In-direct methods for estimating the size of the drug problem

GAP Toolkit Module 3: Prevalence Estimation

Pre-publication version
November 2002

UNDCP
Vienna International Centre
Austria

United Nations
New York, 2002

The contents of the *GAP Toolkit Module 3: Indirect methods for estimating the size of the drug problem* was produced by the United Nations International Drug Control Programme as part of the activities conducted under the Global Assessment Programme on Drug Abuse (GAP).

For further information visit the GAP website at *www.undcp.org,* email *gap@undcp.org*
or contact: Demand Reduction Section, UNDCP, P.O. Box 500, A-1400 Vienna, Austria.

This document has not been formally edited.

United Nations International Drug Control Programme
Printed in Austria, 2002

Preface

GAP Toolkit Module 3: Estimating Prevalence – indirect methods for estimating the size of the drug problem, has been prepared with by the United Nations International Drug Control Programme as part of the activities of the Global Assessment Programme on Drug Abuse (GAP). The main objective of GAP is to assist countries collect reliable and internationally comparable drug abuse data assists with building capacity at a local level to collect data that can guide demand reduction activities, and also to help improve cross-national, regional and global reporting on drug trends.

The GAP epidemiological has been produced to assist UN member states to develop systems to collect drug information that are culturally appropriate and relevant to their country, and to support existing drug information systems to conform to internationally recognised standards of good practice, and focus on harmonization of drug abuse indicators.

Module 2 of the GAP Toolkit forms one component of a compendium of methodological guides that have been developed to support data collection activities. Other modules currently under development provide support in the following areas: developing an integrated information system, school surveys, data interpretation and management for policy formation, and basic data manipulation using SPSS.

Other GAP activities include provision of technical and financial support to the establishment of drug information systems and support and coordination of global data collection activities. For further information on GAP Toolkit Modules contact gap@undcp.org or visit the GAP website at www.undcp.org.

The philosophy behind the toolkit is to provide a practical and accessible guide to implementing data collection in core areas. The toolkit modules are designed to providing a starting point for the development of specific activities, referring the reader to more detailed information sources on specific issues, rather than being an end resource itself. GAP toolkits are based on principles of data collection that have been agreed upon by an international expert panel and endorsed by Member States of the United Nations. Models presented are based on existing working models that have been found effective, however, a key principle is that approaches have to be adapted to meet local needs and conditions. This module of the toolkit therefore provides specific examples to guide the reader on the nature of how general principles and models can be adapted to specific contexts, and is not intended to reflect the complete range or diversity of current drug information systems or data collection methods.

For further information visit the GAP website at *www.undcp.org,* email *gap@undcp.org*
or contact: Demand Reduction Section, UNDCP, P.O. Box 500, A-1400 Vienna, Austria.

Table of Contents

Acknowledgements

UNDCP would like to acknowledge the support of many national counterparts in piloting and providing feedback on the toolkit, and the support of institutions and individuals in providing examples of data collection forms, mechanisms, and other related material. In particular, thanks go to the Community Epidemiology Work Group, the Pompidou Group of the Council of Europe, the European Monitoring Centre for Drugs and Drug Addition, the SADC Epidemiology Network on Drug Use, the East African Drug Information System and the Caribbean Drug Information Network.

The contents of the *GAP Toolkit Module 3: Estimating Prevalence – indirect methods for estimating the size of the drug problem* was produced with the support of the United Nations International Drug Control Programme as part of the activities conducted under the Global Assessment Programme on Drug Abuse (GAP). The report has been prepared by a technical panel of experts. Particular thanks are due to Colin Taylor who coordinated the project and acted as editor with Mathew Hickman who acted as co-editor, and to Rebecca McKetin who co-ordinated and advised on the final stages.
The Technical Panel advisory members were
Colin Taylor, NAC, London (Co-editor)
Mathew HIckman, IC, London (Co-editor).
Michael Lynskey, NDARC, NSW, Australia
Lucas Wiessing, EMCDDA, Lisbon
Paul Griffiths, Rebecca McKetin, Kamran Niaz, UNDCP
Anindya Chatterjee, UNAIDS, Thailand
with assistance from Mathew Warner-Smith, UNDCP, Southern Africa.

The manual has been written as a collaborative effort by a technical expert group. The technical expert group and its meetings were designed to provide an informal editorial advisory board to the editor The function of the group has been

(i) to advise on the content and structure of the module;

(ii) to suggest potential contributors for the different sections of this module.

The initial meeting constituted a major contribution to the project, formulating both the areas to be covered and the structure of the manual. The co-operative efforts of the technical expert group are recognised as invaluable to the project.

1 General Introduction

Background

Although there are countries that can claim successes in controlling the demand for illicit drugs, abuse throughout the world continues to grow. In particular, illicit drug abuse in some developing countries has increased dramatically. However, knowledge of the scale of illicit drug use is still inadequate, and understanding of the patterns and trends limited.

To provide effective policies to reduce drug abuse, governments need data about when, where and why people use illicit drugs. Patterns of drug use transcend national borders as users in all regions of the world get access to a greater variety of drugs, and as social trends, particularly among young people, spread more rapidly than before via better communications. The globalisation of drug abuse means that demand reduction policies also have to be global, as must the information system on which they rely.

In 1998, Member States of the UN adopted a Political Declaration[1] to eliminate or reduce significantly the supply and demand for illicit drugs by the year 2008. This is the first time that the international community has agreed on such specific drug control objectives. However, the systematic data that is needed to monitor and evaluate progress towards these goals are not yet available. For this reason, the UN General Assembly requested the United Nations International Drug Control Programme (UNDCP) to provide Member States with the assistance necessary to compile comparable data. UNDCP was asked to collect and analyse these data and report them to the UN Commission on Narcotic Drugs. As a response to this need UNDCP launched the Global Assessment Programme on Drug Abuse (GAP). GAP has been designed to:

- support Member States to build the systems necessary for collecting reliable data to inform policy and action
- encourage regional partnerships to share experiences and technical developments
- facilitate a better understanding of global patterns and trends in drug abuse by encouraging the adoption of sound methods to collect comparable data.

These aims reflect the challenge posed in the Guiding Principles of the 1998 Political Declaration, which calls for:

> "*demand reduction programmes should be based on a regular assessment of the nature and magnitude of drug use and drug-related problems in the population… These assessments should be undertaken in a comprehensive, systematic and periodic manner, drawing on results of relevant studies, allowing for geographical considerations and using similar definitions, indicators and procedures to assess the drug situation*."[1]

The main objective of the Global Assessment Programme on Drug Abuse (GAP) is to assist Member States build the capacity to collect internationally comparable drug abuse data and assess the magnitude and patterns of drug abuse at country, regional and global levels. Development of these national and regional information systems should not only assist with building capacity at a local level to collect data that can guide demand reduction activities, but also to improve cross-national, regional and global reporting on drug trends. To support this process, the GAP Toolkit Module 3. *Estimating Prevalence – indirect methods for estimating the size of the drug problem* has been produced to assist UN member states to develop systems to collect drug information that are culturally appropriate and relevant to their country, and to support existing drug information systems to conform to internationally recognised standards of good practice, and focus on harmonization of drug abuse indicators. Standardization of indicators and the wider adoption of sound methods for data collection will result in an enhanced analysis of trends in drug abuse in both the industrialized and developing world. For more information on GAP visit the GAP website at *www.undcp.org*, email *gap@undcp.org* , or contact: Demand Reduction Section, UNDCP, P.O. Box 500, A-1400 Vienna, Austria.

---

[1] Special Session of the General Assembly Devoted to Countering the World Drug Problem Together, 8-10 June 1998.

The Role of Prevalence Estimation

Two key policy questions asked of those collecting information on drug abuse are: how is how many members of the population of a country are using drugs; and is this number changing. Understanding the number of drug abusers is helpful in assessing the likely impact drug abuse will have on society and what levels of responses are likely to be necessary. For example, knowing how many injecting drug abusers there are allows for the calculation of the necessary level of provision of services to reduce behaviours associated with HIV infection and for gauging whether sufficient drug treatment places are available. Understanding something about the dynamics of the drug problem not only allows for the likely impact to be assessed but also to alert policy makers to a worsening situation, or alternatively provide evidence that prevention and other initiatives may be working. In many countries, and especially where drug problems are a relatively new phenomenon, having an estimate of the size of the drug problem is a powerful tool in focussing the minds of both policy makers and the general public on the need for action and resource investment.

In epidemiological terms the two questions relate to prevalence and incidence estimation. Epidemiologists use the term 'incidence' for rates of newly diagnosed cases of a disease or health problem and the term 'prevalence' for the total number of cases, a figure that includes newly and formerly diagnosed current cases. To calculate these rates requires a population base that is defined by geographical location, by time period, and by some other well-defined characteristic such as age, race or gender.

Although the need for information on the scale of the drug problem is clear, in practice these data are extremely hard to generate. Many countries in the world are not able to estimate the number of drug abusers and where estimates do exist they often appear to be of questionable accuracy. There are a number of reasons why countries experience difficulties in this area and why estimating drug use prevalence is more difficult than estimating the extent of many other medical conditions. The use of prohibited psychoactive substances is not simply reducible to a disease model. While concepts of dependence and problem drug use are subject to rigorous diagnostic criteria, these are often difficult to operationalise in social survey work. Furthermore many of those who consume drugs do not fall into these diagnostic categories, but are still of interest to policy makers. The range of different substances abused, the different routes of administration used, the different patterns found in terms of dose and frequency, all complicate matters further. In addition, the illicit nature of drug abuse and the fact that the behaviour is socially stigmatised poses particular challenges for the drug researcher that are not found in most other areas of epidemiology.

Given the range of drug taking behaviours that individual engage in, the first question that has to be resolved in any drug prevalence exercise is what exactly is the target behaviour to be measured. Usually, it is most practical to use time-based consumption measures classified by different drug types. Additional information on injecting drug use is also typically required. One common approach to gathering this information is the general population survey. However, for a number of reasons general population surveys can perform poorly in assessing some types of drug abuse. In particular, the rare but most deleterious patterns off drug abuse, such as heroin or cocaine addiction or drug injection, are often not measured well by household surveys and under-reporting can be a problem. Household surveys are also technically complex, resource intensive undertakings and in many developing countries simply are not practical. An alternative approach is to use indirect estimation techniques to estimate the number of drug abusers in these categories. These approaches are the main subject of this module of the GAP toolkit. This document has been written with the needs of developing and transitional countries in mind. However, it should be noted that these kinds of methods are now widely recognized as an important tool for improving the estimate of the size of the drug problem, even in those countries where large-scale household surveys are also conducted.

2 Introduction to Drug Use Prevalence Assessment

Introduction to Drug Use Prevalence Assessment

This module offers an overview of some of the methods that have been developed to address the difficulties of estimating drug prevalence through standard survey techniques. The methods described are quite general, in that they can be used for estimating prevalence of use of any drug, but the setting in which they are described here is in the context of hard-core drug use, such as estimating illicit opiate use. The central point in this setting motivating these procedures is that illicit opiate use is relatively rare when the number of users is considered as a fraction of the entire population, it is a stigmatised behaviour and its illicit nature implies that there can be no national, full-coverage recording system for it.

Standard statistical survey methods applied to general population surveys usually do not fare well in these circumstances. Techniques such as multiplier/benchmark methods and capture-recapture methods provide designs for prevalence estimation that can be used where standard methods fail, and they are presented here as an alternative to general population surveys for estimating drug use prevalence.

Assessing prevalence and the importance of assessing harm

There is a growing need and demand from policy makers for prevalence estimates to inform and guide policy (see European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). Scientific Monograph Series No. 1. Estimating the prevalence of problem drug use in Europe). In turn this has been associated with a growth in manuals and review papers on methods (EMCDDA 1997,2000; Hickman, Taylor et al.,2002;; IWG DMM 1995; Frischer, Bloor et al., 1993; Hser, Anglin et al., 1992; Reuter, 1993). Where this module differs from others is its focus on developing countries, and in providing a manual that brings together examples and methods applicable to developing countries.

Prevalence estimates – such as estimates of the number of heroin users or injectors in the population – are required for several aspects of policy-making.

- Planning and allocation of resources for control, treatment and prevention of problem drug use and its consequences. To argue for an increase in funding often needs a prevalence estimate as supporting evidence.
- Monitoring key targets of drug policy: e.g. having information on the proportion of problem drug users in treatment or the extent of coverage by any harm reduction activities. One way of measuring these targets is use estimates of the number in contact with services, for example, in relation to prevalence estimates of the totl drug-using population..
- Interpreting key harms associated with drug use. The burden of HIV, HCV, fatal overdose, and drug related crime associated with drug use in the population as a whole is related to the level of risk and other behaviours experienced by problem drug users and the prevalence of problem drug users.

Assuming that policy makers are most interested in behaviours that produce the greatest public health harm, the core problem is to find a method for estimating the number of heroin users, injectors, or crack-cocaine users. This then is the focus of the present manual.

There are some notable and intended omissions from this manual. The intention here is to bring together indirect methods of estimating the levels of problem drug use and its attendant problems. In particular, school surveys, which directly assess the problem as it begins at the outset of people' lives, are excluded. These and other specialist studies of particular sections of the population are dealt with in different modules under the overall GAP project.

Prevalence estimates and the role of general population surveys

Prevalence estimates arising from surveys of the general population - surveys that are usually based on sampling households, using self-reporting of drug taking behaviour - are widely accepted as inadequate for estimating the extent of illicit drug use, particularly of heroin use. Survey methods, in this classical sense, are broadly suitable for a restricted number of questions where biasses and errors in responding and general mis-reporting can be held to a small percentage of the overall prevalence rate. In the drug prevalence field this means questions about, for example, prevalence of lifetime cannabis use**,** tobacco smoking, and perhaps alcohol consumption are viable candidates for general population survey studies. Deriving estimates of prevalence for these behaviours is usually accomplished by adding - piggy-back style - questions into the questionnaires already used in existing general population surveys. These added questions, for example, are often aimed at eliciting one-year period-prevalence rates of nicotine smoking or frequent alcohol consumption.

The inadequacy of standard survey methods in relation to estimating the prevalence of hard-core drug use arises principally from two areas -

> *(i) under-coverage:* hard-core drug users occur in significant numbers outside household units (NIDA, 1994) and there is therefore a failure to cover key drug-taking groups that are not included in survey sampling frames: for example, street-dwellers (the 'permanently' homeless), those arrested and possibly those in residential treatment;
>
> *(ii) under-reporting of use:* drug users may simply decline to answer or answer in the negative, a point demonstrated for example by recent work with simultaneous self-reporting and hair analyses; this problem is exacerbated for more stigmatised behaviours: for example, heroin use is usually thought to be more often under-reported than is marijuana use.

These factors attenuate what is, in relative terms, an already low rate of occurrence of hard-core drug use. The resulting low prevalence figures themselves, for example the British Crime Survey and Psychiatric Morbidity Surveys, and in USA the NHSDA all attest to this. Corrective factors can be applied through ratio estimation techniques, as for instance described in some of the NIDA reports (Wright et al, 1997) and discussed below, but generally effective estimation requires procedures that are more specifically tailored to the drug-use context.

Methods for correcting under-estimation within general population surveys

It has been noted (Wright, 1997) that factors related to drug use, such as arrest information and treatment for drug use are also under-reported or inadequately covered in general population surveys. When there is known national, information on these factors it can be used to correct the survey figures for them, by weighting up the number of respondents who do report arrest or drug treatment to what is known to be the correct figure. Using the national figures as a benchmark in this way to adjust for under-estimation in national survey data analysis is a standard application of statistical 'ratio estimation'.

The relevance for our purposes is that when weighting up the respondents for these drug-related behaviours, in so doing the associated drug abuse numbers will also automatically be weighted upwards as a consequence.

Some caution should be attached to this procedure, since the resultant adjustment to the drug prevalence figures is a perquisite from a method designed to adjust the related factors, not the prevalence figures themselves. It is suggested that the results be considered an improvement upon, rather than a correction to, the original estimates.

Drug use prevalence and other epidemiological methods

What characterises drug user populations then is that they are what are termed hard-to-access populations. They are populations that, as far as official institutions an their records are concerned, are 'hidden' with respect to access; and because of the potential effects of stigma and illegality, they are also 'hidden' with respect to the accuracy in the responses that might be provided. Traditionally, survey sampling in the main depends traditionally on using some form of lists of the general population to provide the sampling frame from which to generate survey members. When we abandon the survey statistician's tools of a general population sampling frame because of its unsuitability we are left with only a few options for assessing prevalence. We briefly consider some of these now.

Area sampling

Not all population surveys though require sampling frames. One exception is area-based sampling in which a region or country is conceptually divided into operational areas, usually equal in size, and a sample is drawn of these areas. To estimate the size of the population overall or the size of a defined sub-set such as problem drug users the researchers must then physically in each area count 'cases' that met the defining criteria. (Alternative formulations of this method use 'line transects' instead of areas.)

In some situations this is perhaps the only method that can be used, and in some circumstances it can be a successful strategy. The counting of cases physically implies that there is an ease with which cases can be identified: clearly such a procedure cannot cope in practice with lengthy or complicated procedures to identify a case that meets the defining criteria. In a survey of 'problem drug users' this is not a trivial exercise. Furthermore, it implies that the social structures in the country are suitable for this approach. One example of such a method are the earlier Pakistan National Surveys of 'drug addicts', in which village mullahs or heads were asked to identify drug addicts in the community.

Specific listed populations

Survey researchers have attempted to address some of the problems of under-coverage that are inherent in calculating drug prevalence by using specialist sampling frames when designing sample surveys, such as records from casualty wards or police arrests. However, in general these are too narrowly focused in their coverage to allow general inferences about problem drug use.

A notable exception here is when a group is itself the focus of particular interest, such as school children. National – and indeed international – surveys such as the ESPAD (European School Survey on Alcohol and other Drugs) series can be regarded as giving reasonable coverage of the target population, especially when provision is made to cover truant pupils.

Note that obvious drug user listings such as clinic treatment records, needle exchange listings and the like, that cannot be used to provide a sampling frame for a prevalence survey since they cover *only* drug users, are sources of information that indirect prevalence estimation methods however are designed to use. Special techniques are usually required to make the best use of them.

Indirect methods of prevalence estimation

It is precisely in order to make good use of the specific drug user lists and lists of drug-related behaviours that are available that we turn to indirect methods of association. These methods begin by recognising the inadequacies of registries and other existing data sources. They acknowledge that the target populations of these registers and sources themselves may be only partially observed, and that certainly no single one of them covers completely the population of problem drug users.

Nonetheless, indirect methods set about counting the number of problem drug users by working from these incomplete lists. In some cases they make use of supplementary studies of the target population of drug users alongside primary data collection, and try to allow for the difficulties of sampling amongst the drug users themselves.

The specific analytic methods described in the manual are simple multiplier methods and capture-recapture methods and a description of the somewhat more complex methods of event-rate models is given. In addition, a brief exposition of extrapolation from local prevalence studies to national prevalence estimates.

Reliability and the use of multiple methods

Indirect estimation methods may be very unreliable. Furthermore, the estimates they make are subject to assumptions that it is often impossible to verify and that, when violated, can lead to bias every bit as serious as that in population surveys..

Some of the standard survey methods using confidence interval estimation are available, if the assumptions of the indirect methods are valid, but this lack of robustness of the estimation procedures is not something that can be addressed by the usual confidence interval estimation approaches.

It is common practice therefore to look for concordance and convergence in estimates made by different indirect procedures. To a great extent the reliability of these approaches is only judged by the extent to which they converge to a common and plausible estimate.

National and local prevalence studies

These indirect methods may be used, theoretically, at a national level to establish prevalence amongst the general population, and in particular multiplier/benchmark methods are sometimes used in this way. More typically, though, they are employed in studies of drug use on a smaller, geographically local scale. At this level they are more easily organised to take advantage of locally available data and to accommodate local variations.

The importance of local estimates on harm and prevalence cannot be stressed too much, since geographical variations in drug abuse habits are very strong.. Nonetheless there is still a need very often for overall national estimates to be made, and one way of doing this is to extrapolate from local prevalence studies to an overall picture. Extrapolation methods used in this way attempt to predict prevalence rates in areas where there has been no local study, by comparing them with areas where the prevalence rates are known - or, rather, have been estimated. The technique of extrapolation therefore requires data that are *related to* drug-taking - drug abuse '*indicator'* information - in absence of actual drug abuse prevalence figures themselves, in order to make these comparisons.

If the drug abuse indicator data can be organised to cover the whole of a country and still be available at a local area level, then local estimates of drug prevalence can be extrapolated across the whole country, area by area. The predictions that can be made for other localities are in themselves useful, but is a simple step to combine them to give an overall national estimate of prevalence as well.

These extrapolation methods are based on statistical regression techniques, and are called sometimes 'multiple indicator methods' or 'synthetic estimation'.

3 Specific methods of prevalence estimation

An overview

Thumb-nail sketches of the principal methods are given here, to allow the reader to identify and focus on the chapters of the module that are of greatest interest and relevance, whilst skipping the rest. All these methods of analysis are designed to produce prevalence estimates for populations that are essentially hidden from view, at least with regard to being listed in an available sampling frame. This section gives therefore a brief summary of how the methods differ, and what the essential data requirements are for their application. The reader's attention is drawn to the EMCDDA guidelines on indirect methods on the world-wide web: http://www.emcdda.org/multimedia/project_reports/situation/guidelines_prevalence_pdu_mb_09-01.pdf

Multiplier/benchmark calculations

In multiplier/benchmark studies the research makes use of pre-existing data - usually on a national level - for some behaviour or event that is common in the target population of problem drug taking, for example, police arrest data for drug use or possession, accident and emergency ward data and, more directly, drug treatment data, drug-related deaths data. This pre-existing information, which can be simply an anonymous count of the key behaviour over a fixed time period, is called the benchmark information. Along with this national data set is required an estimate of the proportion of the target population who have experienced this event, that is, who have been arrested, died, and so on; the inverse of this proportion is called the multiplier. Estimating the associated multiplier requires, usually, a small, separate sub-study and again, usually, anonymous records are sufficient.

