UNITED NATIONS
Office on Drugs and Crime

Estimating Prevalence:

Indirect Methods for

Estimating the Size of

the Drug Problem

Estimating Prevalence: Indirect Methods for Estimating the Size of the Drug Problem

2

UNITED NATIONS OFFICE ON DRUGS AND CRIME
Vienna

Estimating Prevalence: Indirect Methods for Estimating the Size of the Drug Problem

Global Assessment Programme
on Drug Abuse

Toolkit Module 2

UNITED NATIONS
New York, 2003

Preface

GAP Toolkit Module 2: Estimating Prevalence: Indirect Methods for Estimating the Size of the Drug Problem, was prepared with the support of the United Nations Office on Drugs and Crime as part of the activities of the Global Assessment Programme on Drug Abuse (GAP). The main objective of GAP is to assist countries in collecting reliable and internationally comparable drug abuse data, in building capacity at the local level to collect data that can guide demand reduction activities, and in improving cross-national, regional and global reporting on drug trends.

The GAP epidemiological toolkit has been produced to assist States Members of the United Nations in developing culturally appropriate systems, relevant to the countries concerned, for collecting drug information, to support existing drug information systems by promoting their conformity to internationally recognized standards of good practice, and to 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 a statistical software package for the social sciences (SPSS).

Other GAP activities include the provision of technical and financial assistance in the establishment of drug information systems and support for and coordination of global data collection activities. For further information on GAP toolkit modules, contact gap@unodc.org or visit the GAP web site at www.unodc.org.

The purpose of the toolkit is to provide a practical and accessible guide to implementing data collection in core areas. The toolkit modules are designed to provide 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 panel of experts and endorsed by States Members of the United Nations. Although the models presented are based on existing working models that have been found effective, a key principle is that approaches have to be adapted to meet local needs and conditions. Module 2 of the toolkit therefore provides specific examples to guide the reader through the process of adapting general principles and models to specific contexts, and is not intended to reflect the complete range or diversity of current drug information systems or data collection methods.

Acknowledgements

Contents

Tables

Case studies

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 through better communications. The globalization of drug abuse means that demand reduction policies also have to be global, as must the information system on which they rely.

In 1998, the General Assembly, at its twentieth special session, devoted to countering the world drug problem together, adopted a Political Declaration (resolution S-20/2, annex) calling for the elimination or significant reduction of the supply of and demand for illicit drugs by the year 2008. That was the first time that the international community had agreed on such specific drug control objectives. However, the systematic data needed to monitor and evaluate progress towards those goals are not yet available. For that reason, the General Assembly requested the United Nations Office on Drugs and Crime (UNODC) to provide Member States with the assistance necessary to compile comparable data. UNODC was asked to collect and analyse the data and report thereon to the Commission on Narcotic Drugs. In response to those requests, UNODC launched the Global Assessment Programme on Drug Abuse (GAP). GAP has been designed to:

(a) Support Member States in building the systems needed to collect reliable data to inform policy and action;

(b) Encourage regional partnerships to share experiences and technical developments;

(c) Facilitate a better understanding of global patterns and trends in drug abuse by encouraging the adoption of sound methods to collect comparable data.

Those aims reflect the challenge posed in the Declaration on the Guiding Principles of Drug Demand Reduction adopted by the General Assembly at its twentieth special session (resolution S-20/3, annex), which states the following:

> "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 ... 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."

The main objective of GAP is to assist Member States in building the capacity to collect internationally comparable drug abuse data and to assess the magnitude and patterns of drug abuse at country, regional and global levels. The development of national and regional information systems should not only contribute to the building of capacity at the local level to collect data that can guide demand reduction activities, but also improve cross-national, regional and global reporting on drug trends. To support that process, *GAP Toolkit Module 2: Estimating Prevalence: Indirect Methods for Estimating the Size of the Drug Problem* has been produced to assist Member States in developing culturally appropriate systems, relevant to each country, for collection of drug information, to support existing drug information systems by promoting their conformity with internationally recognized standards of good practice, and to 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 developed and developing countries. For more information about GAP, visit the GAP web site at www.unodc.org, e-mail gap@unodc.org, or contact the Demand Reduction Section, UNODC, P.O. Box 500, A-1400 Vienna, Austria.

Role of prevalence estimation

The two key policy questions asked of those collecting information on drug abuse are as follows: "How many members of the population of a country are using drugs?"; and "Is that number changing?" Understanding the number of drug abusers is helpful in assessing the likely impact of drug abuse on society and what level of response may be required. For example, knowing how many injecting drug abusers there are allows for the calculation of the level of provision of services required to reduce behaviours associated with infection by the human immunodeficiency virus (HIV) and for gauging whether sufficient drug treatment places are available.

Understanding something about the dynamics of the drug problem makes it possible not only to assess the likely impact of the problem, but also to alert policy makers to a worsening situation, or alternatively to provide evidence that prevention and other initiatives may be working. In many countries, and especially where the drug problem is a relatively new phenomenon, having an estimate of the dimensions of the problem is a powerful tool in focusing the minds of both policy makers and the general public on the need for action and the investment of resources.

In epidemiological terms, the two above-mentioned questions relate to the estimation of prevalence and incidence. 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 those 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, the data are, in practice, extremely hard to generate. Many countries 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 the prevalence of drug use 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, those criteria are often difficult to operationalize in social survey work. Furthermore, many of those who consume drugs do not fall into the diagnostic categories, but are still of interest to policy makers. The range of different substances abused, the different routes of administration used and 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 stigmatized pose particular challenges for the drug researcher that are not found in most other areas of epidemiology.

Given the range of drug-taking behaviours, 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 the 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 of drug abuse, such as heroin or cocaine addiction or drug injection, are often not measured well by household surveys, and underreporting can be a problem. Household surveys are also technically complex and resource-intensive undertakings that are simply not practical in many developing countries. An alternative approach is to use indirect estimation techniques to estimate the number of drug abusers in the different categories. Those approaches are the main

subject of the present module of the GAP toolkit, which was prepared with the needs of developing and transitional countries in mind. However, it should be noted that the methods covered 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.

