to this may include under-reporting of commercial sexual activity by men participating in the household study, and failure to capture parts of the client population in the household study. This latter would be true if many clients were on the road, at sea, in barracks or other institutions at the time of the survey. A little more research could probably add this information, and future estimates could be adjusted accordingly.

But the critical point about these estimates, which could be presented as a range, is that two very different methods have given us estimates in the same order of magnitude. Certainly they are a far cry from the "up to 1.4 million" being claimed by some NGOs. The similarity of the estimates using different methods greatly increases our confidence that they are representative of the actual situation.

Make a national estimate of the number of clients at risk for HIV

We have already made an estimate of the numbers of clients: 6.2 million. Clients who always use condoms are not at risk for HIV. So if we believe the 7 percent of men who report always using condoms in commercial sex, we can subtract them from those "at risk". The remaining 93 percent are exposing themselves to the risk of HIV and STIs through their commercial sexual activity.

6,231,576 x 0.93 **5,795,366**

Estimated number of men at risk for HIV in commercial sex 5.8 million

Estimate how many screening kits are needed in Gulagula

We have already made an estimate of the total number of brothel-based sex workers in Gulagula. If we inflate for sex workers based outside brothels (as we did at a national level) we get the following:

1,575 + (1,575 x 1.08) **3,281**

Achieving 80 percent coverage would mean reaching

3,281 x 0.8 2,625

When estimating for service provision, it is wise to adjust upwards rather than downwards.

Estimated number of test kits needed for pilot screening programme in Gulagula, per month: 2,800

Exercise 2: Estimating the number of MSM in a cruising area

The city of Zaluso has a population of 12 million. There are several locations were men who have sex with men (MSM) gather to meet potential sexual partners or have sexual encounters. One of the most frequented locations is an island in the park next to the National Stadium. The National Stadium is located in an area with about 1.5 million population and several Universities. Many of the inhabitants are young single male students. There are no other MSM meeting venues in the direct environment. Hundreds of men visit the island every day, but their attendance varies by day and time of the day. Some of the attendees are male sex workers, others are clients, but most are seeking sex without pay. Some men move around in these categories, depending on the situation. The total number of men and men in different categories visiting the island is unknown. It is assumed that a substantial number of men attending the island are infected with HIV.

Fratelli Gay, a local NGO, has received a small amount of money from the local government to set up a clinic near the National Stadium with health and outreach services for MSM. To establish the size of the clinic, the number of doctors, nurses and outreach workers as well as the supplies needed during the first year, Fratelli would like to have an estimate of the total number of potential MSM clients in the area and the number that would like to attend their clinic. They think that with the number in hand, they can convince the government to give them more money. One of their volunteers, a marine biologist, comes up with the idea to do a capture recapture study to estimate the total number of MSM in the area and to ask them whether they are going to use Fratelli services.

Fratelli volunteers come up with the following plan to get the numbers they need. They divide the day in four periods of 6 hours and for one week at the bridge to the island in the park they attempt to interview every third single male. They repeat this procedure after 4 weeks.

In the first sample Fratelli is able to approach 633 men, with a response rate of 76%. In the second sample, Fratelli approaches 587 men, with a response rate of 63%. All men are provided with small media outreach materials, condoms and lubricants. Men in the second sample were asked whether they ever had received condoms and lubricants when entering the MSM area in the park. Ninety-seven said they had. Thirty percent of the total number of men approached said they would seek services at the Fratelli Clinic.

Use the relevant parts of the information provided to:

- Make an estimate of the number of MSM in the area that Fratelli plans to cover, with a 95 percent confidence interval
- Make an estimate of the likely size of the client population for Fratelli
- Discuss the strengths and limitations of the estimates

Possible solutions to Exercise 2

Estimate the number of MSM in the area covered by Fratelli

Fratelli workers have attempted a capture-recapture method to estimate the overall number of MSM who hang out near the National Stadium. The most straightforward solution is to take the numbers at face value, and put them into the normal capture-recapture formula, which, to remind ourselves, is:

N = (number in first capture x number in second capture)/ number in both captures.

We first have to calculate how many men were actually included in the samples, by multiplying the number approached by the response rates (633 x 0.76, and 587 x 0.63)

This gives us

(481 x 370) / 97 **1,834**

Estimate of number of MSM in National Stadium area: between 1,500 and 2,100

Estimate the likely client population for Fratelli

Since 30 percent of those who were recruited said they would use the services, the easiest estimate would simply be to assume that 30 percent of the total number of MSM estimated would use Fratelli's services.