An early paper by Hartnoll et al. (1985) illustrates the application of the simplest technique, using deaths amongst drug users. To apply the multiplier procedure to estimate the number of drug users in a given year, he uses two things:

a) the number of deaths to drug users in that year, say 3000; this acts as the fixed 'benchmark' in the calculation

b) the death rate amongst drug users in that year, say 2% over the year, or 1 in 50 dying in the year; this provides the multiplier in the calculation.

The estimate of the number of drug users in that year is calculated from these two figures as the population size required for a 2% death rate to result in 3000 deaths. If 1 in 50 die then the overall population must have been 3000 x 50 = 150,000. The calculation is notable for its simplicity and directness.

The proportion of the target population in the benchmark may be obtained separately and independently by interview/questioning or by other specific studies. Sometimes it is possible to use figures from already published data, if they appropriate for the target population, or even from a general population survey itself if it contains a high number of drug user respondents from the target population. There are a range of different types of multiplier study that can be carried out – including nomination studies, mortality multiplier, treatment multiplier and others.

Capture-recapture methods

Capture-recapture studies also make use of pre-existing lists, but this time lists of the target population of drug users and moreover lists in which the individuals can be identified (by name or id number or otherwise). Of course, these list are incomplete lists of the target population, and the method compensates for this by using more than one such list, for example a list of drug users arrested and a list of drug users treated.

CRC methods were originally used in the study of animal populations to estimate the size of the population and the terminology has persisted, inappropriately, while the methods have migrated from the study of animal populations into a sociological setting, and particularly into drug use prevalence estimation. 'Capture' is equivalent to the drug user being listed as attending a treatment centre or being on police arrest records, for example, where a 'capture list' of identified individuals can be derived. 'Recapture' is then equated with appearing on two (or more) such capture lists when the records are cross-referenced. It is important to recognise that the lists themselves almost certainly do not cover the whole of the target population, and the capture-recapture calculation gives the number of drug users *not on either/any of the list.*

An early example of the method was given by Ghodse (1980), in which two U.K. official data sources were used. The U.K. Home Office Addicts Index was a register of Class A drug users who had come to the attention of the medical authorities in some way, and which was criticized as commanding poor compliance. From the U.K. death register it was possible to identify the deaths due to drug addiction, and these were presumed to be deaths of drug addicts. The first of these capture lists was itself considered an incomplete figure for the total number of drug addicts 'notified to the U.K. Home Office Addicts Index'. The second capture list, by cross-referencing with the first, identifies amongst the drug-related deaths the proportion of addicts notified to the Home Office Index. Assuming this notification rate to be the same amongst living addicts, it provides the requisite proportion for identifying the extent to which the Home Office list falls short in ascertaining the total number of U.K. addicts.

The method is customarily extended to incorporate the use of any two (or more) lists of drug users that are independently derived or constructed, even though they may not be chronologically sequenced as they were in the original animal studies. The general principle is that official data - any routinely collected lists of drug users - are always incomplete with respect to covering the whole of the drug user population. These methods aim to calculate the extent to which the drug using population is *incompletely ascertained* in any list.

The requirements are then that cases must be identifiable for matching across the lists used so that the proportion of matches can be identified directly from the data sources used in the study. In this procedure, no interviews are necessary, no specialist studies of the target population are required and the capture lists used may have incomplete coverage of it.

Extensions and advanced event-rate models

Event-rate methods are based initially a set of method that parallels the multiplier methods and the ‘events’ are usually those same institutional contact events that can be used in multiplier/benchmark or in capture-recapture studies. The two types of information that need to be obtained are:

(i) the overall rate amongst *all* drug users (both those who are in contact and those who are not in contact) at which contact events are generated, and

(ii) the numbers of institutional contacts events made.

These two pieces of information allow us to infer the total number of drug users who were active during the data collection period. It is simply the sum over the number of events of a given kind divided by the corresponding rate at which such events are generated.

To estimate the rate at which contact events are generated is not a simple matter, especially since what we really want to know is what proportion of drug users will generate no contacts. The research trick used is of course to manage to calculate the rate from interviewing only drug users who *are* in contact. If it is feasible to make particular, vigorous assumptions about the relative frequency a drug user *makes multiple contacts,* it is possible to estimate from the pattern of repeated contacts what proportion of drug users never makes contact.

In practice, the estimation procedure becomes more complicated than this simple multiplier formula, because the pattern of multiple contact events under study can be complex and because

these methods typically would allow for different groups of drug users having different rates of contact. But the point of the method is that it allows us to make inferences about the size of the drug using population in total by interviewing people found at a few carefully selected places.

Issues in making a choice of method

Apart from the specific data availability that is needed for each of these methods, there are further practical considerations that influence what choice of study can be made. The first of these relates to the scale or scope of the study. Although any method can target a local or national population of drug users, capture-recapture studies are often more easily mounted on a local, small scale if only because of the matching of identifiers that is required. A second consideration is whether there are resources enough to conduct a specialist study to estimate the multiplier. Depending upon what official data sources there are, mortality studies to estimate a death rate can sometimes take a long time; if there are suitable studies in which the information is already published, these provide make a second-best option. A third factor is whether any interviewing can be don at all - for detailed information this is usually necessary - or whether existing records are adequate. With regard to accuracy and reliability, all these methods have high uncertainty attached to the resulting estimates: it is better to have a large study in which the number of observed drug abusers is as large as possible and the number unobserved as small as possible. It is usually desirable to have more information on the drug users than less: suitable information on gender, age, and so forth can be used to improve the estimates in all cases.

Extrapolation and synthetic estimation

The sections on describing specific methods are accompanied by a section on extrapolating known prevalence of drug use in some regions into estimates for other regions. Typically this is used for generalizing from a series of local studies to provide estimates of national prevalence.

Extrapolation methods are not really a specific method of prevalence estimation as such, but in principle they fulfill the same function when some prevalence information is known for some areas. In general these methods come under the heading of 'synthetic estimation' although technically they are no more than standard statistical regression procedures. The method sometimes known as 'multi-indicator' estimation is also essentially the same principle and is used in similar circumstances.

The centrally important element of synthetic estimation and any other extrapolation method is that it makes use of known prevalence figures in certain regions to estimate prevalence in other regions To do this these regions - the 'target' regions - must have some data sources that are the same as (or very similar to) the regions - the 'anchor' regions - for which prevalence estimates do exist, although of course they lack the regional prevalence figure itself. These data sources are referred to the drug indicator data: measures that are related to the prevalence of drug abuse but cannot in themselves provide a satisfactory figure for it. It is the comparison of target with anchor regions on these indicator variables that provides the basis of generalizing the known prevalences.

This procedure is then an essentially data-rich procedure that uses data that is available over a geographical breakdown of the country.

The role of case studies in the following sections

The document overall and the guidelines are grounded in practical examples from the existing research literature. Although it is difficult to separate particular methods in isolation in the research literature, the research examples have been threaded throughout the individual chapters, so that they are to be found where they are most relevant. These examples and commentaries constitute about half the manual; the remainder is intended as a major piece of work, pulling together the various strands of indirect estimation methods and highlighting their connections.

The methods described here are all illustrated by case studies in each section of the manual. These case studies are intended to be a crucial part of the description of the methods and they

give a great deal of practical information as well as theoretical perspective on how they are implemented. They have been chosen wherever possible to be studies conducted outside the European forum, in areas where existing data and research facilities are less than perfect. They have also been chosen as examples of the estimation methods being applied in difficult circumstances, so that they are studies that report flaws in and approximations to the desiderata of the methods, and do not present themselves as models of perfection.

The list of examples documented in the manual is given at the end of the introduction section.

Using Multiplier/Benchmark Methods

Introduction

Of all the methods of indirect estimation the multiplier/benchmark approach is probably the easiest to implement and probably the one with the longest history of use in the drug epidemiology field. There is a flexibility in how it is applied that makes it useful in many circumstances. In the standard application it uses information about the known size of an identifiable sub-section of the target population of drug users, and generalizes from this sub-section to give an estimate of the complete target population by applying a multiplying factor.

We begin with a simple illustration of the method, and then discuss more generally the strengths and weaknesses of the method in a wider context.

Simple multiplier technique

The essence of the multiplier/benchmark calculation is that we have some information on a subset of the target population - usually some count of the number of drug abusers making contact with a particular agency - that we try to use to estimate how many more there are in the overall target population. If for example we know how many drug abusers are in treatment in 2001, and we know that approximately 1 in 10 abusers attended treatment in 2001, then we can 'multiply up' this treatment figure by a factor of 10 to get an estimate of the total number. These two components - the *known figure* in treatment contact (the treatment benchmark) and the *estimated proportion* of abusers who were in treatment contact (giving the treatment multiplier) are what gives the method its name.

The difficulties involved in this method and their consequences are discussed at length, but its strength is that the principle underlying the calculation is very simple, and that it can make use of a variety of data. Clearly other benchmark groups than 'those in treatment contact' can be used, provided we know the corresponding benchmark and multiplier figures. The Table below lists a set of likely data sources that may be available to be used as a benchmark in a prevalence estimation exercise.

The case study 1 illustrates the basic calculations required for making a multiplier/benchmark estimate of the number of injecting drug users in Toronto in 1996.

Potential data sources for benchmarks for multiplier methods to estimate prevalence of problem drug use[2]	
Data Source	**Example**
Specialist drug treatment	Drug users on methadone, attending treatment agencies, or in residential care
Specialist drug treatment	Drug users on methadone, attending treatment agencies, or in residential care
Low threshold drug agencies	Drug users attending drop-in sites or contacted by out-reach workers
Needle exchange	Drug users registered at needle exchange programmes
Casualty	Drug users attending casualty because of an overdose
Laboratory	Drug users tested for HIV, HCV or HBV
Police / Prison	Drug users arrested or imprisoned for drug offences, Drug users arrested or imprisoned for other crimes
Probation	Drug users on probation
Social services - assessments	Drug users assessed by local social services
Hostels for drug users	Drug users living in hostels
Addict Registers	Drug users reported to a central register
Surveys of problem drug users	Community surveys of drug users
Overdose deaths	Number of deaths due to opiate overdose

[2] From Hickman et al. (2000). Estimating prevalence of problem drug use: Review of methods and their application. *United Nations Bulletin on Narcotics, Drug Abuse Epidemiology: Science and Practice*. Volume No. LIV.

Case Study 1 Toronto Multiplier Study based on HIV tests[3]

A basic multiplier calculation

Archibald and colleagues outlined a multiplier method of estimating the prevalence of injecting drug use (IDU), making use of information from laboratories of the number of HIV tests by IDU, and data from surveys of IDU of the proportion that had had an HIV test in a given year. Below we describe the findings for one city and one year: Toronto in 1996. The example requires two elements; the first is:

> a) A known benchmark figure. Here this figure is the number of HIV tests made on injecting drug users in Toronto in 1996 which was recorded in routinely collected information as 4050. This represents the known part of the population of injectors.

To find the total number of injectors then we need to know what fraction of them are unknown to HIV testing records. The second element the method needs is therefore:

> b) A multiplier that tells us how many more injecting drug users in Toronto have *not* had HIV tests in 1996. This figure can be worked out simply if we find out the *proportion* of drug users who have had these HIV tests in the period. In this example this proportion of users tested for HIV was known from other studies to be 25%, or 1 in 4..

The calculation illustrated in the Table below then made simply by noting that if 1 in 4 injectors have been tested then the total number of injectors must be 4 x 4050 or 16,200 people.

The method assumes that we have an unbiased estimate of the multiplier. Ideally this estimate will be obtained from a representative sample of problem drug users and collected over the specific time period and place corresponding exactly to the time period and geographical location of the benchmark figure that is going to be used.

In practice this rarely happens. In this case study in Toronto, the authors used a multiplier from a survey of injectors carried out in a different city and assumed that it would be the same in Toronto for 1996.

Using HIV test numbers to estimate the number of injecting drug users.

Benchmark (B)	Number of HIV tests by injecting drug users in 1996 Source: laboratory reports	4050
Multiplier (M)	Proportion of injectors reporting getting an HIV test in the previous year Source: community recruited survey of injectors	25%
	Multiplier Calculated as: 1.0 / 0.25 (i.e. One in four)	4.0
Population estimate	Benchmark times multiplier (B * M)	16,200

Note: For reasons of presentation and explication the numbers in the table in this example have been changed slightly from those in the original publication.

Reference

Archibald C.P. et al. (2001) Estimating the size of hard-to-reach populations: a novel method using HIV testing data compared to other methods, AIDS, 15 (suppl3): S41-S48

3

What is noteworthy here in case study 1 is that it uses standard, routinely collected data – the numbers of drug injectors who have had HIV tests in the past year, from official data sources in this instance – to provide a benchmark figure for this 'officially visible' part of the drug injector population. And to multiply up to get the size of the whole drug injector population from this officially visible part, it uses information from other published studies. In this example therefore no new research study was carried out to provide any information – it was all available in some form already.

Of course, as was pointed out in the case study, a heavy compromise was made with accuracy in assuming that the multiplying factor of 4 that was used was relevant for Toronto HIV-tested drug injectors, when in fact it was derived from another location at another point in time. The following, second case study uses a common alternative benchmark and multiplier system, based on official statistics of recorded deaths of heroin addicts. As a benchmark figure this is again something that is available as an existing data source, but the study makes similar compromises in establishing the 'deaths multiplier' value.

Case Study 2 New South Wales overdose deaths

We give a second simple illustration of the multiplier/benchmark method, this time from a study using a 'deaths multiplier'. This study looked to confirm estimates of the number of regular heroin users in New South Wales that had been derived from other methods and studies. It set out to do this by applying multiplier methods to the national heroin overdose data, specifically to heroin-related overdose fatalities. Here we report only this deaths multiplier method, although this is in fact only part of a much more extensive study. (See other case studies below)

Obtaining a multiplier

The most difficult part of a multiplier estimate is usually estimating the multiplier itself. In this case, no direct nor separate study was carried out to get such an estimate, and instead a heroin overdose deaths multiplier was obtained from the existing publications on the subject. Specifically, pooling the results of a number of cohort studies of regular heroin (e.g. Frischer, 1998; Reuter, 1993) users indicated that between 0.8% to 1.0% of people who use heroin regularly will die of a heroin overdose in any given year. This implies that somewhere around 1 in 100 or so heroin users die each year as a result of an overdose, and so a 'deaths multiplier' would have a value between 125 (0.8%) and 100 (1%).

Using the benchmark figure

The number of heroin overdoses recorded in NSW was on average about 360 per annum over the period. Applying the likely multipliers to the available data on overdose indicated therefore that there would need to be between 36,000 (applying the multiplier of 100) and 45,000 (applying the multiplier of 125) people in New South Wales who were regular heroin users in order to generate the observed mortality rates. This broad estimate is similar to the previous estimates made by other means.

Using overdose deaths to estimate the number of regular heroin users in NSW based on a 1% p.a. mortality (figures are averaged over a period of five years).

Benchmark (B)	Number of heroin overdose deaths per annum Source: available mortality records	360
Multiplier (M)	Proportion of regular heroin users who die of an overdose annually Source: quoted mortality rate in research papers	1%
	Multiplier Calculated as: 1.0 / 0.01 (i.e. One in one hundred)	100
Population estimate	Benchmark times multiplier (B * M)	36,000

Caveats

Of course, this procedure is not without its limitations. In particular, the bulk of the cohort studies used to produce the estimate of the annual rates of overdose fatality were conducted outside Australia and covers a span of time only loosely congruent with the benchmark data. There may be time changes in rates of overdose deaths and our method would not be able to account for such regional or temporal differences in rates of overdose. While useful as a first approximation in areas where accurate mortality data is readily available, the relatively crude nature of the multiplier used means that this method can produce, at best, only a rough approximation. Improvements in the accuracy of this method could potentially be made by conducting region- and time- specific studies of rates of overdose mortality among regular heroin users, even though the multiplier used, through lack of more detailed information, would have to be the same in all cases.

The simple statement of the procedure ignores the various compromises that must be made in practice: the definition of drug user of course needs to be precisely specified the "number of drug-users dying in the year" is replaced by the "number of deaths recorded as drug-related"; the "mortality rate" from published cohort mortality studies is presumed to be a reasonable estimate of the "ratio of drug-related deaths to the number of drug users for the year in question", even though it was calculated over a different time period and in a different country.

The particular difficulties with a deaths multiplier is that the estimate of the mortality rate amongst drug users ideally requires a specialist, longitudinal, local study to estimate the mortality rate which, unless a very large study, will take a long time to find the result. A further problem is that the multiplier itself is very large: if only 1% of the population is visible then the unreliability of the estimate is obviously greatly increased.

Nonetheless, the convergence of this estimate with other previous estimates strengthens our confidence in these estimates and highlights the benefits of combining estimates from a range of methods, each of which may have numerous limitations in themselves, to give some reassurance in the resulting estimate.

References

Frischer, M. (1998) Estimating the prevalence of drug abuse using the mortality multiplier method: an overview. in EMCDDA (1997) *Estimating the prevalence of problem drug use in Europe*. Scientific Monograph No. 1.

Hall, W., Ross, J., Lynskey, M., Law, M., & Degenhardt, L. (2000). How many opioid users are there in Australia? *Medical Journal of Australia,* **173**, 528-531.

Reuter, P. (1993) Prevalence estimation and policy formulation. *Journal of Drug Issues,* **23,** 167-184.

Again the preceding case study demonstrates that the estimate of the total population of regular heroin users is made by multiplying the known or 'officially visible' number (the numbers of drug users dying in the year) by the deaths multiplier (the reciprocal of the annual mortality rate amongst drug users). In countries where drug deaths statistics are not easily available, a commonly used alternative benchmark is number of drug users in treatment. To use the 'in treatment' subgroup to provide the benchmark requires (a) the total number of drug-using population who were in treatment at some point during the year in question, and (b) an estimate from some sample survey of the proportion of the drug-using population who were in treatment that year (for example, one in five has been a commonly cited multiplier in U.K. research publications).

Case Study 3 New South Wales treatment and arrests data as a multiplier

Multiple estimates using different data sources

In Australia during the 1990's (and in many other countries) there appeared to have been a rise in the availability and use of heroin. Heroin use and issues associated with it had become a major public and political issue and there was intense media debate about the extent of the problem and potential strategies for ameliorating or reducing heroin –related harm. These concerns and the surrounding debate, however, appeared to be happening largely in the absence of any data on the number of people who used heroin or who were heroin dependent. In fact, despite widely held beliefs that the number of people who used heroin had increased dramatically, there appeared to be no solid data available on this point. Indeed, suggestions that heroin use had increased in Australia were based largely on four types of information:

(i) Evidence from overseas that the worldwide production of opium had increased.

(ii) Local evidence, based both on police intelligence and on interviews with heroin users and other key-informants, that the street price of heroin had reduced, and that this reduction was paralleled by a rise in both the purity and availability of heroin.

(iii) A gradual, but steady, increase in the number of people seeking treatment for heroin dependence.

(iv) Finally, perhaps the most compelling evidence of a rise in heroin use came from well-documented evidence of a steep rise in fatal heroin overdose during the 1990s

Against this background, a group of researchers was asked by the government to attempt to estimate the number of people in Australia who were regular or dependent users of heroin.

Data sources

There were a number of reasons why this task may be relatively easier in Australia than it would be in many countries. In particular, the Australian Bureau of Statistics collects comprehensive mortality data, coded accorded to ICD-10 disease codes, including drug-related fatalities and both the police and treatment agencies keep comprehensive records. Nonetheless, even given these advantages, estimating the size of this "hidden population" is, as discussed above a difficult problem.

While we primarily interested in trying to establish an estimate of the number of regular heroin users in the whole of Australia, it was easier to first estimate the number of regular users of heroin in New South Wales, the most populous of Australia's eight States and Territories. We had data available to us from two principal sources for the State of NSW:

1. *Arrest data.* The NSW police service provided data on arrests for heroin-related offences (possession, supply of heroin) for the period 1997-1999.
2. *Methadone maintenance data.* The primary mode of treatment for opioid dependence in Australia is Methadone Maintenance Treatment (MMT). Methadone is prescribed by physicians working in specialist treatment centres and also by primary care physicians. People prescribed methadone are first registered with the Pharmaceutical Services Branch (PSB) of NSW Health to insure that individuals do not receive methadone from multiple sources.

Multiplier calculations

It was possible to use the methadone maintenance treatment data as a benchmark for estimating the number of regular heroin users in New South Wales. Standard records gave the total number of people entering MMT over the period who were regular heroin users as 13,000.

Using this as a benchmark figure, it was then necessary only to find the appropriate treatment multiplier to estimate the total number of users, including the number of untreated users. Earlier studies had suggested that about one third of heroin users interviewed in those studies had entered MMT in the preceding year. From this total the overall number of regular users could be estimated, multiplying by a factor of 3.0, as 39,000.