Global Assessment Programme on Drug Abuse project

Aims of the Global Assessment Programme on Drug Abuse toolkit concept

The purpose of the present toolkit is to provide a practical starting point for those wishing to use indirect methods to produce an estimate of drug prevalence. Those methods are most commonly used and are most appropriate for estimating the numbers of chronic or problematic drug abusers, such as street heroin addicts or drug injectors. That is because the methods described here often rely on the drug user appearing, or having the chance to appear, on a register of some kind, such as a police list, a treatment report or even a death report. The toolkit is not intended as a full technical manual; rather, it is intended to raise the major methodological and practical issues that need to be addressed in order to conduct a successful estimation study. Pointers are given to more in-depth technical material on each specific topic. Some of the techniques will require the input of a trained statistician, in which case the reader is alerted to the technical demands of the chosen method. Part of the rationale of the toolkit is to allow the technical resources required for an estimation study to be fully understood, and thus to facilitate the planning and implementation process.

Additional materials

The present toolkit is not the only resource of its kind, nor is it intended to be used as the only technical reference for planning an estimation study. In its preparation, a deliberate effort was made to link and complement some of the other material available in this field. References to technical papers can be found throughout the text. For general reference material, the following resources are recommended:

(a) A comprehensive overview of issues in drug epidemiology is contained in the World Health Organization (WHO) Guide to Drug Abuse Epidemiology (2000) (WHO/MSD/MSB 00.3). The Guide can be downloaded from www.who.int/ substance_abuse/PDFfiles/EPI_GUIDE_A.pdf;

(b) A more detailed description of issues relating to prevalence estimation can be found in the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) Scientific Monograph No. 1, *Estimating the Prevalence of Problem Drug Use in Europe* (Luxembourg, Office for Official Publications of the European Union, 1997);

(c)   EMCDDA also has a number of technical tools that provide a further detailed elaboration of the methods described in the present toolkit. The report of a European Union technical working group on prevalence estimation that includes draft guidelines for estimating the prevalence of problem drug use at the national level, can be found at www.emcdda.org/situation/themes/problem_drug_use.shtml or obtained by writing to EMCDDA, Rua da Cruz de Santa Apolónia, 23-25, P-1100 Lisbon, Portugal.

Further research materials and web sites are given in the annex to the present toolkit.

Annual reports questionnaire

Part II of the annual reports questionnaire

In deciding on which criteria to use for defining the population of interest for a prevalence estimation exercise, one criterion should be to adopt commonly used categories, wherever possible. At the global level, there is one mechanism that is designed to provide an overview of the world drug abuse situation, the annual reports questionnaire. The annual reports questionnaire is the mechanism used by Member States, under the international drug control treaties, to report on various aspects of the illicit drug problem to the Commission on Narcotic Drugs. Further information concerning the questionnaire can be found at www.unodc.org.

Part II of the annual reports questionnaire is most relevant to monitoring global patterns and trends in drug consumption. A revised version of part II was adopted in January 2002 to reflect the agreed core indicators of drug consumption (see an overview of drug consumption indicators under "Principles of data collection" at www.unodc.org/pdf/drug_demand_gap_lisbon_consensus.pdf) and to ensure that the questionnaire was sufficiently versatile to allow reporting from countries with different levels of data collection capacity. Specifically, the revised questionnaire provides for global data collection on an agreed set of core drug consumption indicators using the following three levels of reporting: summary expert opinion; unstandardized or partial quantitative data; and standardized quantitative data. While the annual reports questionnaire is intended to provide only a summary data set, it does provide a useful vehicle for encouraging the adoption of multi-source data collection methods on a harmonized set of core indicators. The questionnaire is not intended to be sufficient for all the needs of policy makers, but it can provide a basic structure for data collection efforts. Countries that adopt the core measures found in the questionnaire also ensure that data collection exercises result in information that is compatible with international standards.

Currently, the picture of the global drug situation is built upon data from the annual reports questionnaire in conjunction with other published material on drug consumption, and it relies heavily on data provided by drug information systems at a national and regional level. A copy of the most recent report on the world drug abuse situation can be found at http://unodc.org/cnd_session_45.html.

Common reporting categories

The preferred annual reports questionnaire categories for providing quantitative esti-
mates of drug use are described below.

Time periods.   For each index drug, the annual reports questionnaire asks whether
it was "ever used" (sometimes called lifetime use) and whether it was used "in the
past 12 months". Other measures typically used in this area include "in the past
month" (sometimes called current use) and the concept of "daily use" in the past
month, which, typically, is operationally defined as "using on 20 or more days in
the month prior to interview".

Drug types.   It is important to specify drug type in any prevalence estimation study.
Choosing which drugs to concentrate on will be determined by the nature of the
drug problem. In the annual reports questionnaire (part II), the list of drug types
provided below is used. The list covers the main types of abused substances but is
not comprehensive. It may therefore be amended by specifying other substances.
The relevant drug types are as follows:

(a)   Cannabis-type type: marijuana (herbal) and hashish (resin);

(b)   Opioids: heroin, opium and other opioids (for example, morphine, fentanyl and
buprenorphine);

(c)   Cocaine: powder (salt) cocaine, crack cocaine and other forms of cocaine;

(d)   Amphetamine-type: amphetamine, methamphetamine and "Ecstasy-type"
amphetamines;

(e)   Sedatives and tranquillizers: barbiturates and benzodiazepines;

(f)   Hallucinogens: lysergic acid diethylamide (LSD) and other hallucinogens;

(g)   Solvents and inhalants: a range of volatile substances such as gasoline/petrol,
adhesives, aerosol products (for example, paint sprays, air fresheners and anal-
gesic sprays), anaesthetics (for example, nitrous oxide), cleaning agents, sol-
vents and "room odorizers" (for example, amyl nitrite and butyl nitrite);

(h)   Other drugs: any illicit substances that do not fall within the above-mentioned
categories (for example, gamma-hydroxybutyrate, anabolic-androgenic steroids
and khat).

Route of administration.   The harmful consequences of drug abuse vary not only by
drug type, but also by the way the drug is consumed (route of administration). In
particular, drug injecting is associated with the most severe consequences, such as
transmission of HIV infection, an increased risk of overdose and an increased risk
of drug dependence. Typical routes of administration include: oral consumption
(eating, drinking and swallowing); sniffing or snorting (inhaling up the nose);
smoking or inhaling sublimate ("chasing the dragon"); and injecting.

Most information systems attempt to differentiate data on drug users according to
the route of administration, especially for drugs that can be injected. Injecting behav-

iour is often considered so important that a separate question will be used to ask about lifetime and current injecting experiences and possibly the extent to which the individual has shared injecting equipment with others.