(1,500 x 0.3) and (2,100x 0.3) **450 - 630**

Estimate of the likely client population for the Fratelli clinic: between 450 and 650 clients

95% confidence intervals can be constructed around the resulting number, using the formula

95% CI = N +/- 1.96 $\sqrt{\text{Var (N)}}$, where Var (N) is calculated as follows:

Var (N) = ((# in first capture x # in second capture)(# in first capture - # recaptures) (# in second capture - # recaptures)) / (# recaptures)^3

Var: ((481 x 370) x (481 - 97) x (370 - 97)) / 97^3

95% CI = 1834 +/- 1.96 x $\sqrt{20425}$

95% CI = **1554 - 2114**

Discuss the strengths and weaknesses of these estimates

For capture - recapture estimates to be valid, certain conditions need to be met.

- The two samples must be independent
- The population must be a closed population
- Members of the population must have an equal probability of being "caught"
- Recaptures must be correctly identified

In addition, there is another condition so basic that it is rarely listed: the people "captured" have to belong to the population of interest. This is the weakest point of the above estimates. Fratelli conducted this exercise in a public place which we know has a large population of young single men, many of whom may not have sex with other men. Fratelli did not actually check that the people they were approaching were MSM. A quarter of men approached at the first round and a third at the second round refused to accept the tokens, which may suggest that a high proportion of those who were not MSM did not participate in the "capture", but we cannot be sure of this. It is possible that a significant number of men included in the exercise were not, in fact, MSM.

Are the two samples independent?

The two samples were taken at the same location in different weeks, and both were evenly distributed throughout the week. There is no reason we know of to believe the samples were not independent.

Is the population closed?

We do not have enough information to assert that the population is closed. A month is a relatively long interval, and many MSM may have moved in or out of the area in that time.

Do members have an equal probability of being "caught"?

No. Men who visit the island more frequently will have a greater chance of being caught. In addition, there may be MSM in the area who would be potential clients for Fratelli but who do not visit the island at all and therefore who have no chance of being included in estimates made using this method.

Are recaptures correctly identified?

Almost certainly not. To identify recaptures, people are asked whether they have ever received condoms and lubricants when entering the MSM area in the park. Since the data collection period for each capture takes place over a whole week, it is quite likely that frequent visitors to the park (who also have a higher than average likelihood of being included in the estimates, received condoms and lubricants previously *within the same week*, i.e. within the same recapture period. This would lead to an overstatement of the true number of recaptures, which would in turn lead to an underestimate of the total population.

Another possibility is that men deny having received condoms and lubricants before because they see these items as desirable and are worried that if they admit having received them before, they will not receive them again. This would lead to an understatment of the true number of recaptures and therefore an overestimate of the total population.

Exercise 3: Estimating the number of IDUs for service planning

Public health authorities in Kashmina are concerned about the growing problem if injecting drug use. They have previously secured funding from local government for a needle exchange programme. This has been operating for some time, but questions are being asked by city counsellors about coverage, and new estimates are needed to plan and secure funding for future needs. Data are available from the health department, NGOs, a needle exchange program, behavioural surveillance among drug injectors, and prison/arrest records.

1. There is a methadone treatment service at the central hospital.
   - On an average day they provide methadone to 38 people
   - In an average week, they provide methadone to 60 people
   - In terms of patients, they report the following:
     - 1999: 267
     - 2000: 203
     - 2001: 364
   - They do both detoxification and methadone maintenance
   - For their detoxification program, they report a relapse rate of 90%

2. There are three NGOs who provide services for IDUs including one that does needle exchange. Two of the NGOs provide residential detoxification services. One NGO called FRIENDS has about 20 clients at any given time. They say they have treated 70 clients in the past year. Another NGO called the SAMARITANS has treated 120 clients in the past year. And a third NGO called AIDSACTION has no residential services, but it has been conducting needle exchange and outreach for the past 3 years. The needle exchange tries to provide clients with enough needles to allow them to use a fresh needle at every injection, and has given out a total of 200,000 needles and syringes in the past year.