Using Methadone Maintenance Treatment patients to estimate the number of regular heroin users in NSW		
Benchmark (B)	Number of patients in MMT at clinics over the year Source: available clinic records	13,000
Multiplier (M)	Proportion of regular heroin users who have had MMT in preceding year Source: published studies of samples of heroin users	33.33%
	Multiplier Calculated as: 1.0 / 0.3333 (i.e. One in thre)	3
Population estimate	Benchmark times multiplier (B * M)	39,000

Robustness of the result

These data were analysed using a variety of techniques, including capture –recapture, as well as multiplier methods. We were also able to use data over a period of many years and methods of back-projection, originally developed to assess HIV/ AIDS, as a means of charting historical changes in the number of people who were regular or dependent heroin users.

The application of these different methods produced a surprisingly narrow range of estimates: the six estimates of the number of people in NSW who were regular heroin users ranged from 32,000 to 45,000 with a median value of 37,000. Given the narrow range of estimates, we decided to simply calculate the median estimate as our "best estimate" but it may also be possible to use a more sophisticated method for combining different estimates by weighting estimates by the relative confidence that you have in their application and conclusions.

Consequences of imperfect procedures

This study drew attention to one of the difficulties of making similar calculations using the police arrests data records. From earlier studies it had been estimated that about 20% of interviewed regular heroin users had been arrested in the preceding year, so that a multiplier of 5.0 was needed to cover those regular users who had not been arrested. The number of arrestees during the study period for heroin-related offences was approximately 2400, a figure obtained from the standard police arrests data base. Applying this multiplier to the arrests total gave an estimate for NSW of 12,000 regular heroin users - a much lower figure compared with other estimates, including the MMT multiplier method.

There were several reasons why the police arrests data might not be suitable for the MBM calculation. It seemed likely that the arrestees amongst the interviewed regular users might have been recorded as arrests for offences other than heroin-related offences. It would also have been possible that the interviewees were a set of users who were more prone to arrest than the general run of regular heroin users, although there was no particular reason for this to think this was the case. This would imply, in the first case, that the multiplier as calculated from the interviews did not exactly match to the benchmark definition; in the second case, the multiplier would not fairly represent the general position. In either case therefore the actual proportion of regular users in general whose arrests were recorded as heroin-related was probably smaller than the figure estimated fro the studies. This would imply that the derived multiplier would be too small, and applying this to the benchmark figure therefore gives rise to an under-estimate of the total number of regular users.

References

Hall, W., Ross, J., Lynskey, M., Law, M., & Degenhardt, L. (2000). How many opioid users are there in Australia? *Medical Journal of Australia,* **173**, 528-531.

McKetin R., Darke S., Hayes A.. and Rumbold G. (1999) Drug trends 1998. A comparison of drug use and trends in three Australian states. NDARC Monograph No. 41. Sydney, UNSW.

Primary collection of new data

We recommend that when ever possible the research conducts a sample survey of the target population – injectors or problem drug users - as part of the prevalence estimation study. This has several advantages.

- These surveys can collect several multipliers and if identifiers are collected for the interviewees it can provide a further data source for capture-recapture methods (see Section 5) of prevalence estimation.
- Key risk and protective behaviours can be collected, the study can be used to estimate the prevalence of blood borne viruses and provide a measure of coverage of harm reduction.
- Finally, interviewing the target population allows questions to be asked giving information that could be used in a multiplier study (see Section 4).

The use, in the previous example, of that multiplier and benchmark - the treatment multiplier calculation - in particular makes it possible to carry out a special study to determine the multiplier value. We now give an example (Case Study 4) where new data are collected for estimating the multiplier along these lines and, moreover, new data are collected to establish a benchmark figure.

Case Study 4 Pakistan National Assessment Study 2000

A multiplier-benchmark study using a treatment multiplier from interviews with key informants:

Background position and assessment

The aims of the overall exercise in Pakistan was to establish the national prevalence of hard drug addiction. Although there were good data for the size and structure of the general population in all the Pakistan provinces, there was little reliable information available on drug use itself. An earlier survey had attempted to estimate the figure using national survey methods but was at this point very out of date.

There was however some information on drug addiction treatment centres, which were specialist clinics for drug users (mostly alcohol abusers or heroin abusers). Drug abusers were also treated at state hospitals. It was decided therefore to try to use the treatment system's data of the numbers in treatment to estimate, in conjunction with a treatment multiplier, to establish total numbers in each of several localities in Pakistan.

Extrapolation of these local study figures to give estimates for each major province of Pakistan is discussed in Section 7 on extrapolation. The treatment multiplier method described here was part of a two-pronged prevalence estimation exercise, where the second prong was to use imprisoned addicts and an imprisonment multiplier. It was hoped that there might be some concurrence of these estimates based on different data, so giving support to the validity of the estimates.

Defining the target population

Target population was defined as heroin users or injectors (of anything) - these people were called 'addicts' for lack of a better term. Cannabis in the plant form or resin (hashish/charros) is most common drug of abuse in Pakistan - probably more even than alcohol, the use of which is illegal - but researchers decided that they could not practically run interviews/surveys/sampling that would be efficient for estimating cannabis and heroin or injecting prevalence at the same time.

Practical considerations lea to the target population for the prevalence estimate to be restricted to males aged between 15 and 45

(i) because social *mores* would not allow women drug users to be visible by and large at the treatment centres, although over-the-counter opiates were thought to be a considerable problem amongst females.

(ii) because previous work suggested this age-band contained almost 90% of hard drug addicts, and that males outside this age range were almost never encountered in treatment

The definitions used for the benchmark figure and the multiplier for 'treatment' was confined to specialist treatment centres for addiction. Treatment in government hospitals was excluded because (a) it was rare (b) it was not clear that drug addiction treatment could be distinguished from treatment of other patients (c) record keeping on drug history and abuse was inadequate.

Establishing the benchmark

All benchmark information on the numbers of people in treatment - at all levels - had to be established by a separate census (with 100% coverage) of all specialist drug addiction treatment centres. This was set up using as a starting point earlier, but now out-dated, lists of specialist clinics; this was updated into a 'National Treatment Register'. It is hoped that this part of the study has a spin-off for future drug use prevalence work in having a national register of these clinics that can be maintained. In total nationally, 73 such clinics were identified

The data gathering procedures at the clinics allowed for various definitions of the benchmark, but the analysis focused on the definition being set as 'numbers of addicts treated as in-patients in the past year'. (See Appendix 1 of the main Pakistan report for alternative definitions that were considered.) All 73 in-patient specialist centres' managers were interviewed either by telephone or in person to collect numbers of addict in-patients in past year. Medical personnel carried out these interviews.

Establishing the multiplier

Establishing the multiplier was again the most difficult part of the study. In this case it was derived from a special survey designed to get the right information for estimating the treatment multiplier. Calculations were based on information provided by a sample of 'key informants' (KI) in personal interviews. See the sub-section on *'Fieldwork'* below for the criteria for eligibility to be a KI. Each was asked:

(i) How many addicts they personally had encountered in the previous 12 months

(ii) How many of these had been for in-patient treatment at a specialist clinic in the previous 12 months, to the best of their knowledge.

In addition they were asked how many had been to prison in the preceding 12 months, but this information was used only in forming the secondary imprisonment multiplier. Because the questionnaire was a lengthy data gathering exercise for each KI, these pivotal questions were asked at the outset of the interview.

When using these figures to derive a treatment multiplier, KI who were involved with treatment services in some way were excluded from the analysis (and for the prison multiplier calculations, those with involvement in the police or prison service were excluded). KI who had encountered less than a threshold number of addicts were also excluded from the analyses.

Sampling in the Key Informant survey

KI were selected on nationwide basis. Thirty-six geographical 'locales' were chosen purposively - not randomly sampled - across Pakistan to represent the general social structure of the country. These were used as first-stage sampling units (clusters) and were stratified as one broadly urban and one matching broadly rural locale in close geographical proximity. These eighteen pairs spread across the four provinces of Pakistan with more in the more populous areas.

The interviewer force (40 professional workers) were instructed to interview in each locale at least five key informants of their own choosing from a list of social status categories: if possible one from each of at least five different social categories in each locale. The categories on this list were: policeman, judge, doctor, health worker, other government official, mayor, councilor, mullah, priest, social worker, teacher, tribal headman, ex-addict). Refusals to be interviewed were not recorded as no formal approach nor sampling structure could be defined within a

locale.

Fieldwork in the Key Informant survey

The organisation of the fieldwork was a difficult task in itself, and a local expert researcher was placed in charge of the survey for all day-to-day management decisions, in order to minimize delays and obstacles. Interviews carried out by staff trained specifically for this survey: medical workers, doctors, social workers, trainees in these professions and associated ex-addicts, plus one or two more suitable volunteers. To ensure comparability across a large survey workforce and to ensure precisely the right information was gathered, questionnaire-based structured interviews were used in all cases.

Again to control quality of the information across such a widely ranging study, there was strict supervision of the questionnaire flow by four designated regional supervisors. They ensured that all completed/spoiled/wasted/unused were accounted for by four regional supervisors, and each questionnaire was signed by the interviewer and countersigned after inspection by the supervisor.

Even with geographical clustering of the data collection into a limited number of eighteen locales across the country, travel between interviews was still a major problem (for example, one interviewer had to travel by camel for two days in one instance).

Analysis and results

Calculations using the benchmark of number of male in-patients aged 15-45 were made at the lowest geographical level possible, that is, for each separate locale in which there was a clinic or clinics. For extrapolation and aggregation into a National Pakistan prevalence, the reader should see Section 7 on Extrapolation.

The primary benchmark calculation was made by the following procedures:

(a) Specialist in-patient treatment clinics existed in 30% approximately of the study locales and none outside the study locales, so that benchmark figures could be calculated only for these 'treatment locales'. Benchmark totals were constructed for each of these locales by counting male heroin abusing or injecting in-patients aged 15-45. An adjustment was made to allow for the fact that about an estimated 10% of these patients did not come from the locale, but were from more remote geographical areas. The adjusted benchmark figure was pooled for all locales in a province:

(i) for the principal city of the province; and

(ii) for all other locales in the province.

(b) Multipliers were derived for each locale by taking the number of addicts a KI had encountered in the past year divided by the number of these who were treated. Of the various ways in which different KI information could be pooled into a single multiplier, the median value amongst the qualifying KIs was selected, being:

(i) a figure that was not unduly influenced by extreme responses, and

(ii) would not depend upon whether the averaging took place over the proportions treated or the multipliers themselves. In doing this, KIs were pooled in the same way as the benchmark calculations were, (i) within the principal city of the province; and (ii) across all other locales in the province.

(c) The pooled benchmark figure was multiplied by the median multiplier to give a total number of addicts in each of the four provinces (i) within and (ii) without the principal city. Two further estimates were made to give some idea of the variability in this procedure by taking the lower and upper quartiles for the multiplier instead of the median. Repeating the entire calculation for these multipliers with the same benchmark, gave a range for the number of addicts in each province.

The separate figures for all the province estimates are not presented here. The interested reader should consult the main Pakistan report.

Caveats

The prevalence estimates were obtained by using the estimated total number of addicts in each locale and dividing by the number of males aged 15-45 in the locale. These estimates were considered plausible, but were very low by comparison with the previous survey. This previous survey however and the updating surveys with which it was subsequently combined made far broader assumptions than this present study. The practice employed of using smaller-scale surveys at various intervals afterwards to provide updating factors to the original survey is certain to inflate the error of estimation, whatever the reliability of the original survey,

The estimation of the multiplier is the most troublesome part of this study, as it usually is in other studies. Here it could be made in several ways and which is chosen may influence (up or down) the answer. For example, given the different KI responses in each locale, it could have been calculated by:

- averaging individual multipliers,
- averaging individual estimates of percent in treatment,
- averaging these at locale level or at province level, or
- pooling numbers of addicts encountered and treated before averaging.

In fact, prevalence based on the last of these was also calculated and gave results very similar to the method actually used.

Furthermore the effect of the exclusion criteria for allowing KI a to qualify for entry into calculations are certain to affect the result:

- effect of threshold number of addicts known to KI, and/or
- effect of excluding treatment-related KI (or prison/police-related KI)

Information gathered from the KIs itself of course would be prone to errors of reporting. The distinction between government hospitals as opposed to specialist treatment may not have been clear in mind of a KI when reporting the estimate of the numbers of addicts encountered who had been in treatment. Furthermore the KIs may have not excluded those treated outside the stipulated one year time slot; and it was impossible to ask or check on where geographically the known addicts had been treated.

Finally, although the census data itself may be accurate, there may be problems with the definition of the treatment centres' catchment areas when computing prevalence rates from general population figures.

With the immense difficulty of making prevalence estimates in the absence of any existing data, the need to use different methods and procedures must be emphasised very strongly.

Reference

UNDCP (2002) Drug Abuse in Pakistan. Results from the year 2000 National Assessment. UNDCP, Vienna.

Other methods for estimating the multiplier value

Once the definitions of the benchmark group have been determined, it is usually a relatively simple matter to determine its size from the relevant source. What is always more problematic, at least in drug use epidemiology, is the estimation of the multiplier. In addition to the cohort studies to estimate death rates, and specialist studies for treatment or registration or arrest ratios, there is a set of designs that are characterised as nomination methods. These arise in connection with chain referral sampling procedures – sometimes called 'snowball sampling' - in which each

respondent in a core sample, personally interviewed, is asked to nominate a number of acquaintances who are then incorporated as a second wave into the sample.

In the present context, further drug use related information can be gathered about these nominees, either from the core respondent or from the nominees themselves. To estimate the proportion of drug users in treatment, for example, the respondent can be asked: "Of your 10 closest drug-using friends, how many have been in treatment in this area in the last 12 months?" Pooling information across core respondents will give an estimate of the proportion of drug users in treatment. A classic paper comparing some of the many alternative ways this type of information can be gathered and checked is that by Parker et al. (1982) reporting a small-scale study in four towns. The reader is referred to this paper for further information.

Heterogeneity and stratification of the population

The foregoing case study 4 targeted only the male population of 15 to 45 years old. In the case of the age restriction of the target, it was made because it was thought that there would be little data available for those outside this range, and this partly because so few 'addicts' were thought to be in that group. Confining attention to the targeted age range helps focus attention efficient methods of estimation of prevalence where it is most needed.

In the case of the exclusion of females, though, there was considerable concern about the levels of addiction but it was thought that combining males and females into a single study would distort the estimation. If the procedure had been applied to females it would certainly have required a different value of the treatment multiplier to be used, because so small a proportion of the women appeared in treatment. This division of the problem into sections – firstly, male and secondly, female – is an example of *stratifying* the population. The aim is to break down a heterogeneous population into more homogenous sub-groups in order to improve the accuracy of the prevalence estimation procedures. Other potential candidates for stratification criteria might be – depending on the target population – injectors vs. non-injectors, employed vs. unemployed, and so on. Any characteristics on which we have information to enable us to construct separate multipliers and benchmarks should be considered. Very little is lost if the division is in fact un-necessary – only when sample sizes become very small will there be any appreciable loss in accuracy – and a great deal is gained if the division into strata make the target sub-groups more homogenous.

In fact the case study illustrates the use of another major type of stratification: geographical stratification. Rather than derive for the whole study just a single estimate for the multiplier and a single figure for the benchmark, separate figures were calculated for each of the four major geographical regions; within those, the principal city was dealt with separately from the remaining areas. This was done because it was suspected that within these eight divisions different values of the treatment multipliers would apply, and pooling them together would be a potential source of inaccuracies. An added advantage in the case of geographical stratification is that separate prevalence estimates can be made for each of the four regions.

Assumptions of the multiplier method – where can it go wrong

The virtue of this procedure is in its general applicability, which is determined by having data on two things:

- th required benchmark – for example the number of deaths amongst drug users; and
- the required multiplier – for example the rate of drug user deaths in the total drug-using population.

The multiplier can be estimated by any available sample method; in particular, random sampling, as in the mortality cohort study instanced above, or a variety of nomination (snowball) sampling methods. This flexibility is the core of the method's acceptability.

The definition of the benchmark subgroup is selected primarily for the convenience of the researcher, and it is simply a stepping-stone to calculating the answer: drug-user deaths, treatment attendance, police arrest records and HIV clinic attendance have each been illustrated as used in multiplier analyses. *Any clear and precise definition* will do. But it must be the *same* definition that is used in both the enumeration of the benchmark and in the sample data collection to determine the multiplier. From the point of view of robustness of the prevalence estimation, this is a strong advantage.

Operationally, it is necessary to define exactly the benchmark and its corresponding multiplier. For example, with a treatment multiplier study, what sort of treatment is being employed: perhaps 'in methadone treatment' is a precise enough definition, or perhaps a list of specific treatment agencies will be preferable. Note that this definition should include specification of the geographical extent of the locality or region being considered, as well as the precise time-slot to which the data apply.

We must also pay attention to the assumptions the method makes. Firstly, we assume that the "benchmark" data are accurate. Unfortunately routine data sources can be notoriously inaccurate, because of under-reporting or incomplete data collection. For example in case study 1 the authors raised the possibility that the laboratories may undercount the number of HIV tests carried out and that clinicians ordering tests do not always specify that someone was an injector. Therefore, the benchmark total may need to be adjusted to take account of this type of under-reporting.

The method assumes that we have a correctly defined multiplier. When calculating the multiplier, the key question is really: is the person recorded in benchmark figure? If so, then the multiplier matches the benchmark successfully (even when the benchmark under-records its target). When using a benchmark like treatment numbers, it may be necessary to specify a list of clinics that are being used in order to ensure precise equivalence of benchmark and multiplier definitions. This would be especially true in studies that used geographical stratification, where drug users in one region may be treated in another.

The method also assumes that we have an unbiased estimate of the multiplier. Ideally this estimate will be obtained from a representative sample of problem drug users and collected over the specific time period and the specific place corresponding to the benchmark you intend to use. This rarely happens. In case study 1, the authors used a multiplier from a survey of injectors carried out in a different city and assumed that it would be the same in Toronto for 1996. Truly random representative samples of problem drug users do not exist and the best available option is to recruit subjects in a way that limits any potential bias. For example, if you wanted an "unbiased" estimate of the proportion of injectors registered with a needle exchange for a multiplier estimate it would be foolhardy to recruit injectors directly outside a needle exchange and ask how many are registered.

If the multiplier information is collected through interviewing, then we must assume that the benchmark event is common enough and significant enough to be remembered or detected in a sample of problem drug users. For example, going to treatment or being arrested for drug possession almost certainly will be accurately reported by a sample of problem drug users. In addition though, the "multiplier question" must be clear, so that if, for example, the benchmark is 'injectors registered at dedicated needle exchanges', those interviewed answer 'no' when they have visited only pharmacies for clean needles and 'yes' when they have visited a dedicated needle exchange agency or have visited both.

One of the key requirements is of course that the multiplier is a fair representation of the connection between the benchmark count and the overall target population. If there is, for

example, marked geographical heterogeneity in the true value of the multiplier - if, for example, treatment rates are very different in urban and rural areas - then putting these areas together in a single multiplier can be misleading. The use of stratification of the population as discussed above in sub-section 4.3 can be used if data are available separately for each stratum to overcome this danger. Identifying when it is necessary is primarily a question of looking at the data and using prior beliefs to make a judgment.

Clearly violation of one or all of these assumptions is possible, giving ample opportunity for the study to go give an inaccurate prevalence estimate. It is unwise to rely upon a single multiplier estimate (see Section 8, on Accuracy and Reliability).

Using Capture-Recapture Methods

Introduction

What is capture-recapture estimation? What it does is to take routine data sources that identify problem drug users and use them to estimate the prevalence of problem drug use.

In epidemiology or studies of humans the data sources tend to be lists of people, in our case specifically, lists that include problem drug users.

The capture-recapture method is an indirect prevalence estimation method that uses information on the *overlap* between incomplete *lists* or data sources. The lists are incomplete lists because no one source will list all problem drug users – if it did there would be no problem in estimating prevalence – and because rarely is any reporting system 100% perfect without any under-reporting of its intended cases. The overlap between lists is the people on more than one list.

What we want to do is to ascertain how many problem drug users should be on the list if it were a complete list of our target population, for example of all problem drug users in our city in this year. Hence, Hook and Regal (1995) refer to capture-recapture estimation as a technique for adjusting for "incomplete ascertainment". It therefore has clear advantages in epidemiological studies that want to estimate the prevalence of a disease from routine data sources.

Rationale

In technical, statistical terms the method sets up the data as an incomplete multi-way contingency table and analyses it using a standard statistical method of log-linear modelling (or Poisson regression). It then estimates the number required to complete the list, according to the model. We will explain what a contingency table is and what this analysis might mean below.

Capture-recapture was developed by animal ecologists as a means of estimating how many deer or fish or other animals there were in an area. Their data sources were samples of animals captured, marked and then released, with a second sample identifying how many animals were marked from the first sample (the overlap): hence the name, capture-recapture.

In human epidemiology their sets of captured and marked animals are replaced by lists of people 'captured' at some routine data source. Capture-recapture has developed now as an important method in drug use epidemiology that can be adapted and applied to most local situations. We advise we also to consult examples of capture-recapture studies and guidelines commissioned by

EMCDDA given in key references, and available electronically (see Section 2 and Section 11 for details).

A general introduction to prevalence estimation including capture-recapture is published in the UN Bulletin of Narcotics (Hickman Taylor et al., 2002). For further information three excellent reviews of the history and epidemiological uses of capture-recapture are also given in the references, which may become available electronically on the UN website (see Section 2 and Section 11).

Capture-recapture with two data sources

In this section we first, we describe the simplest example, a two data source study, followed by a more complex multiple data source study and end with capture-recapture studies made in the absence of routine data sources.

Let us start with an example – the theoretical assumptions required to justify the calculations will follow after illustrating the method..