Problem drug use. The methods described in the present toolkit have been designed primarily to try to better estimate the more hidden and problematic patterns of drug use that are particularly difficult to observe by other means. That includes the regular use of heroin or crack cocaine or drug injecting. Sometimes a composite definition of drug consumption behaviour is used for prevalence estimation, which combines a number of behaviours into a single category. For example, in the European Union, estimates are made of "problem drug use", which is defined as "injecting drug use or long-duration/regular use of opiates, cocaine and/or amphetamines" (for further information, go to www.emcdda.org).

Developing assessment tools

It is hoped that the toolkit will assist researchers in achieving recognized and internationally accepted standards when planning their prevalence estimation studies. The structure set out can help position those studies against the background of differing national, cultural, social and geographical settings. The methods and definitions should be applicable to both highly developed and less developed national administrative structures and data collection systems.

One aim of the toolkit is to make best use of the existing data in different countries. Moreover, part of the objective is to encourage the systematic development of the data in forms that will facilitate future monitoring of drug prevalence. And importantly, it is hoped that the toolkit will encourage and support new data collection studies and new data-recording systems that lend themselves to establishing, in future, coordinated results on patterns and levels of drug use. It is the development of these tools of long-term data-recording systems and the capacity for ad hoc or repeated assessment studies that will in the long run be the most important issue from the GAP project and this toolkit.

Need for guidelines and good practice

Development of expertise in drug abuse prevalence estimation methods is regarded as a priority issue within the overall project and is the linking theme throughout the toolkit. Expertise in technical analysis often requires expert technical advice, in the present case from experienced statisticians. The toolkit points up the rationale and the issues of those methods of analysis to enable the researcher to identify where such advice and support is required.

But there is a more general level of research expertise that is as important as the technically detailed command of particular epidemiological analyses, one involving

the general structuring of epidemiological research studies and the resources required for them. In that connection, there is a role to be played by recognized guidelines for achieving effective research. Those guidelines take the form of practical advice at a non-statistical level in research methods. This toolkit recommends that they should form the basis of research thinking and the approaches to any epidemiological studies that might be organized within a country. The central point here is to develop internal support mechanisms in the countries themselves, between personnel and groups involved in monitoring drug abuse. A further point is to encourage international consultation on those matters within a common framework of ideas and principles. From that point of view, the growth of consultation mechanisms within an accepted set of standards should be considered as important as the development of a common research language among the different groups involved in the work.

The above-mentioned issues are dealt with in an excellent EMCDDA publication, available at www.emcdda.org/multimedia/project_reports/ situation/guidelines_ prevalence_pdu_mb_09-01.pdf, to which the authors of the present toolkit acknowledge a considerable debt.

Drug use prevalence assessment

Chapter I

The present module provides 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, as in the case of estimating illicit opiate use. The central point motivating procedures in such a setting 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 stigmatized 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 such 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 the prevalence of drug use.

Prevalence assessment 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 EMCDDA Scientific Monograph No. 1, *Estimating the Prevalence of Problem Drug Use in Europe*). There has been an associated growth in the number of manuals and review papers on methods (EMCDDA 1997 and 2000; Hickman and others, 2002; International Working Group for Disease Monitoring and Forecasting, 1995; Frischer and others, "Mortality among injecting drug users ...", 1993; Hser and others, 1992; and Reuter, 1993). Where the present module differs from others is in 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, including the following:

(a) 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;

(b) Monitoring key targets of drug policy, for example, 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 those targets is to use estimates of the number of problem drug users in contact with services, in relation to prevalence estimates for the total drug-using population;

(c) Interpreting key harms associated with drug use. The burden of HIV, hepatitis C virus (HCV), fatal overdose and 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 harm to public health, the core problem is to find a method for estimating the number of heroin users, injectors or crack-cocaine users. That is the focus of the present manual.

There are some notable and intended omissions from the manual. The intention 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 life, are excluded. Such 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 and 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 that classical sense, are broadly suitable for a restricted number of questions where biases and errors in responding and general misreporting can be held to a small percentage of the overall prevalence rate. In the drug prevalence field, that means that 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 those behaviours is usually accomplished by incorporating additional questions into the questionnaires already used in existing general

population surveys. Those 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 in the following two areas:

(a) *Undercoverage.* Hard-core drug users occur in significant numbers outside household units, 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;

(b) *Underreporting 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. The problem is exacerbated for more stigmatized behaviours; for example, heroin use is usually thought to be more often underreported than is marijuana use.

Those 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, those provided by the crime survey and psychiatric morbidity surveys of the United Kingdom of Great Britain and Northern Ireland and by the National Household Survey on Drug Abuse in the United States of America) all attest to that. Corrective factors can be applied through ratio estimation techniques, as described in some of the reports of the National Institute on Drug Abuse and discussed below, but generally effective estimation requires procedures that are more specifically tailored to the context of the drug use.

Methods for correcting underestimation within general population surveys

It has been noted that factors related to drug use, such as arrest information and treatment for drug use, are also underreported or inadequately covered in general population surveys. When national information on those factors is available, it can be used to correct the survey figures, by weighting up the number of respondents who do report arrest or drug treatment to what is known to be the correct figure. Using national figures as a benchmark to adjust for underestimation in national survey data analysis is a standard application of statistical ratio estimation.

The relevance for the present toolkit is that when weighting up the respondents for those drug-related behaviours, the associated drug abuse numbers will also automatically be weighted upwards as a consequence. Some caution should be attached to that procedure, since the resultant adjustment to the drug prevalence figures is derived 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 characterizes drug user populations is that they are what are termed hard-to-access populations. They are populations that, as far as official institutions and 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 the tools of a general population sampling frame used by the survey statistician is abandoned because of its unsuitability, only a few options are left for assessing prevalence. Some of those options are briefly considered below.

Area sampling

Not all population surveys 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 those areas. To estimate the size of the overall population or of a defined subset such as problem drug users, the researchers must then physically count in each area the cases that meet the defining criteria. (Alternative formulations of that method use line transects instead of areas.)

In some situations, that is perhaps the only method that can be used and, in some circumstances, it can be a successful strategy. The physical counting of cases implies that cases can be easily 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, that is not a trivial exercise. Furthermore, it implies that the social structures in the country are suitable for such an approach. One example of such a method is provided by the earlier national surveys of drug addicts in Pakistan, 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 undercoverage that are inherent in calculating drug prevalence by using specialist sampling frames, such as records from casualty wards or police arrests, when designing sample surveys. However, such an approach is in general too narrowly focused to allow general inferences about problem drug use.