3. The police report that they have arrested 1700 people on drug possession charges in the past year, and there are 157 people in prison right now who were arrested for drug use. Seven of them are females.

4. BSS data is available for a random sample of 300 IDUs reached at locations where drug users congregate. According to the data from this survey:
   - Injectors say they inject an average of three times a day
   - 5% report receiving needles from outreach workers or at the drop-in centre at AIDSACTION in the past month
   - 12% report they tried to detox at least once in the past 12 months and among those who have tried, 20% report having tried in at least two different treatment centres
   - 40% report that they have been arrested and spent time in prison for drug use in the past year
   - 60% reported that the last time they injected, they used a needle and syringe that someone else had used before them

Use the relevant parts of these data to:

- Estimate the number of IDUs in Kashmina

- Estimate how many needles would be needed to eliminate sharing over a one year period

Possible solutions to Exercise 3

Estimate the number if IDUs in Kashmina

Approach 1: Use the needle exchange and BSS data

The needle exchange has given out 200,000 needles in the last year, or an average of 548 a day. They say they try to give out one needle per injection, and we know from the BSS data that IDUs in Kashmina inject on average 3 times a day. So we can estimate the clients of the needle exchange as

548 / 3 **183**

From the BSS data, we know that 5 percent of a random sample of 300 IDUs have recently used the needle exchange. This gives us a multiplier of 20.

183 x 20 **3,653**

Estimated number of IDUs in Kashmina : 3,700

Approach 2: Use the detox and BSS data

The central hospital has treated 364 people in the last year, and we know that at least 60 of those are on methadone maintenance. At a maximum, then, 304 people went through detox at the central hospital in the last year. The NGOs doing detox report 70 and 120 clients in the past year respectively, for a total of

304 + 70 + 120 **494**

But we know from BSS that 20 percent of those who have been in detox in the last year have been in more than one place, so we need to adjust the total for that double counting:

494 x 0.8 **395**

We also know from BSS that 12 percent of IDUs have been in detox in the past year. This gives us a multiplier of 100/12, or 8.3.

395 x 8.3 **3,293**

Estimated number of IDUs in Kashmina : 3,300

Approach 3: Use the prison and BSS data

The police have arrested 1,700 people on drug charges in the past year, and 157 are still in prison. In the BSS, 40 percent of a random selection of IDUs contacted on the streets say thy have been arrested and spent time in prison for drug use in the past year.

We can only use this information to make an estimate if we know how many of the 1,700 people arrested by the police on drug charges were actually imprisoned. It is quite possible that a high proportion of those arrested did not actually do any prison time, either because they bribed their way out or because the arrest did not lead to a conviction. If we do assume all those whose arrests were registered actually did prison time, then we could try to make an estimate.

We know that 157 of the people arrested in the last year are still in jail. This means that they are not on the streets, and could not have been included in the respondents who because the arrest did not lead to a conviction. If we do assume all those whose arrests were registered actually did prison time, then we could try to make an estimate.

We know that 157 of the people arrested in the last year are still in jail. This means that they are not on the streets, and could not have been included in the respondents who said they did jail time. They must therefore be removed from the number of people to whom we apply the multiplier (which is 100/40, or 2.5)

(1,700 -157) x 2.5 **3,858**

Estimated number of IDUs in Kashmina : 3,900

The three approaches give us remarkably similar figures. The mean of our three estimates is 3,600. Provided the BSS was indeed in a random selection of Kashmina's IDUs, we can say with a great deal of confidence that there are between 3,000 and 4,000 IDUs in the city.

Estimate how many needles would be needed to eliminate sharing over a one year period

Some 60 percent of drug injectors said in the BSS that they used a needle previously used by someone else at last injection. If we take the mean estimate of IDUs, that means needles were shared by

3,600 x 0.6 **2,161**

injectors. If each of these people injects on average 3 time a day, then their annual needle needs would be

2,161 x 365 x 3 **2,366,033**

Of course it in necessary to add this to the number of needles already being provided by the needle exchange (since the needle exchange must also continue to serve its existing clients, who we hope are not among the 60 percent that reported sharing at last injection). So the *minimum* estimate for the number of needles needed would be

2,366,033 + 200,000 **2,566,033**

Estimated minimum number of needles per year needed to eliminate sharing: 2.6 million.