Case Study 5 Bangkok study - A Worked Example

Below is a worked example based on a study in Bangkok in 1991 by Mastro and colleagues. They used two data sources:

- Lists of opiate users enrolled in methadone treatment programs from Bangkok's local specialist drug clinics. These were routinely collected attendance records for April-May 1991.

- Opiate-positive people arrested in Bangkok police stations between June and September 1991. Study teams made 891 visits to 72 police stations and collected urine and information on arrestees to identify opiate users: those positive for opiates became the second data source.

They found 4,064 opiate users in the specialist methadone maintenance clinics' records; and from 8212 people in the survey they conducted in police stations they identified 1,540 opiate users. These two sources are the incomplete lists of opiate users in the Bangkok at that time – incomplete because neither is a complete list of all the opiate users in the population.

To obtain the overlap in the two data sources of opiate users, the full name, sex, and dates of birth were used to match between the data sources. They found 171 people on both data sources - that is, opiate users reported as being on methadone treatment and that had been arrested and whose urine was positive for opiates.

Table 1 shows the numbers observed in the two data sources and the overlap between them, setting them out as a contingency table, and then shows the calculations needed to estimate the prevalence. For the two data sources, methadone treatment attendees and opiate urine-positive arrestees, we assume that *amongst the total number* of opiate drug users the proportion who can be found in the clinics is the same as the proportion found *amongst arrestees* and also *amongst non-arrestees*. Filling in the table on this assumption gives an estimate as shown of 36,600 injecting drug users (0.5% of total population) in Bangkok in 1991.

Estimating number of opiate users in Bangkok, 1991

S1 : Methadone Maintenance clinic attenders	S2 : Arrestees with urine positives	Number of people identified	
Found in S1	**Found in S2**		**Key**
Yes	Yes	171	matched in S1,S2 (m)
Yes	No	3893	found in S1 only (c)
No	Yes	1369	found in S2 only (b)
No	No	?	'hidden' (x)
Total in S1	**Total in S2**	Total	
4064 (m + c)	1540 (m + b)	?	**Number of opiate users in the population (N)**

Estimating number of opiate users in Bangkok, 1991 – calculations

number observed = m + b + c = 171 + 3893 + 1369 = 5433

number hidden (x) = 1369 * 3893 / 171 = 31,166

population estimate, 5,433 + 31,166 = 36,599

95% CI. = $1.96*\sqrt{(1540 * 4064 * 3893 * 1369)/171^3}$ = 4516

Rounded Estimate of IDU in Bangkok in 1991 = 36,600 (32,000 to 40,800)

In our example the researchers took two different time periods April-May and June-September, mirroring an ecological study in which the second capture takes place after the first. But in epidemiological studies this is not essential. They could have taken two data sources over the same time period (see **Assumptions** below). Furthermore, they estimated the total number for 1991 as a whole. Strictly speaking they estimated the total number of injecting opiate users in Bangkok over the study period, April to September, but it is reasonable to assume that this corresponds to the annual number. In other words we can assume that the potential number of opiate users not included by the estimate was negligible i.e. the number of new opiate users or the number of opiate users that ceased drug use, died, or left the area. (see **Assumptions** below).

Going back to the definitions in the introduction, the table shown is an example of a contingency table – in this case a two-by-two contingency table with four cells. It is incomplete because we do not know the number who are *not on either data source:* the 'hidden' or unobserved injecting drug users. That is, the two data sources do not "completely ascertain" the target population, the total number of injecting drug users.

Capture-recapture methods use the table to estimate the number required to complete the two lists. We note below the assumptions required by this method, but let us mention one in particular immediately. In two-sample capture-recapture we assume that being on one data source is independent of being on the other. That is, we have assumed above that amongst arrestees and non-arrestees alike, the users are equally likely to have been for treatment at the clinic; and conversely having had methadone treatment does not lead a problem user to be more or less likely to be arrested.

The study observed 5433 users in the combined data sources and estimated that were 36,600 altogether - in other words that about 1 in 7 (5433/36,600) opiate users were observed in the study. This is a comparatively large ratio and its validity rests on the assumption that the two data sources are independent. If they were not independent – something we cannot tell with a two sample study – the larger the ratio between the observed and the total population the larger the scope for error.

It is important to bear this point in mind when planning a study. Ideally, we can collect data on a substantial number of people to reduce the ratio of observed to unobserved people, but this is not always possible in studies of problem drug use. Alternatively we can turn to a multiple data source capture-recapture study, outlined in the next section. It is still possible to do a two-sample study, but if this is done it is advisable to seek information on whether the two data sources can be considered independent, or judge the scale of dependence so that the estimates can be treated as minimum estimates if if the data sources are positively dependent and as a maximum estimate if the data sources are negatively dependent (see below on independence).

Reference

Mastro T.D., Kitayaporn D, Weniger BG, Vanichseni S, Laosunthron V, Thongchai U, Uneklabh V, Vhoopanya K, Kimpakarnjanarat K. Estimating the number of HIV-infected injection drug users in Bangkok: a capture-recapture method. Am J Public Health 1994; 84: 1094-9

Capture-recapture estimation with multiple data sources

By multiple data sources we mean three or more lists: e.g. problem drug users in treatment, arrested, in homelessness hostels, or attending accident and emergency clinics. Matches are then made across these multiple data sources to identify the overlaps of those people on each combination of two or more data sources.

The resulting data are then analysed using a Poisson regression or log-linear modeling program. Filling in the contingency table in the previous example is an instance of very simple log-linear modeling – so simple that it does not require a program, nor even a computer at all, to do it. We will avoid describing the calculations using three or more data sources for estimating the number of problem drug users. Statistical software packages can do these calculations, and written down they look extremely complicated! The assumption of independence used in the previous example and that was vital for the calculations now has several alternative forms and these are discussed in the next section.

For those readers interested in the estimating equations, (see Bishop et al, and Hook and Regal). In general, for capture-recapture calculations with multiple data sources it is advisable to get on-site statistical support to explain or to run the Poisson regression (or log-linear modeling) program that is necessary to analyse the data. EMCDDA guidelines give detailed examples of the use of GLIM and SPSS in capture-recapture and we have added an appendix available electronically on how to use Stata in capture-recapture.

Dependencies between data sources

A dependency between any two of the sources (S1 and S2) is said to be a

- positive dependency if a person in S1 is more likely to be on S2 than someone who is not in S1, or
- negative if a person on S1 is less likely to be on S2 than someone not on S1.

These dependencies can be tested as 'interaction terms' in a Poisson regression model analysis and need to be taken into account and checked when fitting a model to the data. More complex interactions or dependencies can be tested, to, for instance involving a combination of three data sources. A number of different Poisson regression models can be fitted to the data, according to which dependencies are included and which omitted. Which is the best is determined usually by seeing which one best predicts the data, as measured by standard statistical 'lack-of-fit' coefficients.

Fitting a model

The practicalities of fitting a model to multi-source data are as follows.

- Collect three or more data sources of problem drug users
- Match the data sources – identifying people that are on more than one data source, and which data sources they are
- Prepare a multi-way table for analysis in a data file (see example below)
- Analyse the table using Poisson regression.
- Fit model with interactions between the data sources corresponding to potential dependencies.
- Select the best fitting model based on standard lack-of-fit measures (a statistician will give advice on this – see EMCDDA guidelines and appendix).
- Use the model to estimate the number in the unobserved part of the population and also to calculate confidence intervals
- Repeat the whole analysis if possible for different sub-groups (males, females, different age-groups etc.)

We have selected an example below of a multi-source capture-recapture analysis from a study in Glasgow in 1990 by Frischer et al.

Case Study 6 The Glasgow study of injecting drug users

A multi-source capture-recapture study with stratification

Frischer and his colleagues defined the target population as injecting drug users, rather than specifically heroin injectors, because in Glasgow it was common to inject a large number of different types of drug, though strictly speaking data were collected on a range of drug types (heroin, other opiates, amphetamine, cocaine, and benzodiazepines).

The researchers collected data from four sources for a one year period:

- positive and negative HIV tests of people whose exposure to risk was reported as injecting drug use
- people who presented to specialist drug agencies for treatment of their drug problem (including heroin, other opiates, cocaine, amphetamine, and benzodiazepines)
- people arrested for a drug related crime (other than cannabis)
- people registered with the local needle exchanges.

In this type of study with multiple recapture data sources, drug injectors have to be matched across all four sources, so that we can calculate how many were found at each combination of sampling points. This is shown in the text box. In total, 3444 records were collected: 508 from Police data source (S1), 1179 from Needle Exchanges (S2), 507 from HIV test laboratories (S3), and 1250 from specialist drug treatment agencies (S4). After matching for duplicates 578 people were found to be on more than one data source (e.g. 4 were on all four, 41 were on S2, S3 and S4, and 147 were on S2 and S4) giving a total number of 2,866 injecting drug users observed in the study.

In the Glasgow study a number of models were fitted, ranging from one in which all sources were assumed to be independent - no interaction terms fitted - to one designed to allow for complex interdependencies in which all three-source interactions were fitted. The complete independence model did not fit at all well.

The best fitting model for the data in Example 2 included interactions between three data sources: HIV tests (S2), needle exchange (S3) and drug treatment (S4). By contrast, police arrests for possession (S1) was included in no interaction terms in this best-fitting model, implying it was independent of all the other data sources. The interaction between sources was generally positive, meaning that those drug injectors on one of the lists were more likely to be on another. In these circumstances where data sources are dependent, then fitting a model that assumes the data sources are all completely independent is likely to lead to an underestimate of the unobserved fraction of the population. In this study, the complete independence estimate was in fact lower than the estimate generated from the final model.

The researchers estimated that, overall, there were 8,500 injectors, which is about 1.35% of adults aged 15-54 in Glasgow. Sufficient data were collected in the study to enable further analyses (i.e. a further modeling exercises) of the data by sex and by age-group.

This is shown in the Table 5.3. This is called stratification and is one of way of dealing with the problem of heterogeneity (see **Assumptions** below).

Multi-source capture-recapture data from a study of injectors in Glasgow, 1990.

S1 Arrested for possession of illegal drug (exc. cannabis)	S2 Registered with needle exchange programme	S3 Tested for HIV with exposure as IDU	S4 In specialist drug treatment	Number of people identified
Found in S1	**Found in S2**	**Found in S3**	**Found in S4**	
yes	yes	yes	yes	4
yes	yes	yes	no	2
yes	yes	no	yes	13
yes	yes	no	no	56
yes	no	yes	yes	8
yes	no	yes	no	17
yes	no	no	yes	50
yes	no	no	no	358
no	yes	yes	yes	41
no	yes	yes	no	52
no	yes	no	yes	147
no	yes	no	no	864
no	no	yes	yes	116
no	no	yes	no	267
no	no	no	yes	871
no	no	no	no	?
Total in S1 508	Total in S2 1179	Total in S3 507	Total in S4 1250	Total Pop. ?

The best fitting model was one that included interactions between sources S2 and S3, S2 and S4, S3 and S4, and a three-way interaction between S2,S3,S4; source S1 was independent of all the others. This model gave the following estimates (Table 5.3).

Estimating number of injecting drug users in Glasgow – numbers and prevalence estimates.

	Observed / Known	Estimate of hidden	Estimated Total	Estimated Prevalence
All	2866	5628	8494	1.4%
Males	1977	3567	5544	1.8%
Females	889	2349	3238	1.0%
15-19	264	640	904	1.0%
20-24	1137	1613	2750	2.6%
25-29	878	1724	2602	2.7%
30-34	342	796	1138	1.4%
35+	245	1273	1518	0.6%

Rather uniquely in the Glasgow study the police source was completely independent of the other sources. Furthermore the other three data sources S2, S3 and S4 were all dependent. They could have been combined as a single 'Other sources' data list and a simple two-source study carried out with the police data. It is worth going back as a simple exercise to case study 1 above and calculating the prevalence using the two-source calculations illustrated there. The data would then be the following:

- a combined list for treatment, HIV register, Needle Exchange – 2508;
- a police arrests list – 508; and
- matches across 'police' and 'combined' lists – 150

This exercise gives the same answer as the Poisson regression above using multiple sources of data! Why? Because the very special circumstances of the results in the multi-source analysis indicate it is equivalent to using two independent data sources. This is quite reasonable: certainly appearing on the treatment list is probably going to make an appearance at the needle exchange more likely, and possibly even more so if that person is on the HIV register – a three-source interdependency. By contrast it is plausible at least that police arrests represent a random sweep across all users, regardless of treatment, needle exchange or HIV register status.

The point demonstrated here is that a researcher could not have known that the police data source was independent or that all the other data sources were dependent without carrying out a multi-data source study and analysing the data using Poisson regression models. Embarking on a two-source study may give the right answer, but it is impossible to check even the most obvious assumptions on which it is based.

This example illustrates a further point: only in rather special circumstances – such as those revealed in this study – is it a good idea to combine different data sources as though they were just one source. In general, although it is alright if different source lists have different chances of 'capturing' problem drug users on their list, it is important that everyone in the population has the same chance as does everyone else of being observed on any given list. This is probably not going to be the case if we combine very different source lists such as HIV registration and police arrests and it will be a violation of the homogeneity assumption (see **Assumptions** below).

Reference

Frischer M, Leyland A, Cormack R, Goldberg DJ, Bloor M, Green ST, Taylor A, Covell R, McKeganey N, Platt S. Estimating the population prevalence of injection drug use and infection with human immunodeficiency virus among injection drug users in Glasgow, Scotland. Am J Epidemiol. 1993 Aug 1; 138(3): 170-81

What software to use

Two sample studies can be conducted using pen and paper, calculator or a spread-sheet. In most cases multi-sources capture-recapture require statistical software. Bishop et al. (1975) do give the formulae for calculating population estimates by hand**Error! Bookmark not defined.**. Guidelines from EMCDDA give examples of the use of the standard packages, GLIM, and SPSS, and we show how to use Stata in the final section of the manual.

Data sources and how to match across them

Criteria for data sources to use

In capture-recapture the best data sources ideally

- identify clearly the target population we want to estimate
- (e.g. heroin users injectors, or problem drug users).
- collect lots of data that can be used for matching
- collect potential stratifying variables,
- provide the data in electronic form so that it need not be collated by hand.

Unfortunately this rarely happens. So what we need are data sources that are good enough to fulfill the criteria noted in the box below.

DATA CRITERIA

- Do they identify the target population that we want to estimate?
- What identifiers are collected, and what will the data owners release?
- How will the data be collected or provided?
- How many cases will be provided?
- How will the data sources fit with others?

The last point refers to one of the assumptions underpinning capture-recapture (see **Assumptions** below): that the data sources are representative of the target population. If we can, then we should select data sources from the criminal justice field and from the treatment field.

Section 4.1 shows routine data sources that can be available. It is best to undertake an inventory in the study area to find out how many of these are available locally, and whether there are any other potential sources. In aggregate form (grouped) these could be used as benchmarks for multiplier methods, and in disaggregate form (ungrouped or one line of data for each report) these can be used for capture-recapture studies. Remember that data sources for capture-recapture do

not have to be complete (i.e include every possible case) but they do have to be accurate and reliable (i.e. correct identifiers and accurate drug information).

Data sources with small numbers can be combined with other data sources, to an extent, although this trick is not without its pitfalls. Capture-recapture studies with more than five data sources can be complex to analyse. This is because of the potential number of models that could be fitted when using a larger number of data sources.

Matching across data sources

Unless we have thousands of reports the best way to match is manually. But spread-sheets or databases can help match by sorting the data sources in different ways to identify people on two data sources. For example, we can take two lists sorted by sex and date of birth to see if someone has the same name or initials; or look at lists sorted by sex and name to see if they have the same date of birth.

It is sensible to decide what constitutes a match – as differences between data sources may be due to keying errors or people giving slightly different names. For example, someone may be William Shakespeare in one data source and Bill Shakespear in another, and in one the date of birth was 26th April 1563 and in another it was 20th April 1563. The more data we have for any one person the easier it is to match, but we will still have to contend with possible errors between data sources. In general if the date of birth and sex is the same and one of the names is the same then most studies would include this as a match.

Whatever is chosen it is important that it is clear and specified. One way to do this is to adopt a rigid definition, do the analysis and then take a looser definition and compare the prevalence estimates. For instance, Mastro and colleagues in the Bangkok study had six criteria based on six different pieces of information for each person:

Match	Sex	Firstname	Surname	Age	Date of birth	Thai ID number
1	same	same	same	same	same	same
2	same	same	same	same	same	missing
3	same	same	same	same	Similar	??
4	same	Similar/ same	Similar/ same	same	Similar/ same	??
5	same	Different	Similar/ same	same	Similar	??
6	same	Similar/ same	Similar/ same	different	Different	??

By ‘similar’ in the above table the authors mean that the dates or names are close to each other, for example, that the month and year were the same. In the first case study above they used matching criteria 1-4 for the prevalence estimate.

Frischer et al in the study in Glasgow used the following matching criteria based on five pieces of information:

Match	Sex	First charcter of surname	First character of firstname	Date of Birth	Postcode
1	same	same	same	Same	same
2	same	same	same	Same	Similar/ different
3	same	same	same	Similar	Similar/ different
4	same	same	Similar/ different	Same	Similar/ different

Studies without routine data sources

What happens if we have no local data sources, or none collect any identifying information, or confidentiality rules prevent their use? Give up – don't do it, or adopt a different method? Perhaps consider doing capture-recapture calculations through surveys of problem drug users? Such a method is very well suited when the population is discrete geographically and have something in common that we can use to identify them: for example, street injectors, or street sex-workers that inject. We give two examples of this method (Case Studies 5.3 and 5.4).

The method used in Bangladesh is a simple two-sample capture-recapture study (outlined above). It did not collect any identifying information on the individuals but did identify the "matches" so it could use the calculations outlined above to estimate the total population. This method is classified as a capture-recapture study because the matches in data source 2 were genuinely identified as the people in data source 1.

The method used in Glasgow did collect some identifiers to match across individuals across separate nights of field-work, but used a different set of statistical methods to estimate the population size. We recommend that the use of such a method be considered only in consultation with a statistician who has experience of open capture-recapture models and estimation exercises.

Case Study 7	The Bangladesh Dhaka Study – estimating the prevalence of sex-workers

Capture-recapture without routine data sources

In this example, the method was used for estimating the number of street based sex workers in Dhaka city in Bangladesh. In absence of any secondary data on sex worker population in the city, and any routine data sources identifying this population, a through ethnographic fieldwork was conducted. Several categories of key informants were interviewed, which included sex workers, pimps, *rickshaw* and taxi drivers, police and local *mastans* (toughs). It was found that the sex work scene in the city was almost entirely street based. The sex workers solicited for clients in certain streets, corners and parks of the city particularly in the late evening. The whole city was carefully mapped for areas where sex workers solicit clients.

Reference

SHAKTI Project, CARE Bangladesh, Dhaka, Bangladesh

Case Study 8 The Glasgow street prostitutes study

Capture-recapture without routine data sources

In Glasgow, fieldworkers surveyed the main red-light district. They identified how many street prostitutes were working over a period of time, observing how many working on each night and if they had been observed on previous nights. In total over seven months 1145 contacts were made with 206 women (147, 71%, injectors). Identifiers consisting of initials and year of birth were collected for each woman contacted giving a "capture history" over the study period. Analysing the capture histories of the women using "open capture-recapture methods" suggested that the population remained constant at around 200 per night but that this population changed at approximately 8% per week giving an annual total of about 1150 prostitutes.

Reference

Frischer M, Leyland A, Cormack R, Goldberg DJ, Bloor M, Green ST, Taylor A, Covell R, McKeganey N, Platt S. (1993) Estimating the population prevalence of injection drug use and infection with human immunodeficiency virus among injection drug users in Glasgow, Scotland. Am J Epidemiol. 1993 Aug 1; 138(3): 170-81

Credibility of the estimate

How do we know if they are credible? The simple answer is that we don't. Things can go wrong (see A**ssumptions** below). But we can use the researchers' own local knowledge to judge whether the estimates are credible, which is how animal ecologists deal with the uncertainty of the method. So,

- Do the estimates fit with what we expected?
- Do they fit with the evidence base - estimates using other methods or from different years?
- Are they ridiculously low or suggest that 1 in 5 of the population is an injecting drug user – if so, they are probably wrong.
- Is the lower 95% confidence interval negative – if so, other methods can be used but check the accuracy of the data sources and the matching
- In multi-data source capture-recapture, are the dependencies between data sources believable?

Case Study 9 The Jersey study

The advantage of multi-data source capture-recapture is that if there are problems, they may lie with only one of the data sources – in which case that source can be dropped and an estimate recalculated. For instance, a prevalence estimation exercise was carried out as part of a Rapid Assessment in Jersey. Jersey is a comparatively wealthy island located in the channel between England and France and with less than 100,000 residents. Five data sources were collected:- a community survey of injecting drug users, heroin users in drug treatment, an addicts register where all doctors were asked to notify opiate and cocaine addicts to the government, heroin overdose visits at the accident and emergency department, and police arrests for drug offences involving heroin. The first model using all five data sources estimated that there were over 2000 injecting drug users, which would give Jersey a very high prevalence (i.e over 2% overall, and over 4% in those aged 15-54).

The authors suspected that interviewees were giving false names to the community survey. This was corroborated by comparing the level of matching between the community survey and drug treatment service (which identified 10% matches) with the proportion of respondents in the community survey that reported being in treatment in the previous year (over 30%). Clearly something was wrong, as we would expect around 30% matching between the two data sources (treatment and community survey). Dropping the community survey produced a much more reasonable and credible estimate of about 750 (0.8% overall).

This is an example of "misclassification error" (see **Assumptions** below). In addition, it points to the importance of local experience and knowledge so that estimates that are unlikely and potential biases can be identified.