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

It should be noted that obvious drug user listings such as clinic treatment records, needle exchange listings and the like, which 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 are designed to use. Special techniques are usually required to make the best use of them.

Indirect methods of prevalence estimation

Indirect methods of association are adopted precisely in order to make good use of the specific drug user lists and lists of drug-related behaviours that are available. Those methods begin by recognizing the inadequacies of registries and other existing data sources. They acknowledge that the target populations of the 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 those incomplete lists. In some cases, they make use of supplementary studies of the target population of drug users together with primary data collection, and try to allow for the difficulties of sampling among 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, there is 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 made are subject to assumptions that it is often impossible to verify and that, when mistaken, can lead to bias quite 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 the 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 those approaches is only judged by the extent to which they converge to a common and plausible estimate.

National and local prevalence studies

Indirect methods may be used, theoretically, at a national level to establish prevalence among the general population, and multiplier-benchmark methods, in particular, are sometimes used in that way. More typically, however, they are employed in studies of drug use on a smaller, geographically local scale. At that level, they are more easily organized 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 very often a need for overall national estimates to be made, and one way of doing that is to extrapolate from local prevalence studies to an overall picture. Extrapolation methods used in that 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 the absence of actual drug abuse prevalence figures themselves, in order to make those comparisons.

If the drug abuse indicator data can be organized 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 it is a simple step to combine them to give an overall national estimate of prevalence as well.

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

Specific methods of prevalence estimation

Overview

Thumbnail sketches of the principal methods are given below, to allow the reader to identify and focus on the chapters of the module that are of greatest interest and relevance, while skipping the rest. All those 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. The present section therefore gives a brief summary of how the methods differ and what the essential data requirements are for their application. The attention of the reader is drawn to the EMCDDA guidelines on indirect methods (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 and data on drug-related deaths. Such 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 that national data set is required an estimate of the proportion of the target population who have experienced the event, that is, who have been arrested, who have died etc.; the inverse of that 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 and others ("Estimating the prevalence of opioid dependence", *Lancet*, vol. 338 (1985), pp. 203-205) 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 3,000; that acts as the fixed benchmark in the calculation;

(b)   The death rate amongst drug users in that year, say 2 per cent, or 1 in 50 dying in the year; that provides the multiplier in the calculation.

The estimate of the number of drug users in that year is calculated from those two figures as the population size required for a 2 per cent death rate to result in 3,000 deaths. If 1 in 50 die, then the overall population must have been 3,000 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 are 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 studies, treatment multiplier studies and others.

Capture-recapture methods

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

Capture-recapture 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 recognize 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 that are not on any of the lists.

An early example of the method was given by Ghodse (1980), in which two United Kingdom official data sources were used. The United Kingdom 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, a register that was criticized as commanding poor

compliance. From the United Kingdom death register it was possible to identify the deaths due to drug addiction, and those were presumed to be deaths of drug addicts. The first of the capture lists was itself considered an incomplete figure for the total number of drug addicts notified to the United Kingdom Home Office Addicts Index. The second capture list, by cross-referencing with the first, identifies among the drug-related deaths the proportion of addicts notified to the Home Office Index. Assuming that notification rate to be the same among 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 United Kingdom 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. Those methods aim at calculating 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 that 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 on a set of methods 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 as follows:

(a)   Overall rate among all drug users (both those who are in contact and those who are not in contact) at contact events that are generated;

(b)   Numbers of institutional contacts events.

Such information makes it possible 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 the aim is really to find out 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 specific, strong assumptions about the relative frequency with which 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 the simple multiplier formula, because the pattern of multiple contact events under study can be complex, and because those methods typically would allow for different groups of drug users having different rates of contact. But the point of the method is that it permits inferences to be made about the size of the drug-using population as a whole by interviewing people found at a few carefully selected places.

Issues in making a choice of method

Apart from the specific data availability needed for each method, there are further practical considerations that influence what choice of study can be made. The first 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 enough resources 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, they provide a second-best option. A third factor is whether any interviewing can be done at all for detailed information that is usually necessary, or whether existing records are adequate. With regard to accuracy and reliability, all the 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 etc. 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, the purpose is to generalize 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 they fulfil, in principle, the same function when some prevalence information is known for some areas. In general, those 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 also essentially involves 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. In that process, the "target" regions must have some data sources that are the same as, or very similar to, the "anchor" regions for which prevalence estimates do exist, although they lack the regional prevalence figure itself. Those data sources are referred to the drug indicator data, using 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 using those indicator variables that provides the basis for generalizing the known prevalences.

The procedure is therefore an essentially data-rich procedure based on data that are available through a geographical breakdown of the country.

Role of case studies in the following sections

The toolkit as a whole and the guidelines presented are grounded in practical examples from the existing research literature. Although it is difficult to separate specific isolated methods in the research literature, the examples have been spread throughout the individual chapters, so that they are to be found where they are most relevant. The examples and commentaries constitute about half of 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. The 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 were chosen, wherever possible, for countries outside the European Union, 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 are not themselves to be regarded as flawless models.

Using multiplier-benchmark methods

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 field of drug epidemiology. 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 subsection of the target population of drug users, and generalizes from that subsection to give an estimate of the complete target population by applying a multiplying factor.

A simple illustration of the method will first be given, followed by a more general discussion of the strengths and weaknesses of the method in a wider context.

Simple multiplier technique

The essence of the multiplier-benchmark calculation is that some information is available on a subset of the target population—usually some count of the number of drug abusers making contact with a particular agency—and an attempt is made to use that information to estimate how many more there are in the overall target population. If, for example, the number of drug abusers in treatment in 2001 is known, and approximately 1 in 10 abusers is known to have attended treatment in 2001, then that treatment figure can be multiplied by a factor of 10 to get an estimate of the total number. Those 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 the method and the consequences thereof 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 the corresponding benchmark and multiplier figures are known. Table 1 lists a set of likely data sources that may be available for use as a benchmark in a prevalence estimation exercise.