This estimate relies on people having accurately reported their sharing at last injection, and also on the assumption that the proportion who did not share last time is broadly representative of the proportion who never share. In practice, needle sharing tends to be under-reported, especially in communities where harm reduction and treatment programmes are in place. It would be safer to based an estimate on the total number of needles needed by all injectors. That would simply be:

3,600 x 365 x 3 **3,942,000**

Estimated number of needles per year needed to ensure no sharing is necessary: 4 million

7 RESOURCES AND REFERENCES

Archibald, Chris and Gaya Jayaraman. Purposes of population size estimation: the policy perspective. Presentation prepared for the workshop on the estimation of the size of populations at risk for HIV, Bali, April 2002.

Archibald, Chris, Gayatri Jayaraman, Carol Major et al. Estimating the size of hard-to-reach populations: a novel method using HIV testing data compared to other methods. AIDS 2001 15 (suppl 3): S41–S48.

Black, J, DG McLarty, D Mtasiwa. Capture – recapture difficult to use in developing countries. BMJ 1994: 308–531 (Letters).

Chatterjee A., M Khaled, S Rahman, and S Sarkar 1998. Drug Injecting and potential for continuing spread of AIDS: A baseline assessment in Dhaka city. CARE Bangladesh. Dhaka, Bangladesh.

Cox, S and M Shipley. Counting the uncatchable? An epidemiological method for counting drug misusers. Social Psychiatry and Psychiatric Epidemiology 1997 32:19-23.

Frank, Ove and Tom Snijders. Estimating the size of hidden populations using snowball sampling. Journal of Official Statistics 1994 10 (1): 53–67.

Hay, G. (2000). "Capture-recapture estimates of drug misuse in urban and non-urban settings in the north east of Scotland." Addiction 95(12): 1795–803.

Heckathorn, Douglas. Respondent-driven sampling: a new approach to the study of hidden populations. Social Problems 1997 44 (2): 174–199.

Indonesia (2003). National Estimates HIV mong Adults in Indonesia, 2002. Report of a national process to estimate the size of populations at risk for HIV. Jakarta, Directorate General of Communicable of Communicable Disease Control and Environmental Health.

Larson, Ann, Adele Stevens, Grant Wardlaw. Indirect estimates of 'hidden' populations: capture recapture methods to estimate the numbers of heroin users in the Australian Capital Territory. Social Science and Medicine 1994. 39 (6) 823–831.

Mastro, Timothy, Dwip Kitayaporn, Bruce Weniger et al. Estimating the number of HIV-infected injection drug users in Bangkok: a capture-recapture method. American Journal of Public Health 1994, (84): 1094–1099.

Neugebauer, Richard and Janet Wittes. Annotation: Voluntary and involuntary capture-recapture samples – problems in the estimation of hidden and elusive populations. American Journal of Public Health 1994, (84): 1068–1069.

Sarkar, S, F Durnadin, AA Quader et al. Estimation of number of street based female commercial sex workers (CSWs) and their HIV situation in the city of Dhaka in Bangladesh. Proceedings of the 4th International Congress on AIDS in Asia and the Pacific, Manila, 1997 A(P) 071, p. 184.

Stephen, Craig. Capture recapture methods in epidemiological studies. Infectious control and hospital epidemiology 1996, 17: 262–266.

UNODC .In-direct methods for estimating the size of the drug problem. GAP Toolkit). Module 3: Prevalence Estimation. Pre-publication version November 2002. Retrieved July 27, 2003. <http://www.undcp.org/pdf/gap_toolkit_module3_estimation.pdf>.

Weir, Sharon, David Wilson, Philip Smith et al. Assessment of a capture-recapture method for estimating the size of the female sex worker population in Bulaweyo, Zimbabwe. February 2003. Retrieved July 29, 2003, from MEASURE Evaluation Web site. <http://www.cpc.unc.edu/measure/publications/workingpapers/w0363.pdf>.