Reference

Fitch., C., Stimson, G.V., Jones, S., and Hickman, M. (2001). Responding to drug and alcohol use in Jersey: key findings report. Drug Strategy Unit, States of Jersey.

Assumptions

Finally, we touch on the assumptions underpinning capture-recapture. These guidelines are intended to be mostly practical, rather than theoretical, so this will be only a brief outline. But we believe it is important for we to have some understanding of what assumptions are being made in order to interpret the findings of capture-recapture exercises. We discuss the implications of each assumption on the practise of capture-recapture.

Population closed – no deaths nor new cases nor migration into nor out of the study area during the study

Clearly this is impossible to guarantee. So the implication is that the length of the study is small in relation to the length of time people are problem drug users. For instance, a study over a one year period is usual, and would not lead to too much violation of this assumption. In comparison a study using five years or more of data risks including substantial numbers of problem drug users that were not using drugs over the full study period either because

- they ceased using drugs,
- they have died, moved out of the area,
- or only recently started using drug drugs
- or recently moved into the area.

In addition, the assumption suggests that when capture-recapture techniques at a time of high incidence - that is, at a time of rapid growth in new injecting drug use or problem drug use – the study should use as short a time as possible, or be postponed until growth has stabilized.

Open capture-recapture models are possible which estimate the dynamics of the population such as migration, death and births. These are usual in animal ecology but demand a different set of equations and statistical expertise. We know of only one study that has employed them in drug use (Case Study 9 above).

No mis-classification – individuals are correctly matched across all data sources.

In human studies this implies that we correctly identify someone who is on more than one list. We must do our best to ensure that the data we collect are accurate and reliable and hope that any false names given by problem drug users are either weeded out or consistent across data sources. Rules of confidentiality may disallow the collection and use of peoples' full names. Instead, information must be collected that allows matching across data sources without disclosing the identity of the person (see Section 9.2 on ethical issues involved in data collection).

In our examples, Mastro et al used the full name to identify matches; while Frischer et al in common with many capture-recapture studies of problem drug use used the person's date of birth, sex, and name initials (first character of first name and first character of surname) for matching. In the Bangladesh example, without named data sources, the study relied upon the recall of the street prostitutes.

In our study in Jersey mis-classification bias led to an over-estimate of the prevalence of heroin use and this will usually be the result. Because some of those interviewed in the community survey gave false names that could not be correctly matched with the names on other data sources, this lead to a number of true matches being missed.

No heterogeneity – all problem drug users have the same chance of being on a data source

Heterogeneity is inevitable in human studies. Some problem drug users because of their gender, age, ethnic group, or some other characteristic are just more likely than others to be listed on a data source. For example, in some countries young black male problem drug users are more likely to be arrested than other problem drug users. So we must live with the fact of heterogeneity. If there are enough people and enough information on them data we can stratify the population in order to make our estimate. We can do separate estimates for males and females, by or age group and ethnic group. If we have statistical support more complex models can be fitted (see Section 6 on advances in basic methods).

Representative – the data sources are representative of the population of problem drug users we want to estimate

There are differences between capture-recapture studies of problem drug use and some other epidemiological studies, like diabetes, which use as their data sources lists of diagnosed cases. Assuming that there is no mis-classification, then the study is aiming to estimate how many diagnoses relevant to each of the specific data sources in the study have not been reported. Hence, this would be a study of "under-ascertainment".

Do the data sources we are intending to use adequately define the target population of problem drug users? Defining problem drug use is different from defining, for example, diabetes. Firstly, it is defined by a range of clinical, social and criminal problems, which means that we should be as inclusive as we can, ideally gathering data from treatment and criminal justice data sources. Secondly, no combination of lists even if complete with respect to its own specific target will ever add up to the total number of problem drug users in the population. This is what makes

prevalence estimation important, exciting, and more difficult than it is for many other public health problems.

No dependence in two-source studies – the two data sources are independent of each other

Two sample studies are easier to do – but this assumption is their main limiting factor. In general, it is unknown whether two data sources are independent of each other and there is no way to test the assumption. If we do know the direction and rough size of the bias then two sample studies can provide very useful indications of the minimum or maximum size of the population. This is because "negative dependence" – the fact that if a person is on data source A they are then less likely to be also on data source B – means that the calculation over-estimates the true size of the population. Positive dependence – where a person on data source A is more likely also to be on data source B – leads to an under-estimate of the true prevalence.

Three sample studies can drop this assumption, but this increases the computational complexity. Log linear models are used to analyse the data, which in most cases means we need a statistical software package and may require statistical support. Multi source capture-recapture assumes that there is *no interaction* between all sources - i.e. if three data sources are used, the assumption is that there is no three-way dependence between all sources. However, if there is evidence of relationships between each pair of sources, this assumption of no three-way relationship may not be credible.

As the number of data sources used increases, the number of possible models that have to be inspected increases even more rapidly. So with three data sources 7 models can be fitted but with 5 there are 31 potential models. The one model that cannot be fitted is what is called the "fully saturated" model – where all data sources interact with each other. So, with three data sources a three-way interaction cannot be fitted, and with four data sources a four-way interaction cannot be fitted, and so on.

It is advisable, if practicable, to obtain at least one data source that is likely to be independent of all the others for any multi-source capture-recapture estimation study.

Summary

To recapitulate on the method: once we have got the data sources, match them; after matching put the data into a multi-way table and get an estimate; then turn the estimates of the number of problem drug users into prevalence estimates – i.e. as a proportion of the total population, the adult population (say, aged 15-54), or for males and females and age-groups separately, if estimates were made in these sub-groups.

We can summarise the key issues as practical advice as follows:

KEY ISSUES

- Be aware of potential problems
- Try to find multiple data sources if at all possible
- Select data sources carefully (try and get a range of different types of data source)
- Check that names are available and reliable
- Seek corroborating evidence – ask whether the estimate is credible? Does it fit with other estimates using other methods
- Improve the data sources for use next time.

Using Advanced Modeling Techniques

There is a further set of modelling techniques that are of a more complex nature, requiring a more statistical analysis than the simple MBM and CRC methods described in Sections 4 and 5. For using these, it is very strongly suggested that the help of an experienced statistician is required. In general, also the data demands are greater, in terms either of routine data or collecting data specifically for the estimation exercise. In this section, therefore, we summarise some of these methods and give some key references in case you want to examine the methods in more detail. We will briefly describe:

- Covariate modelling in capture-recapture
- Enhanced event based multiplier studies
- Truncated Poisson
- Back-calculation models
- Dynamic modeling

Covariate models in capture-recapture methods

Dr Kate Tilling has developed a variation on the capture-recapture methodology that allows covariates (such as age, sex, area of residence, ethnicity, etc) to be fitted within the modelling procedure to adjust the total prevalence estimate and derive separate estimates for the covariates. That is the modelling procedure tests for and adjusts for heterogeneity. Traditionally heterogeneity is dealt with by stratifying the data set into sub-groups and running separate models. But with each stratum more information is being used and it is feasible only if there are sufficient numbers of records to allow sub-group models to be run. The covariate model is much more efficient. But, the price for this advantage is that the modelling is more complex and demands statistical support. Moreover, examples using the method in drug abuse epidemiology have not yet been published. So it is a method to look out for in the future.

Reference: Tilling K, JAC Sterne. Capture recapture models including covariate effects. Am J Epidemiol 1999; 149 (4): 392-400.

Event-based and related models

Simeone et al. (1997) proposed a modified and enhanced multiplier study, which they piloted in Chicago, and is being used in in a border town in Mexico (Dr Elena Medina-Mora, personnel communication). The method uses "events" as the multiplier and benchmark. Before briefly describing the method consider the distinction between observed data that count a drug user – a person – and those that count repeated events – the number of times that person (and others) is observed. The events considered are usually institutional contacts or deaths: for example, data on multiple treatment attendances that record not only how many people attend in a year but also how many times each person attends in that year, although some 'events' of course are intrinsically non-repeatable (e.g. deaths). Person-counts usually can be regarded as a simple form of event data, where we work with an ever/never (in the time period) definition rather than a count. Sampling events is sometimes easier than sampling people if we want to collect personal or drug history information from the drug user. For example, if we sample people at a fixed site such as a treatment clinic then the 'attendance events' are fairly easily arranged to be a random sample of all clinic attendances, or at least a good approximation to that. As a sample of people, however, even if we count each person only once, no matter how many attendances are made in the period, we are of course more likely to be interviewing people who tend to turn up frequently than those who tend to turn up only occasionally. In addition, some routine data sources only collect events and do not have the ability to count the number of individuals, which is an important consideration for this method.

In the example in Chicago the benchmarks were the number of hostel attendances, the number of incarcerations, and treatment episodes by a heroin or crack-cocaine user. The multipliers were the annual rate of hostel attendance, incarcerations, and treatment episodes. Data for the multiplier and to adjust the benchmark were collected by interviewing samples of hostel users, prisoners, and those in specialist drug treatment. Self-reported use of drugs coupled with hair analysis to validate and adjust self-report were used to estimate how many incarcerations, hostel attendances, and treatment episodes were generated by heroin and crack-cocaine users.

Advanced statistical and modelling techniques were used to generate "unbiased" estimates of the event rates. Since each survey will be biased, the researchers used information from all three surveys to adjust each other to derive an overall estimate of the event rate for heroin and crack-cocaine users.

The method is theoretically interesting, but needs to be demonstrated in more sites. The results from the Mexico study should be available shortly.

Reference: Simeone R, Rhodes W, Hunt D, Truitt L. A plan for estimating the number of "hardcore" drug users in the United States. Washington: Drug Policy Research Group, Office of National Drug Control Policy, 1997

Truncated Poisson stochastic models of event counts

Truncated Poisson methods have been proposed. These use information on repeat events to estimate the size of the population with 0 events.

Note that this probably makes better use of data provided of course the counts are recorded accurately – something that usually requires a good memory on the part of the respondent to recall precise detail. The truncated Poisson is in fact the simplest of the stochastic models used in this field.

Restrictive assumptions

There are a number of assumptions required in adopting a truncated Poisson distribution as stochastic model. We need to:

- Assume repeat events are independent of event history of that individual. i.e., there is the same probability of being, say, arrested at every point of time throughout the study, regardless of previous arrest history
- Assume individuals all equally likely to generate an event; although we can relax this assumption with some difficulty if we have enough individuals with common values; (even then we need to assume something about the 'zero-attendance' group's personal history values, since if some individuals have an almost zero probability of being arrested then the procedure must of necessity miss counting/estimating them).

These unreasonable – and essentially un-checkable – assumptions make it dangerous to use as a method for estimating the number of 'hidden' drug users, probably less desirable than even a two-source CRC study.

Reference: Hay G, Smit F. Estimating the number of drug injectors from needle exchange data. Addiction Research and Theory.

Back-calculation methods

Law and colleagues adapted back-calculation methods to estimate the incidence and prevalence of heroin use. The back-calculation method developed in AIDS epidemiology works on the basis that the incidence of a relevant disease end-point (the incidence of AIDS, in the case of HIV) and the infection process resulting in the end-point are related through the incubation time

between the infection and the development of the end-point. Knowledge of any two of these three components allows estimation of the third. Typically, the distribution of the incubation time and the incidence of the end point are assumed known and the infection process underlying the observed incidence is estimated. The estimated infection process is then used with the same information on the incubation time to predict the incidence and prevalence of the end-point of interest. In the examples for drug abuse epidemiology the observed incidence of the end-point was opiate overdose death, with trends over time provided by routine mortality statistics. The incubation distribution is the distribution of the time between starting and stopping injecting, where the stopping process is the result of either a fatal overdose or the actual cessation of injecting. The data demands are considerable: including reliable mortality statistics to identify the number of opiate overdose deaths, and data on the opiate overdose and other drug related mortality rate of injecting drug users and information on the cessation rate from injecting.

Reference: Law M, Lynskey M, Ross J, Hall W. Back-projection estimates of the number of dependent heroin users in Australia. Addiction 2001; 96: 433-443; **and:** Hickman M, Shuying S, De Angelis D. Estimating long-term trends in the incidence and prevalence of injecting drug use (IDU) and the number of ex-IDU: the use of opiate overdose deaths and back-calculation methods. Forthcoming.

Dynamic models

Finally, one rapidly developing area of research into prevalence estimation methods is the field of dynamic models. Broadly speaking these require a great deal of data that might be described as drug indicator information taken from as wide a range of different data sources as can be found. These data are analysed jointly in a single dynamic model, in which the principle structure is determining how data at one point in time can lead to the observations at a subsequent point in time. These are specialist models that need expert tailoring to each situation.

Extrapolation from local to national prevalence estimates

The preceding sections describe specific methods for estimating prevalence. Extrapolation methods are not a specific method of prevalence estimation in themselves, but when some prevalence information is known they are used to extend this information into areas - usually, literally, other geographic regions – where the prevalence information is not known.

For the most part the specific methods described in the preceding sections are most easily applied on a local level - that is, within a relatively small geographical region, sometimes even a single city. This is partly because data sources required for estimating drug use prevalence are often more readily obtainable and able to be manipulated on a local level. For instance, capture-recapture methods require matching a list of names - or some other form of identification - of drug users from one data source with a list of names from another. This matching of identifiers can be an arduous task, if there are no automated computer facilities, even at a local level and nationally the task might be prohibitively difficult. Moreover, it is likely that local drug users are only found on local sources in the main, and searching national lists for them is a wasted effort.

Furthermore, the type of data available and their method of collection are sometimes consistent only on a local, regional level. For example, if we consider the use of treatment clinics as a source for research data, then in highly urban areas of a country there may be more information available from treatment clinics than is available in more rural areas. It may be that in some regions a service such as the provision of specialist treatment clinics is completely absent and

drug users rely instead of on the government's general hospitals and associated institutions where there are less detailed records. In particular, where treatment is a privately run enterprise, the quality of record keeping in different areas or cities may vary considerably from one organisation to another, with no overall structured uniformity nor consistency.

Finally, there is sometimes a question of the *ownership* of data that is difficult to organise at a national level. In particular it is often the case that in more rural areas policing records are kept differently; but there is sometimes that the data are owned by the local authorities and may not be combined or aligned with data from other localities. The same is true of data on deaths in some countries – at least, information regarding deaths of addicts.

All these reasons and many more suggest that it is much easier to mount an estimation research study at a local level. The ability to mount a nation-wide exercise may therefore be very limited. Of course, in many practical ways it is the local figures for drug use and the attendant problems that have the greatest usefulness and importance. These local figures are in their own right a vital piece of information. Nonetheless their are obvious advantages in being able to produce a national figure as a sort of summary of the overall position - an indispensable aid to central policy making. The question then is whether the results and estimates of prevalence from local studies can be used to extrapolate to a national prevalence figure. It is worth noting specifically that the same considerations apply when, for example, we want to generalise results from studies in specific cities to give estimates for the region in which the city is located.

The important element of synthetic estimation and any other extrapolation method is that it makes use of known prevalence figures in certain regions to estimate prevalence in other regions To do this these regions must have some data sources that are the same as (or very similar to) the regions for which prevalence estimates exist, although of course they lack the regional prevalence figure itself. The general principle is then to use data that are similar across the separate localities to project figures for drug use prevalence from localities where it is known to localities where it is lacking.

Extrapolating from a single local estimate

We consider first an example of the simplest kind – extrapolating from one local area to one other area, although in fact this area is in this example the rest of the whole country (see Case Study 10)

Even in this basic example there are several fundamentally important points demonstrated.

Firstly and most obviously the researchers choose not to simply apply the prevalence rate for drug abuse in NSW (termed the 'anchor point') to the rest of the country (sometimes termed the 'target' area), where it is unknown. When can regional prevalence rate can be considered also the national rate? or the rate for another area?. Many factors, such as varying social structures and demographic distributions, as well as extent of urbanisation and proximity to drug supply routes, can make this assumption that one local area is typical of another too simplistic to be useful. In this case study, a simple *drug abuse indicator* was used instead to make the extrapolation.

The data structure required can be thought of in the following manner, shown in Table 7.1. In the anchor point area we know how many people there are and how many overdose deaths there are and how many drug users there are, and this enables us to calculate the prevalence rate - say, per 100,000 of the adult population - for overdose deaths (the indicator) and for drug abuse. In the target area we have information that allows us to calculate the indicator – the overdose death rate per 100,000 of the adult population, but need to estimate the prevalence rate for drug abuse.

Case Study 10 Extrapolation in the NSW Study

While a local study had produced an acceptable estimate for New South Wales of approximately 37,000 of regular opioid users (see case study 2 and 3), there was also considerable interest in extrapolating this finding to the whole of Australia.

One very simple way to do this would have been just to multiply our estimate of the prevalence for NSW by the number of people in Australia. This would effectively assume that the prevalence per 100,000 of the population in NSW was the same as the prevalence rate for Australia as a whole. Therefore, given that approximately one third of the entire Australian population reside in NSW, it would be concluded that there were 111,000 (3 X 37,000) people in Australia who were regular heroin users.

A major problem with this approach, however, would be that it ignores potential regional differences in rates of heroin use by assuming that the proportion of people who regularly use heroin is the same in NSW as it is in the rest of the country. However, it is well known that rates of heroin use have historically been much higher in NSW, and particularly in Sydney, than in other parts of the country. National data on both methadone maintenance treatment and fatal heroin overdoses have consistently shown that approximately half the people entering treatment for methadone and half of those dying from heroin overdose are from New South Wales.

Faced with the choice of assuming that the numbers of regular heroin users is proportional, geographically speaking, either

a) to the size of the general population, or

b) the numbers heroin overdoses recorded,

the second is clearly preferable because it is a more specific relationship. The first assumes that the prevalence of use is constant geographically, and the second assumes that the fatal overdose rate is constant geographically.

Therefore, we used this multiplier (of 2.0) to estimate the number of people in the whole of Australia who are heroin dependent. This gave an estimate of 2 x 37,000 or 74,000 people heroin dependent in Australia as a whole. This figure was then used to calculate the prevalence rate for the part of the population that was thought to be almost exclusively at risk of overdose death, those aged between 15 and 54. The figure of 74,000 then equates to a population rate (per 1,000 people aged 15-54) of 6.9.

References

Hall, W., Ross, J., Lynskey, M., Law, M., & Degenhardt, L. (2000). How many opioid users are there in Australia? *Medical Journal of Australia,* **173**, 528-531.

Lynskey, M. and Hall, W. (1998) Jurisdictional trends in opioid deaths 1988-1995. *Australian an New Zealand Journal of Public Health,* **23**,519-524

Technically the steps involved are:

(i) to calculate the drug indicator prevalence rate for the anchor (a) and for the target (b);

(ii) to calculate the drug abuse prevalence rate for the anchor (c);

(iii) to extrapolate from the anchor value to the target by assuming the relationship between drug abuse and the indicator is the same both for anchor (c) and target regions (d);

(iv) to calculate, if we want them, the actual numbers of drug abusers, as opposed to the abuse rate, in the target area (e).

(v) to calculate overall prevalence rates, if wanted, by first summing the estimated count of drug users in each area, then summing the population figures, and then taking the ratio.

Calculations made in this way will mimic the calculations of case study 10; they also show the assumption involved and enable us to generalise the procedure in several ways - as discussed in points 2) and 3) immediately following.

Data structure for extrapolation from known anchor point drug prevalence levels to target areas where it is unknown

	Population	Drug Abuse Indicator (Overdose deaths)		Drug abuse prevalence	
	Numbers	**Numbers**	**Rate**	**Rate**	**Numbers**
Anchor point					
NSW	known	known	calculated (a)	calculated (c)	known
Target					
Rest of Country	known	known	Calculated (b)	unknown (d)	unknown
Overall					

Secondly, then, we can see that although only one area is used as a target, it would be a simple matter to list several target areas in the lower part of the table, provided we have the drug abuse indicator information for them, and to produce different estimates for, say, each territory in Australia.

It is also possible to add in several more anchor points to the upper part of the table if drug indicator and drug abuse information is available, with the aim of establishing a broader base to the relationship between drug abuse prevalence and overdose deaths prevalence. This raises a new element – how to combine the different pieces of information offered by the anchor points that have been added. In fact, this is a standard statistical regression analysis problem, and can be easily handled by any computer software package. Essentially, an average relationship between the two rates is established across the anchor points, based on the assumption that the same relationship holds in all of them. In case study 10, the calculations in the example are greatly simplified by having only one anchor and one target point.

Thirdly, it is clear that the role of the drug abuse indicator in these calculations is non-specific, that is, any indicator no matter how it is derived will do provided that

(i) it is relevant to drug abuse, and is likely to have a stable or uniform relationship to it across anchor and target points, and

(ii) we have the information for it to be calculated in a uniform and parallel way across anchor and target points.

Finally we note that if there are several different estimates of prevalence from several different local studies for a single area that can be used as an anchor, then there is no difficulty in simply including each estimate of the prevalence rate as a separate anchor point in Table 7.1.

Before moving on, we should note that there are some technical, statistical considerations that apply once we use more than one anchor point in a regression analysis.

(i) There is a question about whether we should use a Poisson regression, which is common when analysing rates of occurrence or prevalence, as opposed to using simple linear regression analysis; and if the latter, whether we use the rates themselves as data or whether we use log-transformed data.

(ii) There is a question about whether we can establish a weighting for the different anchor points to represent the reliability of the drug abuse prevalence estimate that is supplied as data to the extrapolation exercise.

(iii) There is a question about whether the relationship between the drug indicator and the drug abuse prevalence rates are similar across all the data points, whether some are too different from others and should be excluded, and how this might be checked. This applies not only to similarities between the anchor points, but also to similarities of anchor points to targets – if some target areas are very different on the drug indicator dimension, then the extrapolation may not be reliable nor valid.