Table 1. Potential data sources for benchmarks for multiplier methods
to estimate prevalence of problem drug use

Data source	Example
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 outreach workers
Needle exchange programmes	Drug users registered at needle exchange programmes
Casualty ward	Drug users attending casualty ward because of an overdose
Laboratory	Drug users tested for HIV, HCV or hepatitis B virus
Police/prisons	Drug users arrested or imprisoned for drug offences or 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

Source: Hickman and others, "Estimating drug prevalence: review of methods with special reference to developing countries", *Bulletin on Narcotics* (United Nations publication, forthcoming).

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

Case study 1. Toronto multiplier study based on HIV tests
(Basic multiplier-benchmark calculation)

Archibald and others (2001) outlined a multiplier method of estimating the prevalence of injecting drug use, making use of information from laboratories of the number of HIV tests by injecting drug users and of data from surveys of the proportion of injecting drug users that had had an HIV test in a given year. The findings for one city in one year, Toronto in 1996, are presented below. The example requires two elements. The first is a known benchmark figure. That figure, in the present case, is the number of HIV tests made on injecting drug users in Toronto in 1996, which was recorded in routinely collected information as 4,050. That represents the known part of the population of injectors.

To find the total number of injectors, it needs to be determined what fraction of them are unknown to HIV testing records. The second element required by the method is therefore a multiplier that tells how many more injecting drug users in Toronto did not have HIV tests in 1996. That figure can be worked out simply if the proportion of drug users who did have HIV tests during the period is determined. In the example, the proportion of users tested for HIV was known from other studies to be 25 per cent, or 1 in 4. The calculation illustrated below in table 2 is then made simply by noting that if 1 in 4 injectors have been tested, then the total number of injectors must be 4 x 4,050, or 16,200, people.

The method assumes that an unbiased estimate of the multiplier is available. Ideally, that estimate will be obtained from a representative sample of problem drug users collected for the specific time period and place corresponding exactly to the time period and geographical location of the benchmark figure to be used. In practice, that rarely happens. In the Toronto case study, 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.

Table 2. Using HIV test numbers to estimate the number of injecting drug users

Item	Applied values	Estimates
Benchmark (B)	Number of HIV tests by injecting drug users in 1996[a]	4 050
	Proportion of injectors reporting getting an HIV test in the previous year[b]	25 per cent
Multiplier (M)	Multiplier calculated as 1.0/0.25 (i.e., 1 in 4)	4.0
Population estimate	Benchmark times multiplier (B*M)	16 200
Laboratory	Drug users tested for HIV, HCV or hepatitis B virus	
Police/prisons	Drug users arrested or imprisoned for drug offences or 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	

Note: To facilitate presentation and analysis, the numbers in the table have been changed slightly from those in the original publication.
[a]Derived from laboratory reports.
[b]Derived from community survey of injectors.

Source: Archibald, and others (2001).

What is noteworthy in case study 1 is that it uses standard, routinely collected data—the numbers of drug injectors who had HIV tests in the past year, from official data sources—to provide a benchmark figure for the officially visible part of the drug injector population. And to multiply up to get the size of the whole drug injector population from that officially visible part, it uses information from other published studies. In the example, therefore, no new research study was carried out to provide any information; it was 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 second case study below uses a common alternative benchmark and multiplier system, based on official statistics of recorded deaths of heroin addicts. The benchmark figure is again derived from an existing data source, but the study makes similar compromises in establishing the "deaths multiplier" value.

Case study 2. New South Wales overdose deaths
(Multiplier-benchmark estimation with a drug use indicator)

A second simple illustration of the multiplier-benchmark method is given on the basis of a study using a deaths multiplier. That study sought 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 that by applying multiplier methods to the national heroin overdose data, specifically to heroin-related overdose fatalities. That deaths multiplier method is reported here, although it 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 the present case, no direct or separate study was carried out to establish such an estimate; 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 (for example, Frischer (1998) and Reuter (1993)) users indicated that between 0.8 and 1.0 per cent of people who use heroin regularly will die of a heroin overdose in any given year. That implies that somewhere around 1 in 100 heroin users die each year as a result of an overdose, and so a deaths multiplier would have a value of between 125 (0.8 per cent) and 100 (1 per cent).

Using the benchmark figure

The number of heroin overdoses recorded in New South Wales 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 (see table 3). That broad estimate is similar to the previous estimates made by other means.

Table 3   Using overdose deaths to estimate the number of regular heroin users in New South Wales based on a 1 per cent per annum mortality

(Figures averaged over a period of five years)

Item	Applied values	Estimates
Benchmark (B)	Number of heroin overdose deaths per annum[a]	360
	Proportion of regular heroin users who die of an overdose annually[b]	1 per cent
Multiplier (M)	Multiplier calculated as 1.0/0.01 (i.e., 1 in 100)	100
Population estimate	Benchmark times multiplier (B*M)	36 000

[a]Derived from available mortality records.
[b]Derived from quoted mortality rate in research papers.

Caveats

Of course, the 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 was 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 the 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 the method can produce, at best, only a rough approximation. Improvements in the accuracy of the 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 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 are that the estimate of the mortality rate among drug users ideally requires a specialist, longitudinal, local study, and that, unless the study is very large, it will take a long time to find the result. A further problem is that the multiplier itself is very large. If only 1 per cent of the population is visible, then the unreliability of the estimate is obviously greatly increased.

Nonetheless, the convergence of that estimate with other previous estimates strengthens confidence in the 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 result achieved.

Source: Frischer, "Estimating the prevalence of drug abuse ..." (1997); Hall and others (2000); and Reuter (1993).

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 of drug users dying in the year by the deaths multiplier (the reciprocal of the annual mortality rate among drug users). In countries where statistics on drug deaths are not easily available, a commonly used alternative benchmark is the number of drug users in treatment. To use the subgroup in treatment to provide the benchmark requires the following:

(a) The total number of the drug-using population who were in treatment at some point during the year in question;

(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 United Kingdom research publications).

Case study 3.   New South Wales treatment and arrests data as a multiplier
(Multiple estimates using different data sources)

Background

In Australia during the 1990s (and in many other countries), there appeared to be a rise in the availability and use of heroin. Heroin use and issues associated with it became a major public and political issue and there was intense media debate about the extent of the problem and potential strategies for alleviating or reducing heroin-related harm. Those 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 that point. Indeed, suggestions that heroin use had increased in Australia were based largely on the following four types of information:

(a) Evidence from overseas that the worldwide production of opium had increased;

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

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

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

Against that 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 the 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, in accordance with the disease codes of the International Statistical Classification of Diseases and Related Health Problems, tenth revision (ICD-10), including drug-related fatalities, and both the police and treatment agencies keep comprehensive records. Nonetheless, even given those advantages, estimating the size of the hidden population is, as discussed above, a difficult problem.