Appendix 1:
Agenda and guidance notes for three day workshop on population size estimation

Agenda

Day 1	
09.00 - 09.30	Welcome, participant introductions
09.30 - 10.00	Aims of meeting and overview of agenda
10.00 - 10.30	Purposes of population size estimation
10.30 - 10.50	Coffee
10.50 - 12.00	Country perspectives on purposes of estimation (See Guidance Notes 1)
12.30 - 14.00	Lunch
14.00 - 14.20	Programme perspective on estimation
14.20 - 15.10	Challenges and issues in size estimation Case definitions Problems (mobility, dynamics etc) Ethics Relationship with other data systems Discussion
15.10 - 15.25	Coffee
15.25 - 16.30	Country experiences with of estimation (See Guidance Notes 2)
	Evening: Optional field visit to IDU outreach programme
Day 2	
09.00 - 09.20	Overview of estimation methods
09.25 - 09.40	Survey methods
09.45 - 10.00	Enumeration Break into three groups
10.05 - 11.05	Exercise 1 based on survey and enumeration data
11.05 - 11.20	Coffee
11.20 - 11.50	Presentation and defence of exercise 1 by group 1, discussion
11.50 - 12.15	Capture recapture
12.15 - 13.15	Lunch
13.15 - 14.00	Exercise 2 based on capture recapture
14.00 - 14.30	Presentation and defence of exercise 2 by group 2, discussion
14.30 - 15.10	Multiplication methods Multiple methods
15.10 - 15.25	Coffee
15.25 - 16.25	Exercise 3 based on multiplication
16.25 - 16.55	Presentation and defence of exercise 3 by group 3, discussion
Day 3	
09.00 - 11.00	Group work Choosing the right estimation method in the circumstances (see Guidance Notes 3)
11.00 - 11.20	Coffee
11.20 - 12.30	Presentation of group work, Discussion
12.30 - 13.30	Lunch
13.30 - 14.15	Country work: Data review and action plan
14.15 - 15.00	Presentation of country action plans
15.00 - 15.15	Coffee
15.15 - 16.00	Summary and next steps

Guidance Notes 1

Country perspectives on population size estimation

Please tell us your thoughts on population size estimation needs in your country.

- What are the major "hard-to-reach" populations in the context of the HIV epidemic in your country?
- Do you need population size estimates for advocacy? Programme planning? Monitoring and evaluation? Reporting? What else? Do these needs differ for different population groups?
- What are the major challenges and opportunities? Does the situation make it easier or harder to make good estimates?

Guidance Notes 2

Country experiences with population size estimation

If your country has done any work in estimating the size of hard-to-reach populations, please take a few minutes to share information for the rest of the group.

- What groups have you tried to estimate the size of?
- In what geographical areas (national, provincial, city etc)?
- What relation did this have with other data collection efforts? HIV and behavioural surveillance? Drug information systems?
- What difficulties did you encounter?
- Lessons learned

Please DO NOT give a detailed description of methods, but discuss methods generally and tell us a bit about your experiences to date.

Guidance notes 3

Group work on choosing appropriate estimation methods for different sub-populations

The purpose of this session is to discuss in greater depth the methods most likely to be appropriate in the general context of information available in the countries in which we work. We recognise that the situation will be different in different countries; the session is meant to facilitate an exchange of ideas between participants and encourage consideration of the potential for better use of existing data.

The groups will be divided by population group: IDU, MSM and a third group considering female sex workers and clients.

Please consider the following questions in your discussion

- What data sources are most commonly available for this group?
- How could they be adapted to increase their utility for the purposes of estimating population size? What other (existing) data collection mechanisms could be used or adapted to contribute information for better estimates in this group?
- What are the major limitations of each data source (tend to overestimate, tend to underestimate? Only available on a local scale? etc)
- Given the data that are available, which do you think are likely to be the most appropriate estimation methods for this population group?

One SUGGESTED way of reporting back would be to complete a table, reviewing the strengths and limitations of each estimation method for the population group that you are considering. POSSIBLE things to consider would be (please feel free to change these headings to capture your discussions)

METHOD	Data needs	Advantages	Limitations	Comments

Appendix 2:
Checklists for data sources that may be used for estimation

Preparatory work for the population size estimation meeting

Attached you will find three similar tables, asking for information about potential sources of data that might be used to help make estimates of hard-to-reach populations. Please fill out each of these tables to the best of your ability, for three populations. Table 1 is for drug-related information, Table 2 (page 59) is for information relating to sex workers and clients, and Table 3 (page 60)is for information relating to men who have sex with men, including male sex workers and transvestites.

You may have to make a few phone calls to help fill out these tables. For example you may need to call the department of corrections and ask whether the prison system keeps records by reason for arrest, and whether those information could be made available to the health department.