(iv) There is a question about whether the assumption that the relationship between these two prevalence rates is valid, whether it is strong enough to be used at all in a regression analysis and how this can be checked in the data.

These questions can be usually be tackled and should ideally be discussed with an experienced statistician.

Extrapolating using several drug indicators

We add one more generalisation to the method outlined in 7.1 for extrapolation by using regression analysis, one that picks up on the role of the drug indicator measure used in that example. The extended method to be described comes under various headings: usually, 'synthetic estimation', or 'multi-indicator' method, or sometimes under the more technical name of 'regression on principal components'.

The intention is to use a variety of drug indicators available in the anchor and target areas in order to improve the predictability of drug abuse prevalence in the target areas. Case study 11, following, sets out to use rates of police drug seizures, drug-related convictions, drug clinic treatment attendance, HIV tests by injecting drug users and drug-related deaths. All of these are represented as rates per 100,000 of the general (adult) population and are available for all anchor and target points. To the reader who is familiar with regression methods, this initially sounds as though it is a straightforward prediction of drug abuse prevalence in a multiple regression equation using five separate predictor variables – an approach that works if there are a large number of anchor points in the data set. Usually in this type of indirect estimation situation, though, there are very few anchor points from which to extrapolate, which results in highly unstable multiple regression predictions and so an amended regression procedure is required.

Since all the drug indicators are, hopefully, related to a central 'drug abuse index', the separate indicators are combined into a single indicator 'index' across all the data points, achieving a good summary, again hopefully, of the set of indicators. This is carried out using a principal components analysis (PCA), which is designed specifically to just that: produce a good, single summary index of the set of indicators. The regression analysis, which has to establish the relationship with drug abuse prevalence rates based on the data from the few anchor points alone, then uses just the one index as a predictor and can give more reliable predictions for the target areas.

Case Study 11 Extrapolation in a U.K. study using multiple indicators

The data available for this study are shown in Table 7.2, where four anchor points for which there is some knowledge of problem drug use rates are to be used in extrapolating to seven target points spread across Great Britain. Note that the anchor points that are used are themselves spread out over the countries as a whole.

Aims

The European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) has produced methodological guidelines for national drug prevalence estimation. This paper pilots the methods to estimate prevalence for Great Britain and provides a commentary on the methods and resulting estimates. Three types of methods were used to estimate prevalence: a) the Multiple Indicator (MI) method, b) multipliers applied to (i) drug-treatment records (ii) HIV estimates and (iii) mortality statistics and c) the British/Scottish Crime Surveys. This case study reports only to the multiple indicator section.

Definitions

The definition of "problematic" drug use in this study follows that of the EMCDDA working group which is "intravenous drug use or long duration/regular use of opiates, cocaine and/or amphetamines. Ecstasy and cannabis are not included". This definition is suitable for studies based on routine sources which do not normally record detailed information about types of drug use. It would be unrealistic, for example, to propose using the current International Classification of Diseases (ICD-10) definition of drug dependence which in part consists of "a cluster of behavioural, cognitive and physiological phenomena that develop after repeated substance use".

Method: The Multivariate Indicator for problem drug use

The aim of this method is to estimate the number of problem drug users in the population by combining information on prevalence that is available only in a few areas (the calibration samples, or 'anchor points') and 'indicators' or 'predictors' of drug use that are available in all areas (Mariani et al., 1994). This method was first used in the United States (Woodward et at, 1984) and has been described more fully elsewhere (Wickens, 1993).

The key assumption of this method is that the relationship between prevalence (dependent variable) and the predictors (independent variables) in the calibration sample is transferable to all other areas. It is also assumed that a single factor underlies the drug related indicators and Principal Components Analysis (PCA) can be used to extract the main factor which explains the largest amount of variance in the indicators.

The steps below summarise the process used to analyse the G.B. data.

Step 1.

We obtain a range of indicators or predictors of the prevalence of problematic drug use that are available for all geographical areas within the country. Great Britain was divided into eleven regions: the regional health authorities in England (1-8), Wales (9) and Scotland divided into Strathclyde (10) and the rest of Scotland (11). The division was based on pragmatic grounds, reflects the availability of the required data.

Step 2.

General population figures for each of the 11 geographical area are obtained from census information.

Step 3.

The following drug abuse indicator variables were collected for all regions the UK pilot project for the time-span of the preceding year:

(A) Convictions for drug offences

(B) Seizures of controlled drugs

(C) People treated for drug abuse as recorded in Regional Drug Abuse Databases

(D) Cases of Human Immunodeficiency Virus (HIV) related to injecting drug use (IDU)

(E) Drug related deaths

Additionally, existing estimates of drug abuse prevalence were obtainable for four regions: North Thames, West Midlands, Wales, and Strathclyde in Scotland. These provide four anchor points for the extrapolation. The data and the sources of these estimates are shown in the Table below.

Drug abuse indicator data for 11 region of Great Britain and drug abuse figures for four anchor regions

Regions	Population nos.	Drug use indicators (See Key)					Drug use obs. nos
		A	B	C	D	E	
England	47,055,204	83,533	92,095	51,850	788	2,371	
North Thames /a	7,190,479	17,696	21,168	7,842	334	352	44410
West Midlands /b	5,150,246	7,125	5398	4,322	26	193	13130
Northern & Yorks	6,600,626	11,356	13,285	9,722	37	344	
Trent	4,606,495	6,451	7,010	3,580	67	207	
Anglia and Oxford	4,521,912	3,761	4,183	3,762	79	216	
South Thames	6,579,403	13,987	16,530	7,774	122	346	
South West	6,131,705	10,600	12,717	5,890	60	311	
North West	6,274,338	12,557	11,804	8,958	63	402	
Wales	2,835,073	6,110	5870	2,282	14	139	
All Wales /c	2,835,073	6,110	5870	2,282	14	139	8357
Scotland	5,134,105	3,008	13,452	8,614	687	267	
Strathclyde /d	2,283,671	943	7,989	4,331	97	127	18000
Rest of Scotland	2,850,434	2,065	5,463	4,283	590	140	
United Kingdom	55,024,382	92651	111417	62746	1489	2777	

Key to table:

A	Convictions for drug offences, 1996
B	Seizures of controlled drugs, 1996
C	People receiving treatment for drug abuse England and Wales (regional drug abuse databases October 1996-March 1997); Scotland (April 1995-March 1996)
D	Cases of HIV related to IDU England and Wales 1996, Scotland 1995
E	Drug related deaths, UK 1995
F	Population
G	Estimated no of problematic drug users (anchor points only)
H	Model Estimate based on Principal Component Analysis
a	Population survey: ever injected drugs, 1991
b	General Practitioners' computer records: extrapolation from diagnoses of drug abuse/dependence, 1996[Error! Bookmark not defined.]
c	Capture-recapture study of 'serious drug users', 1994, [Error! Bookmark not defined.]
d	Projection from computerised drug prevalence estimation program [Error! Bookmark not defined.]

Step 4

At the next step, each variable, (A) to (E) and each anchor point drug abuse figure is converted into a rate per 100,000 of the population.

Step 5

We derive a drug abuse indicator for each geographical area using Principal Component Analysis (PCA) with the five indicator rate variables. The analysis was checked to see that the derived index represented a satisfactory amount of the variation between regions in the indicator variables (in excess of 45% in this instance), and whether a second index should be used. The first component showed high correlations with the crucial indicators: convictions (0.049), seizures (0.842), treatment (0.789), HIV (0.147) and deaths (0.864). In contrast, component 2 (not used in this analysis) is more strongly associated with convictions and HIV.

Step 6

Finally, regression analysis is used to on the anchor point data (known drug abuse prevalence rate regressed on the calculated indicator index); the regression model is then used to make predictions of the estimated level of drug abuse prevalence rate in each of the seven target areas. This is then converted into an estimate of the number of drug abusers in each region.

The data for the multiple indicator method are shown in the Table below.

GB data and multiple indicator prevalence estimates, 1996.

	Population Nos	Indicator index	Calc. Obsd. drug rate	Extrap'd drug rate	Extrap'd drug nos
England					*(225,358)*
North Thames /a	7,190,479	57	0.618%	0.573%	41,213
West Midlands /b	5,150,246	19	0.255%	0.187%	9,643
Northern and Yorks	6,600,626	56		0.557%	36,786
Trent	4,606,495	31		0.313%	14,410
Anglia and Oxford	4,521,912	28		0.279%	12,600
South Thames	6,579,403	58		0.581%	38,234
South West	6,131,705	47		0.473%	28,997
North West	6,274,338	69		0.693%	43,475
Wales					*(12,629)*
All Wales /c	2,835,073	45	0.295%	0.445%	12,629
Scotland					*(30,221)*
Strathclyde /d	2,283,671	75	0.788%	0.749%	17,110
Rest of Scotland	2,850,434	46		0.460%	13,111
United Kingdom					*(268,208)*

There are two technical points we should draw attention to in this example. Firstly, initial inspection of the drug indicator data over the 11 data points shows that drug-related convictions are *negatively related* to all the other drug indicators, mainly as a result of the unusually low conviction rates in the two Scotland regions. This is a plausible scenario, in that where convictions are high there could be a resulting drop in visible drug use, but it does open up the possibility of alternative analyses. One option is simply to omit the convictions indicator and use PCA to produce a single index of the remaining four, positively correlated indicators and proceed as before. Table 7.2 would then be based on an indicator index derived from only four indicators.

The published paper however take a more sophisticated route and identifies *two* summary indices from the set of five variables, using a PCA that requires the extraction of two obliquely related components, one of which primarily covers 'convictions' and the other 'treatment'. Whether one or both of these indices is subsequently used in a regression analysis is a matter of judgment, but with only four anchor points for the analysis it is probably better to lean towards using just one index, as the researchers report in Tale 7.3. Again in this data set, as chance would have it, there is no great difference in the resultant overall extrapolation whichever of these alternatives is implemented. These more complex amendments to the basic procedure should however not be attempted without experienced statistical advice: this point cannot be too strongly stressed.

A second point is that we can make a direct comparison of the extrapolated - predicted - values of estimated drug prevalence derived from the regression model *for the anchor points themselves* with the actual data provided for these points. This is a useful side check on whether the regression model is doing well its job of establishing the indicator-abuse relationship, and in general terms is called 'residuals examination'. The researchers point out that in this case study the observed drug abuse prevalence rate value for Wales is not very well fitted at all by the model and since this is one quarter of the anchor point data, this must give some cause for caution when interpreting the results. Examination of residual should be a routine part of any extrapolation exercise.

Limitations

The method assumes that the unobserved prevalence is related to the observed indicators, and that the relationship between the indicators and the anchor points is similar for other geographical areas. However, other factors also have a bearing on the indicators that may invalidate this assumption and the derived results. For example,
a) the number of drug users in treatment may be restricted by the capacity of treatment services, or be affected by the level of under-reporting that can vary across the country,'
b) the level of policing and attention given to drugs offences may vary across the country; and
c) the prevalence of HIV infection among injectors varies by geographical area,
d) the variables should be reported by geographical area of residence, whereas some (e.g. police statistics) are reported by area of report.

Reliability and validity of estimates for the anchor points is of critical importance. While we used four anchor points for the UK study they were obtained through various techniques that are subject to various assumptions we were not in a position to evaluate. Furthermore, only two of the four estimates were for the same time period as the indicators (1996).

One of the key findings in the present study is the regional variation in estimates of problematic drug use arising from the multiple indicator method. The estimates range from 19 per 1000 population in the West Midlands to 75 per 1000 population in Strathclyde (see Table 7.3). In interpreting this range, it should be remembered that the estimates are dependent on the available anchor points which may represent different forms of problematic drug use. In two areas there is a considerable difference between the anchor point observed prevalence and the MI model prevalence predictions: in the West Midlands and in Wales.

Conclusions

Of the methods used in the pilot study, the multiple indicator method produced the highest, but perhaps the most valid, estimate for problematic drug use in Great Britain. The method is cost-effective as it does not require new data collection, unless separate studies are needed to estimate new anchor points. One problem encountered during the study was obtaining data from multiple government and other agencies. Sub-national estimates were made for Regional Health

Authorities because it was not possible (in the study time frame) to obtain data for smaller more meaningful populations within the UK. Regional Health Authorities are large heterogeneous populations containing urban and rural parts, and, therefore, dilute geographical differences. In future, it would be better to examine other ways of aggregating populations in order to reveal geographical differences e.g. to separate inner-London, and outer-London from Thames RHA, and to have separate estimates for other large cities such as Liverpool, Manchester and Birmingham.

In conclusion, the current study suggests that a more differentiated response to the problem of drug abuse may be possible, although more work is required to provide more detailed breakdowns in terms of demographic characteristics.

Reference

EMCCDA Project (CT.97.EP.04) entitled 'Study to obtain comparable national estimates of problem drug use prevalence in all EU member states'.

Frischer, M., Hickman, M., Kraus, L., Mariani F. and Wiessing, L. A comparison of different methods for estimating the prevalence of problematic drug abuse in Great Britain

Frank, B, Scmeidler, J., Johnson, B and Lipton, S. Seeking truth in heroin indicators: the case of New York City. *Drug and alcohol dependence* 1978; 3: 345-358.

The example as published uses seven target and four anchor points. Although subtotals are eventually required for Scotland and for England, the data used in the principal components analysis should be only the 11 distinct areas, excluding any subtotals of these areas.

The published study uses all information transformed into rates per 100,000 of the population of each area (whether per 100,00 or per 1000 or per 100 makes n difference to the analysis, provide it is a *rate* and not an absolute count that is used), and the analysis uses these rates directly. An alternative would be to use logarithmically transformed rates. Although in this example doing so makes little difference to the extrapolated values, it is often safer when analysing rates to work with log-transformed data from the outset, in particular if the rates cover a wide range. If nothing else, this step ensures that estimates of problem drug use rates are greater or equal to zero (i.e. are technically plausible estimates).

Assumptions

In conclusion, we note that the use of synthetic estimation – or multiple indicator estimation – can have many minor variants in its application, depending on what data is available and on what aims the estimation exercise has. For example, since only prevalence rates are used in the analyses, extrapolation can be based on just these, if they are obtained without knowing the associated absolute counts. The basic assumptions however are clear in all these potential variants:

- analysis should be base on prevalence rates of indicators and problem drug use
- it assumes that there is a stable relationship between this prevalence and the indicator set, and this implies too that the indicators are defined and constructed in a similar manner for all data points
- the extrapolation is weaker in its validity and reliability if some data points do not conform to this presumed relationship
- the extrapolation is dependent on the form of the regression model fitted (Poisson, linear, log-transformed)

- the extrapolation is weaker, with doubtful predictions, if more than one indicator is used, unless there is a large number of anchor points ("at least three for every index used" is a good rule of thumb)
- the principal component analysis should have an adequate number of data points to support it ("at least two to three times the number of indicator measures" is a good rule of thumb).

Accuracy of Estimates of Prevalence Levels

Indirect estimation methods can be very inaccurate. It is essential therefore to assess the accuracy of any results in some way. Traditionally statisticians distinguish between reliability and bias. 'Reliable' means the method will give the same result if it was repeated on the same population in the same way time after time. The results are reliable if any small discrepancies between different occasions can be explained by chance i.e because different samples select different individuals and although they are representative of the total population, the individuals by chance may differ slightly between samples.. 'Bias' means that there will always tend to be a discrepancy between the result and the true answer, no matter how reliable the method is. The extent to which bias can be produced by the assumptions of the method failing to be reflected in the actual research situation is called the 'robustness' of the method - how fragile the method of estimation is when dealing with situations for which is was not designed.

Reliability is, at one level, determined intrinsically by the method of estimation, but it also fundamentally depends on the scale of the study, the smaller ones in general being less reliable than the larger ones. In indirect estimation methods a good crude guide to how unreliable a prevalence rate estimate is likely to be is the ratio of how many drug abusers are *estimated* to exist in relation to the number of records of drug abusers used in the estimation exercise. Particular methods of course vary within this broad rule of thumb and other related factors come into play. For example, mortality multiplier studies which tend to multiply the number of deaths by 100 or more are likely to be less reliable than multiplier studies with more common benchmarks and lower multipliers. Equally, the more data and overlap between data sources found in capture-recapture studies the more robust the estimate. Given, of course, that the study is not "biased".

In the matter of bias though, the considerations are different. Study size is of little help in removing the distortion from the estimates produced by a method. For this reason most methods in common use are developed because they produce unbiassed results when properly applied. Bias is introduced because of a failure to maintain good practice in applying the research procedures and because breaches of the assumptions on which the statistical analysis depends go unchecked or unobserved. Good practice in research methods, is covered in the next two sections of the manual. Breaches of assumptions for an analysis are discussed under Robustness immediately below (Section 8.3).

But first, in the next paragraphs, the use of formal statistical procedures for establishing confidence intervals around an estimate of prevalence are outlined.

Confidence intervals from formal statistical theory

A confidence interval (CI) estimate simply gives a range of values within which it is likely that the true answer - the actual population prevalence - lies. This most helpful presentation of reliability - the wider the interval then obviously the less precise the estimated value - gives an indication only of sampling variation - the chance deviations of the estimate from the true value. There are standard formulae for setting confidence bands around the spot estimates of prevalence

derived by most statistical procedures; a computer software package will produce them as a matter of course. A CI can be expressed in terms of absolute numbers of abusers (e.g. 12,000 to 27,000) or as a factor applied to the estimate (e.g. 18,000 within a factor of 1½). Note that the validity of CI calculations depends on viability of the assumptions used in the analysis, and that it only reflects sampling variation, not bias. For this reason, given that indirect prevalence estimation methods are subject to many potential biases, CI are less important than using multi-methods or comparing any prevalence estimate with others.

Multiplier/benchmark confidence intervals

The calculation of CI for an estimate of prevalence derived by multiplier methods can done easily if the estimation is considered as a Poisson (log-linear) regression procedure as described in Section 4. Calculating the estimate through a computer software package will automatically give a CI. Broadly speaking when estimating the number of drug abusers in a country or region the CI (expressed as a factor) indicates that an estimate becomes more unreliable with increasing size of the multiplier, and with increasing unreliability in the value of the multiplier itself. The latter factor - unreliability in the calculation of the multiplier itself - can be gauged crudely from the size of the sample study that derived the multiplier: the larger the better.

The requisite information for calculating a CI is not often available and few multiplier/benchmark studies are able to quote one. This is not though a really serious drawback since there are more damaging forms of uncertainty in an estimate than sampling variation, as will be discussed in a following section (8.3) under Robustness.

Capture-recapture confidence intervals

For capture-recapture studies the CI can be calculated most simply for models of any complexity by the computer software package used to fit the model. Bishop et al give equations for calculating CI by hand. Confidence intervals can also be calculated using the "likelihood interval approach" or "goodness of fit" approach where values for the lower and upper CI are derived by trial and error based upon getting the nearest values for the "unobserved" population that give a difference of 3.84 (95%) in the G^2, which is a measure of the fit of the model. See Hook and Regal (1995) for further discussion of this method. . In capture-recapture estimation the statistical measures of uncertainty are increased by smaller sizes of the samples taken at the capture points, smaller proportions and numbers of matches between data sources, and the complexity of the model required to describe the data.

Again, it is important to note that the CI says nothing about the viability of the model and therefore is not the only thing to consider in determining how certain the estimate is as a measure of the true population figure.

Confidence intervals around any estimate can usually be calculated analytically, as described in Section 8.1, or very often, as an alternative, by data simulation methods, which make fewer assumptions than analytic methods. Although Monte Carlo simulation techniques and other data-led methods of assessing reliability could be applied, in practice this is seldom done in practice. General purpose data-led methods such as bootstrap and jack-knife estimates are plausible but not always satisfactory methods of assessing reliability. In general these methods are labour-intensive and tend to make so much work that the actual analysis becomes a minor part of the effort required. As a result the methods are very seldom used; and if they are, the services of an expert are always required to assist in the programming.

Robustness of prevalence estimates

The standard survey methods of calculating confidence intervals that are available depend upon the validity of the assumptions of whatever indirect estimation method is being used. However it is not sampling variation but robustness that is the real issue and departure from assumptions the key difficulty.

It is much more likely that errors in the resultant estimates are not for the main part simply sampling variation in the traditional statistical sense, but are generated by failures of the situations to meet the required assumptions. These failures tend to produce biassed results that can sometimes be seriously distorted, and the resulting lack of accuracy in the estimation procedures is not something that can be addressed by the standard confidence interval estimation approaches.

Traditionally, multiplier/benchmark methods seldom use statistical theory to derive confidence intervals. Instead, an upper and lower estimate of prevalence is generated by varying the multiplier or the benchmark, based on information about their uncertainty. For example, Hartnol et al. (1985) used a mortality multiplier of 100 - based on a presumed mortality rate of 1% - and then repeated the calculation with a multiplier of 50 - based on a presumed mortality rate of 2%. In this way a range of estimates is produced that is not related to confidence interval theory, but instead reflects in a completely informal manner the uncertainty in the information being used.

Given the number of assumptions involved in CRC and the potential for violation it is always possible that the model used in the analysis is wrong, despite the best endeavours of the researcher, and the estimate therefore is not a true reflection of the prevalence. Specific assumptions that lead to difficulties have been highlighted in the appropriate sections on the method, for example:

- the dangers of using only a two-capture CRC design
- the necessary assumption that at least one (highest-order) interaction is null

There is also a standard test in the statistical literature for checking whether log-linear Poisson models are in fact adequate for describing the data (Pregibon's test).