While the primary concern was 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 the eight states and territories of Australia. Data were available from the following two principal sources for the State of New South Wales:

(a)   *Arrest data.* The New South Wales police service provided data on arrests for heroin-related offences (possession and supply of heroin) for the period 1997-1999;

(b)   *Methadone maintenance data.* The primary mode of treatment for opioid dependence in Australia is methadone maintenance treatment. 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 of New South Wales Health to ensure 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 methadone maintenance treatment over the period who were regular heroin users as 13,000.

Using that 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 methadone maintenance treatment in the preceding year. From that total, the overall number of regular users could be estimated, multiplying by a factor of 3.0, as 39,000 (see table 4).

Table 4. **Using methadone maintenance treatment patients to estimate the number of regular heroin users in New South Wales**

Item	Applied values	Estimates
Benchmark (B)	Number of patients in MMT at clinics over the year[a]	13 000
	Proportion of regular heroin users who had methadone maintenance treatment in the preceding year[b]	33.33 per cent
Multiplier (M	Multiplier calculated as 1.0/0.333 (i.e., 1 in 3)	3
Population estimate	Benchmark times multiplier (B*M)	39 000

[a]Derived from available clinical records.
[b]Derived from published studies of samples of heroin users.

Robustness of the result

The data were analysed using a variety of techniques, including capture-recapture, as well as multiplier methods. Data over a period of many years and methods of back-projection, originally developed to assess HIV/acquired immunodeficiency syndrome (AIDS), were used as a means of charting historical changes in the number of people who were regular or dependent heroin users.

The application of those different methods produced a surprisingly narrow range of estimates: the six estimates of the number of people in New South Wales 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, the median value was simply calculated as the best estimate, but it may also be possible to use a more sophisticated method for combining different estimates by weighting estimates according to the relative confidence in their application and conclusions.

Consequences of imperfect procedures

The study drew attention to one of the difficulties of making similar calculations using the data records of police arrests. From earlier studies, it had been estimated that about 20 per cent 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 those arrested for heroin-related offences during the study period was approximately 2,400, a figure obtained from the standard police arrests database. Applying the multiplier to the total arrests gave an estimate for New South Wales of 12,000 regular heroin users—a much lower figure compared with other estimates, including the methadone maintenance treatment multiplier method.

There were several reasons why the police arrests data might not be suitable for the multiplier benchmark calculation. It seemed likely that the arrestees among 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 to think that such was the case. That would imply, in the first case, that the multiplier as calculated from the interviews did not exactly match 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 from the studies. That would imply that the derived multiplier would be too small, and applying that to the benchmark figure therefore gives rise to an underestimate of the total number of regular users.

Source: Hall and others (2000); and McKetin and others (1999).

Primary collection of new data

It is recommended that, whenever possible, the researcher should conduct a sample survey of the target population—injectors or problem drug users—as part of the prevalence estimation study. That has several advantages:

(a) The 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 "Using capture-recapture methods", below) of prevalence estimation;

(b) 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;

(c) Finally, interviewing the target population allows questions to be asked, yielding information that could be used in a multiplier study (see "Using multiplier-benchmark methods", below).

The use, in the previous example, of that multiplier and benchmark—the treatment multiplier calculation—makes it possible, in particular, to carry out a special study to determine the multiplier value. In the following example (case study 4), new data are collected for estimating the multiplier along those lines, as well as for establishing a benchmark figure.

Case study 4. Pakistan National Assessment Study 2000
(Multiplier-benchmark study using a treatment multiplier from interviews with key informants)

Background position and assessment

The aims of the overall exercise in Pakistan were 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 provinces of Pakistan, there was little reliable information available on drug use itself. An earlier survey had attempted to estimate the figure using national survey methods, but the results were far out of date.

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

Extrapolation of the local study figures to give estimates for each major province of Pakistan is discussed elsewhere (see "Extrapolation from local to national prevalence estimates", below). 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 the estimates based on different data, thus giving support to the validity of the estimates.

Defining the target population

The target population was defined as heroin users or injectors (of anything)—people referred to as "addicts" for lack of a better term. Cannabis in the plant form or resin (hashish/charros) is the most common drug of abuse in Pakistan—probably even more common than alcohol, the use of which is illegal. But

researchers decided that they could not practically conduct interviews, surveys and sampling that would be efficient for estimating cannabis and heroin or injecting prevalence at the same time.

The target population for the prevalence estimate was restricted to males aged between 15 and 45 because of the following practical considerations:

(a) Social mores would not allow women drug users to be visible at the treatment centres, although over-the-counter opiates were thought to be a considerable problem among females;

(b) Previous work suggested that the age range selected contained almost 90 per cent of hard drug addicts, and that males outside that age range were almost never encountered in treatment.

The definitions used for the benchmark figure and the multiplier for "treatment" were confined to specialist treatment centres for addiction. Treatment in government hospitals was excluded for the following reasons:

(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 per cent coverage) of all specialist drug addiction treatment centres. The starting point of the process was earlier, but now outdated, lists of specialist clinics, updated into a national treatment register. It is hoped that that part of the study has a spin-off for future drug use prevalence work through maintaining a national register of those clinics. A total of 73 such clinics were identified nationwide.

The data-gathering procedures at the clinics allowed for various definitions of the benchmark, but the analysis focused on the definition being set as the "numbers of addicts treated as inpatients in the past year" (see appendix 1 of the main Pakistan report for alternative definitions that were considered). The managers of all 73 inpatient specialist centres were interviewed either by telephone or in person to collect the numbers of addict inpatients during the past year. Medical personnel carried out those interviews.

Establishing the multiplier

Establishing the multiplier was again the most difficult part of the study. 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 in personal interviews. See the section on fieldwork below for the criteria for eligibility to be a key informant. Each was asked:

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

(b) How many of the addicts had been for inpatient 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 that information was used only in forming the secondary imprisonment multiplier. Because the questionnaire was a lengthy data-gathering exercise for each key informant, those pivotal questions were asked at the outset of the interview.

When using those figures to derive a treatment multiplier, key informants 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. Key informants who had encountered less than a threshold number of addicts were also excluded from the analyses.