The tables are fairly self-explanatory, but here are some notes to help fill them in:

Data source

These are potential sources of data which can be used, either alone or (usually) in conjunction with other sources of data, to make estimates of at-risk populations. Every country has different data sources, and we have listed some of the most common, but you may have other ideas or know of other sources. Please do add any other sources you can think of to the list.

Does it exist?

Does this source of data exist at all in the country where you work? If you do not know whether it exists or not, please try to find out. If you cannot find out, please mark this section "don't know". Sadly, we expect to see many "no" s in this section.

For what cities/geographical areas?

If the data type does exist, please not the geographical coverage. For example prison records data may be available nationally, where as BSS may be available only for major urban areas. It is important to know this because it affects the scope of the estimates we can hope to make.

Data quality

Please give a judgement of the data quality if you can. For example, you may know that the department of social affairs keeps a register of sex workers, but that it is rarely updated and in any case only covers a small portion of sex workers. In this case the answer would be "incomplete". If you don't know enough about the quality of the data to make a judgement, say "don't know".

Is it available for estimations work?

In some cases, we may know that data are collected, but they may be impossible to access. For example, some countries treat data on drug seizures as a state secret. Of course most of these data types will not be a matter of public record, but please note if you think they could be made available to the health ministry to help with surveillance and population size estimations after negotiation.

How feasible would it be to collect?

If a type of data does NOT exist in your country, please note whether you think it would be feasible to collect it. Would it be possible, for example, to set up a BSS system where there is none, or to systematically collect data from STI clinics serving sex workers?

Can you bring examples?

We will prepare some case studies but we would like discussion to centre as much as possible on what is feasible in the countries attending the meeting. If you can bring examples of any of the types of data listed it would be very very helpful in ensuring that the exercises we undertake are relevant. Again, please do not worry if your table is almost blank or you can not bring much in the way of data. We expect this. But without a realistic picture of the data limitations we cannot provide help in developing methods that are feasible in our situations.

Thank you very much for your cooperation.

Table 1: drug-related data

Data source	Does it exist?	For what cities/areas ?	Data quality	if it exists,is it available for regular estimation work ?	If not, how feasible to set up data collection ?	Can you bring examples to the meeting ?
Behavioural Sentinel Surveys (BSS)						
Other focused surveys						
Household surveys						
HIV surveillance						
Register of addicts						
Treatment centre data						
Police data, by reason for arrest						
Court data, by reason for trial						
Prison data, by reason for conviction						
Hospital data on drug-related admission						
Mortality data on drug-related death						
Health service data on positive drug tests						
Employment agency data on positive drug tests						
Data from needle exchange programmes						
Data from outreach programmes						
Research studies						
Anything alse ?						

Table 2: Data related to sex work

Data source	Does it exist ?	For what cities/areas?	Data quality	if it exists,is it available for regular estimation work ?	If not, how feasible to set up data collection ?	Can you bring examples to the meeting ?
BSS among sex workers						
BSS among clients						
Other focused surveys among sex workers						
Mapping of brothels/ red light districts						
Registry of brothels						
Police data on brothels						
Registry of "short stay" hotels						
Registry of karaoke bars or "entertainment places"						
Other focused surveys among clients						
Household surveys of consumption of commercial sex						
HIV surveillance						
Register of sex workers						
STI clinic data						
Police data, by reason for arrest						
Court data, by reason for trial						
Prison data, by reason for conviction						
Condom sales or distribution data						
Data from outreach programmes						
Research studies						
Anything else?						

Table 3: Data related to MSM/transvestites

Data source	Does it exist ?	For what cities/areas?	Data quality	if it exists,is it available for regular estimation work ?	If not, how feasible to set up data collection ?	Can you bring examples to the meeting ?
BSS among MSM/transvt						
BSS among clients						
Other focused surveys among MSM/transvt						
Mapping of cruising areas						
Registry of bars or "entertainment places"						
Household surveys with questions about same-sex experience						
HIV surveillance						
MSM clinic data						
STI clinic data						
Data from outreach programmes						
Research studies						
Anything else you can think of						

Note

Note

Note

Family Health International
Institute for HIV/AIDS
2101 Wilson Boulevard, Suite 700
Arlington, Virginia 22201 USA
Telephone: 1.703.516.9779
Fax: 1.703. 516.9781
URL: http://www.fhi.org

ISBN 974-91495-0-5