There has been little formal work done on the precise extent to which these different indirect methods of estimation are sensitive to departures from assumption, but a general consensus appears to be that estimates can go badly wrong under adverse circumstances. As a result it is common practice to take a very simple approach to the question and to look for concordance and convergence in the estimates made by different indirect procedures. By and large, the accuracy of these approaches is only judged by the extent to which they converge to a common estimate in any situation. Again, formal criteria for judging such convergence and reconciliation of different estimation procedures are not available, so the judgement is informal.

Given than that concordance of different methods probably gives best indicator of a satisfactory estimate being derived, the following recommendations can be made:

(i) use multi-methods:

- capture-recapture and multiplier/benchmark methods if possible
- use multiple multiplier methods, ideally with the multipliers generated from more than one source (in case the sample used to derive the multiplier is biased)

(ii) use different models within analysis of any one set of data to give a potential range of answers:

- in CRC, comparison of different sub-optimal log-linear models

(iii) look for plausibility and consistency of estimates of different behaviours or across different sub-populations:

- injecting prevalence and heroin use prevalence should have some relation, unless there is good reason to think otherwise (a good reason might be that heroin smoking is the main route of administration, for instance);
- the model selected in capture-recapture estimation in terms of the proposed interactions between data sources should be credible.

4 Guidelines for producing research based estimates

General guidelines

The data used in the indirect estimation methods described in this manual are obtained by two means. The first of these is data collation from existing sources; the second is specialised primary data collection for the research projects. The corresponding responsibilities involved in obtaining the data are different, but there is an important overlap at a very general level that is worth recommending: any research undertaken should provide an opportunity for co-operation and for building research capacity. The intention is to use any research study to pave the way for future work and to lay down the building blocks to make it easier. The following sections therefore stress the importance of consultation with other drug-related researchers and other institutions, partly to fulfill the immediate research priorities and partly to develop networks of interested parties for research in the future. Using a research study in this way, taking the opportunity for co-operation, should encourage the building of a sustained capacity for drug abuse monitoring and research through the development of a common body of knowledge and information and through the establishment of channels for co-operation amongst otherwise isolated groups of workers. It is in this spirit that the following broad guidelines to 'good practice' are formulated.

Minimum requirements in a prevalence assessment project

Good quality research cannot be undertaken lightly. The mix of skills required is considerable, and a team undertaking is essential. The following check-list of skills, capacities and general expertise is of course only a broad guideline, but the range of requirements that might be considered minimal speaks for itself.

CHECK LIST	**Minimum requirements in expertise and resources:**
	• contact with the international scientific (epidemiology) community at large;
	• expertise in project management and planning ;
	• an advisory panel or group
	• consisting of key savants with previous experience;
	• statistical consultants available during the design stages;
	• questionnaire design skills;
	• field workers with experience in drug research or similar work;
	• training of field workers;
	• computer access and computer skills in data entry and data file design;
	• statistical advice in later analysis;
	• web-based or other discussion groups with interested other researchers;

Many of the items listed on this check-list would be obvious to any researcher, some perhaps may only have been thought of after the event?

The first point – contact internationally – is more important than just being an obvious background requirement. A great deal of expertise, experience and knowledge has been built up in drug research bodies around the world and being in touch with this source at the outset is vital – it is vital in helping establish the aims of a new project and in helping with the fundamentals

of design and in assessing feasibility. To embark upon a new project that cannot be done is as un-necessary as embarking upon a new project that has already been done in some other form; to embark upon a design that is difficult when an alternative design achieves almost the same end much more efficiently is making unnecessary difficulties that might endanger successful completion. The particular factors, the particular situations, the particular conditions of any project to assess prevalence are far too varied to allow generalised recommendations in a manual of this sort and discussion with people with prior experience cannot be underestimated in its importance. It is hoped that the numerous examples in part B of this manual give some idea of the different local conditions that need to be accommodated in a project and to which the design must be adapted.

The need for expertise in management and planning is something that ought to be self-evident, but is one of those skills that is taken for granted too often. Research is not easy. Drug prevalence research is hard. Forward planning is only part of what is required – meeting contingencies and being sufficiently flexible without compromising the validity of the methods is a skilful balancing that always requires help. The use of an advisory panel of experts to help in both in setting out the project's aims and methods and in assessing the project's progress is very strongly recommended. This panel should be able to assist in making decisions to cope with difficulties that arise is the course of the project and in dealing with flaws in the project that are only revealed once in progress. It is unwise to presume that all the careful planning that took place before the project 'hit the field' will in fact have foreseen all eventualities, and the ability to put in place corrective measures where judged necessary must be one of the features of a successful project.

All too often a statistician's help is sought only when there is a set of data that has been collected and advice some method of analysis is required. In fact at the design stage of a project this type of advice is even more essential than at later stages. A design that does not permit a proper statistical analysis cannot be corrected after the data collection phase has been completed, and there are usually a variety of pitfalls that can render a study unusable, from failing to sample properly, failure to ask questions in the necessary form, failure to assess the feasible and the necessary size of the project, and so on. These types of shortcoming are often a little too technical for the non-statistical researcher to spot easily, but an experienced statistician should always be able to give guidance on them. Attention to and advice on design should also cover the data recording or collection forms – data coding sheets – and, if used, questionnaires. For those studies that require an interview with drug abusers, a statistician can be a great help in designing the detail of the questionnaire.

When all the planning that can be done is complete, then the burden of research falls upon the field workers and the field organisation. Good fieldwork is in the final analysis the cutting edge of data collection, be it collection from institutional records or through personal interviews. To this end it is highly desirable to use experienced fieldworkers, particularly if there are to be interviews with drug abusers. The need for drug users to feel comfortable with the interview, that the interviewer is not judging them morally, and that confidentiality of response will be respected is essential if good and reliable data are to be collected. People who work in the drugs arena are useful in this respect, and employing ex-drug abusers should also be considered; for obvious reasons people associated with the law in some way are not good interviewer material. Whether collecting from official sources, semi-official records or by interview, co-operation should be the touch-word for the fieldwork force.

Whether the fieldwork force is experienced or not, it will be necessary to train them in the particular methods, questions and procedures that will be used in the data collection phase of the study. It should not be presumed that the data collection procedures are obvious or self-explanatory. Ensuring that criteria, definitions and interpretations are uniformly applied if there are several field workers is a further aspect of the necessary training. Some degree of supervision and organisation will be required in co-ordinating the field workers. This serves both to control the quality of the data collected and to ensure that procedures are correctly followed, and also to control the physical flow of information – data sheets or questionnaires –

back to the study headquarters. The tendency for questionnaires and data sheets to be lost or spoiled after collection is surprisingly high. In the final analysis the validity of the data used in the study depends upon the quality of the fieldwork organisation and the fieldwork force.

No modern prevalence studies can be conducted without access to a computer and knowledge of how to operate software programs and how to design and construct data files. Data entry into a computer format can be a lengthy procedure and, if the study is a large one, the researchers should consider entering data as the study progresses. If the primary researchers do not have these computer skills then it is essential to have the co-operations of someone who does. A statistician can always advise on these matters as an addition to advice on statistical analysis. It is hoped that this manual will give guidance on statistical analysis for those researchers with some expertise in these matters and the text does indicate where the more complex aspects of analysis will require more expert input. For those with no experience, it will be essential to seek statistical advice when implementing these indirect estimation methods.

Finally, when the analysis is under way and writing up the findings is the focus, it is important this is not done in isolation. Statistical help is only one aspect of the benefits that come from team-work at this stage: the advice of others – be it informally obtained through contact with peers or obtained through more formally organise discussion groups, even internet-based discussion groups

Ethical considerations in different social settings

Care needs to be taken always in research that involves acquiring personal details and information from respondents in a survey or any other type of study. Personal interviews with respondents themselves are especially rich with pitfalls. The social customs and mores of each country must determine what considerations might be required to identify ethical as opposed to unethical modes of comportment, good manners and socially acceptable procedures. The following headings classify what might be primary concerns in many countries and in many social settings.

Social Censure

This type of research is dealing with socially stigmatised behaviour; and therefore must always be handled with great sensitivity. The respondent in any interview should be able to expect confidentiality and anonymity, both in respect of the answers he or she makes and in the way he or she is contacted. It is vital that the researcher should be in a position to assure respondents of the anonymity and confidentiality of their responses not only because it is more likely to produce valid data, but also because of the simple ethics of the way in which the researcher is likely to impinge on a respondent's life. Different groups of people will need different and specialised assurances in may cases - for example the issues relating to young people and to women are in some countries highly charged, requiring especial attention

Interview settings

The manner of contacting respondents in drug abuse inquiries is important: it should be discreet and as private as is possible. This consideration extends to the interview setting for questioning respondents: it is usually recommended that this be done in a private and isolated room or area. If this is not possible then the researcher must pay special attention to family issues, not only concerning the role of women in the family in general – which in many countries requires especial attention - but also the interactions between husbands, wives and children.

A second significant factor in interviewing is the role that can be played by social status in the country. The interviewers need to be aware of any social status differentials between themselves and the respondents and in particular this would apply when help is sought from key socially prominent individuals. There can be no general guidelines that can apply across the many different social setting in different countries. The best this manual can do is to draw attention to the issue to alert the researcher to develop approaches that are suitable.

Security considerations

There are intrinsic difficulties for the researcher dealing with illegal behaviour. Foremost is the type of relations the researcher has with the police and judicial systems in the country, and with any other government record offices. It is vital that the researcher should be in a position to physically ensure anonymity and confidentiality of all information offered by way of the research. Furthermore, it is part of this confidentiality that there is a separation of data and record keeping from the more accessible administrative documents and writings. If any form of identification of the respondent is retained, this should be restricted in its detail as far as is practical, and even this restricted identity information should be held in a secure setting. It is often the case that government records held in one data system become 'contaminated' by identification possibilities across to another, separate government information data system. It is important to control access to the gathered data and other project records - and especially to the responses of interviewees - during the entire project and also – importantly – after the project has finished. To ensure anonymity and confidentiality after the project is finished would usually require the destruction of all administrative information and personal identification information, retaining at most only the research data counts and completely anonymous (i.e., de-identified) records.

Technical Guidelines

In this section brief guidelines are offered for treading the sequence of steps in a research study that lead to actual data collection and subsequent publication. These span the study from its beginnings of targeting, planning, and choosing the appropriate method through to selecting the type and place of final publication of the results.

Target definitions

Terms of reference

There are a number of targets within any study that need to be clearly set out at the start of the project, if 'project drift' is to be controlled. Not least amongst these is the need for a clear statement of the terms of reference for the overall project, under which it will operate. From the financial point of view, most funding bodies will want to see that the project has delivered what is set out under its terms of reference. From the planning point of view, there is a need to curb the natural tendency for a project to expand its aims as more and more interesting questions are brought into focus: a project can easily become too ambitious is its aims in both financial and practical aspects. The tendency for the more difficult research aspects to be shelved is also a natural development, which needs to be controlled. A project that has several strands to it – either several strands of data collection or several methods of estimation being applied under the same exercise – needs to ensure that these are kept as distinct as is practical and that they progress in a unified framework.

Target and reference population

The target population is, in these prevalence estimation studies, the population of drug abusers, somehow defined – in other words the numerator in the prevalence rate. The reference population for a prevalence estimation is the baseline group of people – that is, the denominator in the prevalence rate - amongst whom the target population is to be found. Defining criteria with respect to individual demographics, geographical location and time period must be the same for both target and reference populations.

Clearly a geographical delimitation\of the populations is the most obvious requirement: the whole country, or a specified part of it if there are indeterminacies about borders or other reasons for defining only part of a country. In some instances it may be impractical to carry out a prevalence estimation exercise in some region of the country, either because of the difficulty of access, or limits on available information or even because the prevalence level of drug abuse is so low that it will be too difficult to measure. It may be that the prevalence estimate is wanted only for a small region, or even for a single city. Precise and clear definition is in all cases a priority requirement.

There will presumably be considerations about nationality and residency, too. For example, are visiting or resident foreign nationals to be included or excluded from the prevalence figure? In some countries this may make no tangible difference to the prevalence figure, in other places it might. When the studies are more locally based, say for example if the target and reference are just one city in the country, then residency issues may become very important. The choice of criteria should be determined by the principal research question under the terms of reference, and the methodological considerations determine that the criteria apply to target and reference alike.

In addition to geographical definition, the exact demographic range of the study will need to be identified: does the study deal with people between 15 and 65 years old, for example, or with the entire age range? Does it cover both males and females in the population? Employed and un-employed? Most often the choice with respect to these demographic and perhaps other social characteristics is determined by convenience and availability of the necessary data, as well as accessibility, and not just by policy preference alone.

Defining the time-span of the study is usually slightly more complex. This will be part of the definition of the target behaviours (see Key Measures, below), along with definition of what behaviours are the criteria for defining a 'drug abuser' for the target population. For example, a standard, recommended time period is part of the definition 'used opioids in the past 12 months'

Key behavioural measures

To identify the target population definitions are also required for the drug subset(s) of interest. Section 1 gave the drug categories that are used in the GAP operational classification of drug abuse. It is of course recommended that several – or all the relevant - prevalence rates, defined by several of these categories, are estimated within the same one study. This is not always possible, particularly when some data sources relate to only, say, opioid treatment, and the likelihood of finding users of other drugs is limited. It may be that it is injecting of drugs – any drugs, not just heroin - that is the key definition, or it may be that definitions are required for the concepts of 'drug dependence' or 'problematic drug use'. Note in this regard that it is usually very difficult to operationalise the ICD or DSM criteria fort these concepts in the field, or through a questionnaire administered by a field worker. Close – but always precise – definitions are often the best that can be achieved.

Section 1 lists the definitions for period of prevalence that might be used. Preferably for compatibility with the ARQ a 1-year period prevalence is the basic requirement, but information on use over other periods can also be obtained at the same time. The location in time of the period is usually the 12 months prior to interview, if interviews are being conducted. When using data from pre-existing records the researcher may have no choice than to use the past calendar year. Where multiple data sources are being cross-linked it is by far the easiest option to use the same calendar period in all data sources, if at all possible. Not doing so requires special and sometimes difficult weighting adjustments to be made in the cross-linking (Simeone et al., 1997; Fischer, Hickman et al., 1999].

Where it is part of the data collection to identify a 'drug abuser' or a 'current addict' or some such similar concept, the key criterion in the definition is always the frequency of use within a relevant period. A person is usually said to have an active career if he or she is using drugs at or

above a stipulated threshold level, in terms of frequency – even if the threshold level is defined as 'any use at all'; amounts of drug used are always too difficult to gauge practically when in the field. As defined in many studies (e.g. Simeone et al., 1997) here, a drug use career begins the first time that an individual uses drugs at or above some operational threshold.

Although the drug user may drop below this criterion threshold frequency and later rise above it again, effectively dropping out and back into drug using, this kind of detail is usually disregarded ,or definitions are chosen in relation to the time period of the study that preclude this possibility (for example, 'any use in past 12 months' in a 12-month study precludes such difficulties). Moreover, drug users experience various kinds of events during the course of their career that are relevant to these estimation techniques, such as arrests, admissions to drug treatment, and stays at homeless shelters, which are considered usually as instantaneous, even though some types of events may engender states that have duration - in jail, in drug treatment, and in a shelter. It is not likely, for example, that someone will be arrested while he is in the hospital, but these types of confounding issues are usually disregarded in analyses. If the study can use this type of information profitably then expert statistical advice should be sought on how to do so (see Simeone et al., 1997).

Data source definitions

The objectives and definitions to achieve then are of course structured and limited by what is available in existing data sources. It has already been mentioned that it is important that the time-span of the study is reflected in each data source, when more than one is being used. It is equally important that all other definitions correspond across all data sources. Where this might be impossible, then tests or calculations should be carried out to see how far the discrepancies between the available data definitions will disrupt the overall analysis of the data. This may be a complicated matter, or it may be quite easy. A simple example would be the definition of a heroin injector: police arrest sources of data may use different criteria to determine this status than would a heroin abuse treatment centre; and if data is collected on HIV tests amongst injectors, it may be that the information relates to injection of any drug, not just heroin. A very common example of definition ambiguity relates to 'attending treatment', with different sources perhaps referring to different types of treatment and, in a local study, to different treatment units such as those in or out of the local area. Instances of these failures to match are discussed in several of the case studies.

Data possibilities and methods

Pre-assessment of the situation

It is vital in the planning stages that effort is put into assessing the existing state of knowledge about drug use. Whether this knowledge turns out to be trustworthy or of dubious validity is not as important as whether it can raise issues and questions that will benefit the thinking in designing a study. Consideration should be given not only to official statistics and publications when making this assessment, but the planners should also consider the use of key informants to help in this process.

At the very least a quick and easy assessment of all available data sources should be made and a general overview of the potential for primary data gathering should be made. There is an important role to be played here through inspecting of similar studies, in the target country or outside it, and of any types of study that have been conducted in similar social circumstances

Audit of national or local routine data sources

It is helpful when beginning a research study to have a very clear idea of what data and information sources are routinely available and which of these can be accessed for extracting information relevant to the study - relevant in any aspect whatsoever. These data sources provide not only figures for epidemiological use, but also they provide lists of institutions, locations and individuals that may be used for sampling purposes in the research. Proper

'sampling frames' are hard to come by in any circumstances, but more so than ever in the drugs research field, and these sources should be checked before any project design is finalised.

Since the availability of data sources like these vary from country to country and within countries sometimes from social setting to social setting, no general statements on what is available for use and what to use can be made. The following check-list may be of use to researchers in helping the start of an auditing process of sources that may facilitate contact with drug users.

CHECK LIST

Possible data sources:

- registers/records from treatment centres
- doctors and medical professionals generally
- general hospitals
- psychiatric hospitals
- specialist drug user services
- HIV or other health registers
- specialist addict registers
- deaths registers
- drug-related deaths registers
- police and judicial records

The EMCDDA guidelines to prevalence estimation (see Additional Sources, Section 1) should also be consulted on this point.

Characteristics of data sources

Whatever sources of useful information are identified, there are certain key aspects that should be checked about them. These relate to content, to structure, and to accessibility. Again, the following checklist gives some general aspects that should be considered, but particular studies will of course have particular demands upon the data they will use. This list is therefore a broad guide of some of the issues about the data source and its contents that may be relevant to a research study.

CHECK LIST

Relevant features of data and information bases

- who owns and maintains the data base?
- who owns the data itself ?
- what type of drug(s) does the information describe?
- does it hold drug injecting information?
- what information on drug use is held?
- are the data based on persons or events/contacts?
- does it distinguish first-time-contact from repeated contact?
- are individuals identifiable across events/contacts within the data?
- are individuals identifiable in terms of other data sources?
- what is the geographical coverage?
- what is the time-span covered?
- in what physical format are the kept?
- are there potential obstacles to accessing the data?
- are there possibilities for future improvement in the data?
- are there networking possibilities arising from is use?

Again, the EMCDDA manual on guidelines to prevalence estimation (see Additional Sources, Section 1) should also be consulted.

Ad hoc data sources and potential for primary data collection

When a study sets out to collect data of its own for any – that is, from sources that are not routinely collected - of the procedures described in this manual, there may be a potential for this ad hoc source to be developed into a permanent source for future monitoring. For example in the Pakistan study (Case Study 4) a register of specialist drug treatment agencies was constructed as part of establishing a benchmark for the treatment multiplier estimation procedure that was used. It was a peripheral target in the study that this register could be kept updated and provide a resource in future for use either by further drug prevalence studies or for any other purpose. Whilst not an official data source, effort was put into leaving this register in hands that would be able to maintain it into the future.

Issues in primary data collection methods

Role of piloting for data recording procedures and questionnaires

The process of data collection requires considerable organisation, co-operation and planning. As part of this, there is a need usually to develop physical (hard-copy paper) forms for data collection, be they forms for transcribing official records into a more convenient format for the study or be they questionnaires for recording interview information. Of course in some cases, with routine data sources, it may be that the data are already held in electronic format that requires only a little computer processing to be suitable for study use. This is not often the case, but when it is, full advantage should be taken of this bonus and any transcribing to paper forms avoided. Usually though, when transcription or collection forms are required, there are general points about these forms that need to be borne in mind when designing them. The first is that almost certainly they will be used as input to a computer data system of some type, so the relevant computer expertise should be called upon to help devise or at least approve them. This will apply to both how the data are recorded and how it will be coded, as well as how the

individuals and forms will be identified or numbered in the computer coding scheme. A second over-arching design consideration relates to who it is that will fill in these forms - is it the researcher themselves, a trained interviewer, or is the form to be completed by the person interviewed?

When designing a questionnaire to be filled in from an interview there are established principles that aid good data collection. Whilst the contents of questions asked are properly the domain of the expert in the research field, the technical points in questionnaire design need to be addressed with professional help or the help of a good reference book [Oppenheim, "Questionnaire Design", for example]. A statistician can be a help in determining both the more general aspects of specific, prevalence-related questions and the variety of design details that should be given attention. Although for prevalence estimation the questions are likely to be more easily determined because they relate only to specific behavioural measures and not attitudes, which are more difficult to measure, nonetheless some experience is required in questionnaire design. A list of some of the detailed points that should be considered is given in the following checklist. An assessment also should be made of the time it is likely to take to ask and record the information in the interview setting - no respondent has unlimited time nor endless patience to help with voluntary research.