Sampling in the key informant survey

Key informants were selected on a nationwide basis. Thirty-six geographical locales were purposely chosen—not randomly sampled—across Pakistan to represent the general social structure of the country. They 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. Those 18 pairs spread across the four provinces of Pakistan, with more in the more populous areas.

The interviewers (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, including, if possible, one from each of at least five different social categories in each locale. The categories on that list were: policeman, judge, doctor, health worker, other government official, mayor, councillor, mullah, priest, social worker, teacher, tribal headman and ex-addict. Refusals to be interviewed were not recorded, since no formal approach or sampling structure could be defined within a locale.

Fieldwork in the key informant survey

The organization 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 were carried out by staff trained specifically for the survey: medical workers, doctors, social workers, trainees in those professions and associated ex-addicts, plus one or two more suitable volunteers. To ensure comparability across a large survey workforce and to ensure that precisely the right information was gathered, questionnaire-based structured interviews were used in all cases.

To control the quality of the information across such a wide-ranging study, there was strict supervision of the questionnaire flow by four designated regional supervisors. They ensured that all completed, spoiled, wasted or unused questionnaires 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 18 locales across the country, travel between interviews was still a major problem (for example, one interviewer had to travel by camel for two days).

Analysis and results

Calculations using the benchmark of number of male inpatients 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 prevalence estimate for Pakistan, the reader should see section E on extrapolation.

The primary benchmark calculation was made by following the procedures described below.

Specialist inpatient treatment clinics existed in approximately 30 per cent of the study locales and none outside the study locales, so that benchmark figures could be calculated only for those treatment locales. Benchmark totals were constructed for each of those locales by counting male heroin-abusing or heroin-injecting inpatients aged 15-45. An adjustment was made to allow for the fact that about an estimated 10 per cent of those 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, including the principal city of the province and all other locales.

Multipliers were derived for each locale by taking the number of addicts that a key informant had encountered in the past year divided by the number of those who were treated. Of the various ways in which different information supplied by key informants could be pooled into a single multiplier, the median value among the qualifying key informants was selected. That value was a figure that was not unduly influenced by extreme responses, and would not depend upon whether the averaging took place over the proportions treated or the multipliers themselves. In that context, key informants were pooled in the same way as the benchmark calculations were, within the principal city of the province, and across all other locales in the province.

The pooled benchmark figure was multiplied by the median multiplier to give a total number of addicts in each of the four provinces with and without the principal city. Two further estimates were made to give some idea of the variability in that procedure by taking the lower and upper quartiles for the multiplier instead of the median. Repeating the entire calculation for those 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. That previous survey, however, and the updating surveys with which it was subsequently combined made far broader assumptions than the 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 the study, as it usually is in other studies. Here it could be made in several ways, and that which is chosen may influence (up or down) the answer. For example, given the different responses of key informants in each locale, it could have been calculated by: averaging individual multipliers; averaging individual estimates of the percentage in treatment; averaging those at the level of the locale or the province; and pooling the numbers of addicts encountered and treated before averaging. In fact, prevalence based on the last-mentioned procedure was also calculated and gave results very similar to the method actually used.

Furthermore, the exclusion criteria for allowing key informants to qualify for entry into calculations are certain to affect the result, including the effect of the threshold number of addicts known to key informants and the effect of excluding treatment-related key informants (or prison- and police-related key informants). Information gathered from the key informants would of course be prone to errors of reporting. The distinction between a government hospital and specialist treatment may not have been clear in the mind of a key informant when reporting the estimate of the numbers of addicts encountered who had been in treatment. Furthermore, the key informants may not have 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 catchment areas of the treatment centres when calculating 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 emphasized very strongly.

Source: Rehmann, Griffiths and Taylor (2002).

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 characterized 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 information related to drug use can be gathered about those 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 the area in the last 12 months?" Pooling information across core respondents will give an estimate of the proportion of drug users in treatment.

Heterogeneity and stratification of the population

The foregoing case study 4 targeted only the male population between 15 and 45 years of age. 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 that range, and partly because so few addicts were thought to be in that group. Confining attention to the targeted age range helps focus attention on 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. That division of the problem into sections—first, male and secondly, female—is an example of stratifying the population. The aim is to break down a heterogeneous population into more homogeneous subgroups 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 versus non-injectors, employed versus unemployed, and so on. Any characteristics on which information is available to construct separate multipliers and benchmarks should be considered. Very little is lost if the division is in fact unnecessary—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 subgroups more homogeneous.

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. That was done because it was suspected that, within those 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 the procedure is in its general applicability, which is determined by having data on the following two things:

(a)   The required benchmark—for example, the number of deaths among drug users;

(b)   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. That flexibility is the core of acceptability of the method.

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, that 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 the 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.

The assumptions on which the method is based should be carefully considered. First, it must be assumed that the benchmark data are accurate. Unfortunately, routine data sources can be notoriously inaccurate, because of underreporting 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 that type of underreporting.

The method assumes that a correctly defined multiplier is available. When calculating the multiplier, the key question is really the following: Is the person recorded in benchmark figure? If so, then the multiplier matches the benchmark successfully (even when the benchmark underrecords 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 the precise equivalence of benchmark and multiplier definitions. That 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 an unbiased estimate of the multiplier is available. Ideally, that estimate will be obtained from a representative sample of problem drug users and collected over the specific time period and for the specific place corresponding to the benchmark to be used. That 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 a multiplier estimate required an unbiased estimate of the proportion of injectors registered with a needle exchange, 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 it must be assumed 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 those areas together in a single multiplier can be misleading. The use of stratification of the population as discussed above (see "Other methods for estimating the multiplier value"), can be used if data are available separately for each stratum to overcome that danger. Identifying when it is necessary is primarily a question of looking at the data and using prior beliefs to make a judgement.

Violation of one or all of those assumptions is clearly possible, providing ample opportunity for the study to give an inaccurate prevalence estimate. It is unwise to rely upon a single multiplier estimate (see "Accuracy of estimates of prevalence levels", below).

Using capture-recapture methods

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, lists that, in the case of the present study, 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 per cent perfect without any underreporting of its intended cases. The overlap between lists is the people on more than one list.