CHECK LIST

Basic questionnaire design considerations

- ensuring each question can be answered properly and sensibly by everyone
- making sure there is always room to answer 'do not know the answer'
- allowing for people who refuse to answer some – or indeed all – of the questions
- 'routing' the person filling in the questionnaire through sections that can be 'skipped' in particular cases
- ensuring that no 'blank-coded responses' occur – some mark or other should always be recorded for each question
- determining what categories of answer are of interest, and that they are exhaustive and mutually exclusive categories
- determining when multiple-response questions can be used
- deciding which questions can best be left be 'open-ended'
- ensuring contingent questions (if 'yes' then what ...?)...are properly structured and routed
- ensuring that the relevant time-frame or time window is specified
- ensuring that, where relevant, the appropriate geographical location specified

With so many design issues in both the form or questionnaire itself and in procedures for interviewing or recording data, it is essential that proper 'pilot' studies be conducted to test out how things might actually turn out when 'in the field'. Although usually small in scale, these studies are invaluable in trapping and revealing errors and obscurities and difficulties in the questioning, recording and procedural steps that have not been foreseen. It is of course important that piloting takes place in conditions that are as close to actual full study conditions as possible in order to make the best use of it as a check.

Non-standard selection of respondents when interviewing

In many drug studies of prevalence where drug abusers are counted or interviewed as part of the data gathering in the estimation procedures, the inaccessibility of drug abusers forces the researchers into sampling them through non-standard means. By 'non-standard' here is meant

methods that, because they do not have access to a sampling frame of individuals that is required by classic probability sampling, use more inventive procedures. Foremost amongst these is site sampling, described immediately below, but a second frequently used procedure is nomination sampling. In essence this method uses a standardly drawn small random sample of respondents, obtained by some means, and expands it by asking respondents for information about acquaintances and friends. 'Snowball' sampling and 'chain referral' sampling are two names under which a variety of procedures based on this principle are know. The reader is referred to the publications by Intraval (Bieleman and de Bie, 1992) and the Wirral studies (Howard Parker et al., 1988) for detailed descriptions of procedures and safeguards in these methods.

Binomial sampling (Site sampling) issues

In designing a data collection strategy, it is often necessary to address a central problem discussed above, namely that people involved in drug use tend to be difficult to locate. Many approaches interview admitted drug users in places where they are most likely to be found, including booking/arrest facilities, public and private drug treatment programs, and homeless shelters. This sort of sampling, that is, sampling that takes place physically at a site, rather than using a pre-constructed sampling frame or list of drug users, is called 'site' or 'binomial' sampling (Bielemann and de Bie, 1992; Goodman 1961). One of the issues that needs to be assessed in site sampling is whether the frequency with which someone attends the site will distort the sample away from a random representation of the target population - since obviously the more frequently a particular type of person attends, the more likely they will be sampled in relation to other types of attenders. And this consideration must be taken in relation to whether it is people whom we want to sample randomly or whether it is events (attendances). For example, in site sampling, if we want to project the results for this sample to the entire population of hardcore drug users, then we must weight the sample event rates by the inverse of the probability that the user was sampled (see e.g. Simeone et al., 1997). Failure to do so means that the sample is representative of attendances at the site, not attenders.

Sampling of sites: selection procedures

When carrying out procedures of any sort that require sampling or enumeration or record collecting at various sites, and when it is impossible to include all relevant sites in the data collection - usually the available resources determine that some sample of them is all that can be included - then as far as possible a formal sampling procedure should be drawn up to select the sites themselves. I it is possible to generate a sampling frame for the relevant sites that has complete, or at least very good, coverage of the population, then stratification and clustering procedures can be used to help with obtaining a feasible, representative selection. Researchers not familiar with probability sampling procedures should consult an expert.

Interviewing and report veracity

A further problem arises in interviewing situations where drug abusers need to be distinguished from non-users, in relation to veracity of answers about (self-reported) drug use. It would be reasonable to assume that people who are willing to admit that they are drug users will be candid in answering other questions about their use. But not all drug users who make contact with these institutions will be forthcoming about their behaviour. If it is felt necessary, some biological tests might be possible to identify recent drug use, or even tests on a random sample of people to estimate the proportion of drug users who are concealing heir use. In practice though, this is not usually a feasible route to follow since people are reluctant to submit to bio-testing - unless perhaps there is a financial inducement - events of each kind that are attributable to members of the target population; in general researchers hope for the best in relation to veracity.

Cohort mortality studies and the deaths multiplier

Specifically in multiplier studies it was mentioned that primary data collection might be needed to establish the value of the multiplier to use in conjunction with its benchmark. In the case of the mortality multiplier, it is unlikely that a study-specific exercise can be launched to establish the death rate amongst the drug abusers being studied. Mortality rates are usually determined by cohort studies, which are both difficult studies and long-term studies. The interested reader is

referred to specific texts on these types of cohort study (see e.g. Frischer et al., 1998) and to the EMCDDA publication on indirect estimation (see Additional Resources, Section 1).

Publication and reviewing by peers

Whether the proposed study is an internal report for a funding organisation or for a government agency, or whether it is intended to provide a publication in the academic literature, it is vital that the report is submitted for reviewing by peers and other experts. No researcher can plan in advance for all contingencies nor see all potential pitfalls and difficulties of their research. Reports and papers benefit enormously from being reviewed by other researchers and experts - this practice of critical reviewing is universal amongst serious and well-respected researchers, academic journals, and funding bodies.

It is of paramount importance not only as a help to the researcher but also in maintaining accepted reporting standards within the international research community. As such, it is the surest way for the research project and its publications to achieve acceptance as a valid study in the professional community and to be recognised beyond the bounds of the researchers' immediate circle of colleagues and financial providers.

When the primary aim is for the research to appear as a publication as governmental report, it is often possible to take advantage of this as an opportunity to follow up with a publication in the academic research community journals as peer-reviewed research. One difficulty with research findings is that they are often buried in lengthy governmental reports and become difficult to access (i) possibly because of required permissions, and (ii) possibly because of the length of the reporting document. A research publication in the public forum however, is often far more succinct and is usually directed at particular points of major interest in the project's findings, which enables these findings to have a wider circulation than they otherwise might. Permissions from the body who commissioned the report will need to be obtained at the point of publishing any follow-up paper, and ensures that there is no further administrative difficulty in accessing the findings.

A final report should have written up in detail the methods used in both the data-gathering and the analysis phases of the study. Moreover, it is important that weaknesses and difficulties in the research, foreseen and unforeseen, should be reported if credibility is to be maintained. In general terms the four aspects of a research study:

- planning the research,
- collecting the relevant data,
- data analysis, and
- writing up the research study.

All carry equal weight and can be each as equally labour-intensive as the others. The writing up of both the data collection procedure and the results of the analyses are equally important. The easiest way to ensure accuracy in reporting is to write up the research *as it proceeds.* Waiting until the end of the exercise before beginning the reporting is likely to lose important detail and important qualifications and trouble-spots in the research. And having written up the research procedures as the research proceeds not only gives a more comprehensive report but in addition lightens the work load in the final stages of the research project.

Further reading and resources

General guidelines on prevalence estimation

The GAP Epidemiology Website - purpose and use

The overall GAP has a website that the reader should visit. The development of this site as a resource for helpful reports, publications and other web-sites is one of the priority areas of the project. See *www.undcp.org,* and look for GAP. See also the Resource Centre at GAP.Gide.net .

Ueful Websites Index

http://www.who.int/substance_abuse/PDFfiles/EPI_GUIDE_A.pdf

http://www.emcdda.org/situation/themes/problem_drug_use.shtml

Estimating Prevalence of Problem Drug Use in Europe (EMCDDA Scientific Monograph Series Number 1. (1997) ISBN 92-9168-006-0 Luxembourg: Office for Official Publications of the European Union).

Guide to Drug Abuse Epidemiology (2000) (WHO/MSD/MSB 00.3), prepared by the Mental Health and Substance Dependence Department, WHO, Geneva.

General documentation index

Reference cited in the text

Archibald C.P. et al. (2001) Estimating the size of hard-to-reach populations: a novel method using HIV testing data compared to other methods, AIDS, 15 (suppl3): S41-S48

Bieleman B and de Bie E (1992). Between The Lines - A Study Of The Nature And Extent Of Cocaine Use In Rotterdam. *Intraval Foundation, 1992.*

Bishop YMM, Fienberg SE, Holland PW. (1975) Chapter 6. Estimating the size of a closed population pp 229-56 In: Discrete multivariate analysis: theory and practice. MIT Press: Cambridge, Massachusetts, US, 1975

European Monitoring Centre for Drugs and Drug Addiction (EMCDDA).

Methodological guidelines to estimate the prevalence of problem drug use at local level. CT.97.EP.05. Lisbon, EMCDDA, 2000. Available at: http://www.emcdda.org/situation/themes/problem_drug_use.shtml

European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). Scientific Monograph Series No. 1. Estimating the prevalence of problem drug use in Europe. Lisbon: EMCDDA, 1997

Fitch., C., Stimson, G.V., Jones, S., and Hickman, M. (2001). Responding to drug and alcohol use in Jersey: key findings report. Drug Strategy Unit, States of Jersey. Frank, B, Scmeidler, J., Johnson, B and Lipton, S. Seeking truth in heroin indicators: the case of New York City. *Drug and alcohol dependence* 1978; 3: 345-358.

Frischer, M. (1998) Estimating the prevalence of drug abuse using the mortality multiplier method: an overview. in EMCDDA (1997) *Estimating the prevalence of problem drug use in Europe*. Scientific Monograph No. 1.

Martin Frischer, Mathew Hickman, Ludwig Kraus, Fabio Mariani & Lucas Wiessing. A comparison of different methods for estimating the prevalence of problematic drug abuse in Great Britain

Frischer M, Bloor M, Goldberg D, Clark J, Green S, McKeganey N. Mortality among injecting drug users: a critical reappraisal. J Epidemiol Community Health. 1993; 47(1): 59-63

Frischer M, Leyland A, Cormack R, Goldberg DJ, Bloor M, Green ST, Taylor A, Covell R, McKeganey N, Platt S. (1993) Estimating the population prevalence of injection drug use and infection with human immunodeficiency virus among injection drug users in Glasgow, Scotland. Am J Epidemiol. 1993 Aug 1; 138(3): 170-81

Goodman L. Snowball Sampling. (1961) *The Annals of Mathematical Statistics, 1961, 32, 148-170.*

Hall, W., Ross, J., Lynskey, M., Law, M., & Degenhardt, L. (2000). How many opioid users are there in Australia? *Medical Journal of Australia,* **173**, 528-531.

Hay G, Smit F. Estimating the number of drug injectors from needle exchange data. Addiction Research and Theory. (in press)

Hickman M, Taylor C, Chatterjee A, Degenhardt L, Frischer M, Hay G, Tilling K. Estimating drug prevalence: Review of methods with special reference to developing countries. UN Bulletin on Narcotics, 2002

Hook EB, Regal RR. (1995) Capture recapture methods in epidemiology: methods and limitations. Epidemiologic Reviews 1995; 17: 243-64

Hser Y, Anglin, MD, Wickens TD, Brecht ML, Homer J. Techniques for the Estimation of Illicit Drug User Prevalence: An Overview of Relevant Issues. Washington: National Institute of Justice, 1992

International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple record systems estimation I: history and theoretical development. Am J Epidemiol. 1995; 142: 1047-57

Lynskey, M. and Hall, W. (1998) Jurisdictional trends in opioid deaths 1988-1995. *Australian an New Zealand Journal of Public Health,* **23**,519-524

Mariani, F., Guaiana, R. & Di Fiands, T. An epidemiological overview of the situation of illicit drug abuse in Italy. *The Journal of Drug Issues 1994;* 24, 579-595.

Mastro T.D., Kitayaporn D, Weniger BG, Vanichseni S, Laosunthron V, Thongchai U, Uneklabh V, Vhoopanya K, Kimpakarnjanarat K. Estimating the number of HIV-infected injection drug users in Bangkok: a capture-recapture method. Am J Public Health 1994; 84: 1094-9

McKetin R., Darke S., Hayes A.. and Rumbold G. (1999) Drug trends 1998. A comparison of drug use and trends in three Australian states. NDARC Monograph No. 41. Sydney, UNSW.

Rehmann, Griffiths and Taylor (2002) in 'UNDCP Rapid Assessment Report on the Pakistan National Assessment Exercise', U.N.

Reuter, P. (1993) Prevalence estimation and policy formulation. *Journal of Drug Issues,* **23,** 167-184.

SHAKTI Project, *CARE Bangladesh, Dhaka, Bangladesh*

Simeone R, Rhodes W, Hunt D, Truitt L. A plan for estimating the number of "hardcore" drug users in the United States. Washington: Drug Policy Research Group, Office of National Drug Control Policy, 1997

Tilling K, JAC Sterne. Capture recapture models including covariate effects. Am J Epidemiol 1999; 149 (4): 392-400.

Wickens TD. Quantitative methods for estimating the size of the drug using population. *Journal of Drug Issues* 1993; 23: 185-216.

Woodward, J.A., Retka, R. and Ng, L. Construct validity of heroin abuse estimators. *The International Journal of the Addictions 1984; 19* (1), 93-117.

Capture-recapture references

Hook E B and Regal R R (1992) Effect of Variation in Probability of Ascertainment by Sources upon "Capture-Recapture" Estimates of Prevalence. *American Journal of Epidemiology, Vol 137, No 10.*

Doscher M-L, Woodward J L (1983) Estimating the Size of Subpopulations of Heroin Users: Applications of Log-Linear Models to Capture-Recapture Sampling. *The International Journal of the Addictions, 18(2), 167-182, 1983.*

Wittes J T. (1974) Capture-Recapture Methods For Assessing The Completeness Of Case Ascertainment When Using Multiple Information Sources. *J Chron Dis 1974, Vol 27, pp 25-36, Pergamon Press.*

Brownie C (1987). Reader Reaction - Recent Models for Mark-Recapture and Mark-Resighting Data. *Biometrics 43, 1017-1022, December 1987.*

Seber G A F. (1986) A Review of Estimating Animal Abundance. *Biometrics 42, 267-292, June 1986.*

Wolter K M (1990). Capture-Recapture Estimation in the Presence of a Known Sex Ratio. *Biometrics 46, 157-162, March 1990.*

Cormack R M (1989). Log-linear Models for Capture-Recapture. *Biometrics 45, 395-413, June 1989.*

Bonett D G, Woodward J A and Bentler P M (1988). A linear model for estimating the size of a closed population , *British Journal of Mathematical and Statistical Psychology (1986), 39, 28-40.*

Cowan C D and Malec D (1986) Capture-Recapture Models When Both Sources Have Clustered Observations. *Journal of the American Statistical Association. June 1986, Vol 81, No 394, Survey Research Methods.*

Chao A (1987). Estimating the Population Size for Capture-Recapture Data with Unequal Catchability. *Biometrics 43, 783-791, December 1987.*

Haber M (1986). Testing for Pairwise Independence. *Biometrics 42, 429-435. June 1986.*

Community and Snowball Sampling references

Kaplan C D, Korf D, Sterk C. (19..) Temporal and Social Contexts of Heroin-Using Populations. An Illustration of the Snowball Sampling Technique. *The Journal Of Nervous And Mental Disease. Vol 175, No 9.*

Biernacki P. (1981) Snowball Sampling. Problems and Techniques of Chain Referral Sampling. *Sociological Methods & Research. Vol 10 No 2, November 1981, 141-163.*

van Meter K M . Methodological and Design Issues: Techniques for Assessing the Representatives of Snowball Samples. *Source not stated*.

Kaplan C D, Korf D, and Sterk C . Temporal and Social Contexts of Heroin-Using Populations. An Illustration of the Snowball Sampling Technique. *The Journal Of Nervous And Mental Disease. Vol 175, No 9.*

Bieleman B and de Bie E (1992). Between The Lines - A Study Of The Nature And Extent Of Cocaine Use In Rotterdam. *Intraval Foundation, 1992.*

Goodman L . Snowball Sampling.(1961) *The Annals of Mathematical Statistics, 1961, 32, 148-170.*

Griffiths,P., Gossop,M., Powis,B. et al (1993) Reaching Populations

of drug users by the use of privileged access interviewers: methodological and practical issues. *Addiction*, 88, 617-1626.

Mortality and Deaths Multiplier References

S, J. and Hall,W. (1996b) Overdose among heroin users in Sydney, Australia: II. Responses to overdose. *Addiction,* 91:3, 413-417.

Davoli,M., Perucci,C., Forastiere,F., et al (1993) Risk factors for overdose mortality: a case-control study within a cohort of intravenous drug users. *International Journal of Epidemiology,* 22:2, 273-277.

Farrell, M., Neeleman,J., Griffiths,P., et al (1996) Suicide and overdose among opiate addicts. Editorial. *Addiction,* 91:3, 321-323.

Frisher,M., Bloor,M., Goldberg,D., et al (1993) Mortality among injecting drug users: a critical reappraisal. *Journal of Epidemiology and Community Health*, 4, 59-63.

Gossop,M., Griffiths,P., Powis,B., et al (1996) Frequency of non-fatal overdose. *British Medical Journal,* 313, 402.

Hammersley,R., Cassidy,M. AND Oliver,J. (1995). Drugsassociated with drug-related deaths in Edinburgh and Glasgow, November 1990 to October 1992. *Addiction,* 90, 959-965.

Home Office (1995) Statistics of drug addicts notified to the Home Office, United Kingdom, 1994. *Home Office Statistical Bulletin 17/95*.

Joe,G,W., Lehman,W. and Simpson,D. (1982) Addict death rates during a four-year post-treatment follow-up. *American Journal of Public Health,* 2:7, 703-709.

Ruttenber,A. and Luke,J. (1984) Heroin-related deaths: new epidemiological insights. *Science*, 226, 14-20.

Walsh,R. (1991) Opioid drug accidental deaths in the Newcastle area of New South Wales, 1970-1987. *Drug and Alcohol Review,* 10, 79-83.

Zador,D., Sunjic,S. and Darke,S. (1996) Heroin-related deaths in New South Wales, 1992: toxicological findings and circumstances. *Medical Journal of Australia,* 164, 204-207.

General References

"European Monitoring Centre for Drugs and Drug Addiction (EMCDDA).

Methodological guidelines to estimate the prevalence of problem drug use at local level. CT.97.EP.05. Lisbon, EMCDDA, 2000. Available at: http://www.emcdda.org/situation/themes/problem_drug_use.shtml

"European Monitoring Centre for Drugs and Drug Addiction (EMCDDA).

EMCDDA Recommended Draft Technical Tools and Guidelines - Key Epidemiological Indicator: Prevalence of problem drug use. Lisbon, EMCDDA, 2000. Available at: http://www.emcdda.org/situation/themes/problem_drug_use.shtml

Hickman M, Taylor C, Chatterjee A, Degenhardt L, Frischer M, Hay G, Tilling K. Wiessing L., Griffiths, P. and McKetin, R. (in press) Estimating drug prevalence: Review of methods with special reference to developing countries. UN Bulletin on Narcotics Vol. No. LIV

Hook EB, Regal RR. Capture recapture methods in epidemiology: methods and limitations. Epidemiologic Reviews 1995; 17: 243-64

International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple record systems estimation I: history and theoretical development. Am J Epidemiol. 1995; 142: 1047-57

Bishop YMM, Fienberg SE, Holland PW. Chapter 6. Estimating the size of a closed population pp 229-56 In: Discrete multivariate analysis: theory and practice. MIT Press: Cambridge, Massachusetts, US, 1975

Mastro T.D., Kitayaporn D, Weniger BG, Vanichseni S, Laosunthron V, Thongchai U, Uneklabh V, Vhoopanya K, Kimpakarnjanarat K. Estimating the number of HIV-infected injection drug users in Bangkok: a capture-recapture method. Am J Public Health 1994; 84: 1094-9Mastro et al.

Fitch, C., Stimson, G.V., Jones, S., and Hickman, M. (2001). Responding to drug and alcohol use in Jersey: key findings report. Drug Strategy Unit, States of Jersey.

SHAKTI Project, CARE Bangladesh, Dhaka, Bangladesh

McKeganey N, Barnard M, Leyland A, Coote I, Follet E. Female streetworking prostitution and HIV infection in Glasgow. BMJ. 1992 Oct 3; 305: 801-4

Ghodse, H.(1977) Casualty departments and the monitoring of drug

dependence. *British Medical Journal*, 1, 1381-1382.

Frank O (1979) Estimation of population totals. *IN: Holland P W and Leinhardt S L eds. "Perspectives On Social Netweork Research New York: Academic (1979).*

Picklands III J. and Raghavachari M . (1987) Exact and asymptotic inference for the size of a population. *Biometrika (1987), 74, 2, pp 355-63.*

Further References

Hickman M, Sutcliffe A, Sondhi A, Stimson G. Surveillance of problem drug use in the UK: a review of a regional drug abuse database *Journal of Public Health* 1999; 21: 271-77

Crabbe T, Donmall MC, Millar T. Validation of the university of Manchester drug abuse database. *Journal of Epidemiology and Community Health* 1999; 154-65.

Unlinked Anonymous Surveys Steering Group (Chair: Metters J) Data to the end of 1998. London: Department of *Prevalence of HIV in the UK: Report of the unlinked anonymous prevalence monitoring programme in the UK.* Health, 1999.

Frischer M, Heatlie H, Millson D, Lowdell J, Hickman M, Chapman S, Bashford J, Norwood J. A comparison of trends in problematic drug abuse from two reporting systems. *Journal of Public Health Medicine* 2000; 22 (3): 362-367.

Bloor M, Wood F, Palmer S. *Estimating the prevalence of injecting drug use and serious drug use in Wales.* Cardiff: Social Research Unit, 1998.

Ditton J, Frischer M. Computerised projection of future heroin epidemics: a necessity for the 21st century? *Journal of Substance Use and Abuse (in press).*