The aim is to ascertain how many problem drug users should be on the list if it were a complete list of the target population, for example, of all problem drug users in a city in a specific 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. What a contingency table is and what that analysis might mean is explained 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 now developed as an important method in drug use epidemiology that can be adapted and applied to most local situations. Examples of capture-recapture studies and guidelines commissioned by EMCDDA have been cited in the present study and are available electronically (see chapter I and the annex for details).

A general introduction to prevalence estimation including capture-recapture appears in the *Bulletin on Narcotics* (Hickman and others (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 United Nations web site (see chapter I and the annex).

Capture-recapture with two data sources

The present section begins with a description of the simplest example, a study based on two data sources, followed by a more complex study based on multiple data sources, ending with capture-recapture studies made in the absence of routine data sources.

An example will first be given, and the theoretical assumptions required to justify the calculations will follow an illustration of the method.

Case study 5.   Bangkok study—a practical example
(Basic two-source capture-recapture study)

Below is a practical example based on a study conducted in Bangkok in 1991 by Mastro and others. The following two data sources were used:

(a)   Lists of opiate users enrolled in methadone treatment programmes from local specialist drug clinics in Bangkok. They were routinely collected attendance records for April-May 1991;

(b)   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 records of the specialist methadone maintenance clinics, and from 8,212 people in the survey conducted in police stations, they identified 1,540 opiate users. Those two sources are the incomplete lists of opiate users in 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 sources of data on opiate users, the full name, sex and dates of birth were used to match between the data sources. They found 171 people in both data sources, that is, opiate users reported as being on methadone treatment, who had been arrested and whose urine was positive for opiates.

Table 5 shows the numbers observed in the two data sources and the overlap between them, setting them out as a contingency table, followed by the calculations needed to estimate the prevalence. For the two data sources, methadone treatment attendees and opiate urine-positive arrestees, it is assumed that, among the total number of opiate drug users, the proportion who can be found in the clinics is the same as the proportion found among arrestees and also among non-arrestees. Filling in the table on that assumption gives an estimate of 36,600 injecting drug users (0.5 per cent of total population) in Bangkok in 1991.

Table 5.   Estimating the number of opiate users in Bangkok, 1991

Found in S1[a]	Found in S2[b]	Number of people identified	Key
Yes	Yes	171	Matched in S1 and S2 (m)
Yes	No	3 893	Found in S1 only (c)
No	Yes	1 369	Found in S2 only (b)
No	No	??	"Hidden" (x)
Total in S1	Total in S2	Total	Number of opiate users in the population (N)
4 064	1 540		
(m + c)	(m + b)	—	

[a]S1 = Methadone maintenance clinic attenders
[b]S2 = Arrestees with urine positives

The number of opiate users in Bangkok in 1991 is therefore estimated as follows:

Number observed = m + b + c = 171 + 3,893 + 1,369 = 5,433

Number hidden (x) = 1,369 * 3,893 / 171 = 31,166

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

Rounded estimate of injecting drug users in Bangkok in 1991 = 36,600
(32,000 to 40,800) (95 per cent confidence interval)

In the 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, that 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 that corresponds to the annual number. In other words, it may be assumed that the potential number of opiate users not included in the estimate was negligible, that is, the number of new opiate users or the number of opiate users that ceased drug use, died or left the area (see "Assumptions", below).

Bearing in mind the original definitions, table 5 is an example of a contingency table, a two-by-two contingency table with four cells. It is incomplete because the number of those who are not on either data source, the hidden or unobserved injecting drug users, is unknown. 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. The assumptions required by that method are noted below, but one in particular should be mentioned immediately. In two-sample capture-recapture it is assumed that being on one data source is independent of being on the other. That is, it is assumed above that among 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 5,433 users in the combined data sources and estimated that there were 36,600 altogether—in other words that about 1 in 7 (5,433/36,600) opiate users were observed in the study. That 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 that cannot be determined 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 that point in mind when planning a study. Ideally, data can be collected on a substantial number of people to reduce the ratio of observed to unobserved people, but that is not always possible in studies of problem drug use. Alternatively, a multiple-data-source capture-recapture study, outlined in the next section, can be used. It is still possible to do a two-sample study, but, if that 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 the data sources are positively dependent, and as a maximum estimate if the data sources are negatively dependent (see section below on independence).

Source: Mastro and others (1994).

Capture-recapture with multiple data sources

The term "multiple data sources" refers to three or more lists: for example, problem drug users in treatment, arrested, in homelessness hostels, or attending accident and emergency clinics. Matches are then made across those 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 modelling program. Filling in the contingency table in the previous example is an instance of very simple log-linear modelling—so simple that it does not require a program, or even a computer at all, to do it. A description of the calculations will be avoided using three or more data sources for estimating the number of problem drug users. Statistical software packages can do those calculations, and they look extremely complicated when written down. The assumption of independence used in the previous example, and that was vital for the calculations, now has several alternative forms, and those are discussed in the next section.

Readers interested in the estimating equations are referred to Bishop and others (1975) and Hook and Regal (1995). 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 modelling) program that is necessary to analyse the data. EMCDDA guidelines give detailed examples of the use of the Generalized Linear Interactive Modelling (GLIM) package (an interactive statistical analysis program) and of a statistical software package for the social sciences (SPSS) 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 in S2 than someone who is not in S1; or a negative dependency if a person in S1 is less likely to be in S2 than someone who is not in 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, 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:

(a)   Collecting three or more data sources of problem drug users;

(b) Matching the data sources—identifying people who are on more than one data source, and which data sources are involved:

(c) Preparing a multi-way table for analysis in a data file (see example below);

(d) Analysing the table using Poisson regression;

(e) Fitting model with interactions between the data sources corresponding to potential dependencies;

(f) Selecting the best-fitting model based on standard lack-of-fit measures (a statistician will give advice in that connection (see EMCDDA guidelines));

(g) Use the model to estimate the number in the unobserved part of the population and also to calculate confidence intervals;

(h) Repeat the whole analysis if possible for different subgroups (males, females, different age groups etc.).

An example is given below of a multi-source capture-recapture analysis from a study in Glasgow in 1993 by Frischer and others.

Case study 6.   Glasgow study of injecting drug users
(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 the following four sources for a one-year period:

(a) Positive and negative HIV tests of people whose exposure to risk was reported as injecting drug use;

(b) People who went to specialist drug agencies for treatment of their drug problem (including heroin, other opiates, cocaine, amphetamine and benzodiazepines);

(c) People arrested for a drug-related crime (other than cannabis);

(d) People registered with the local needle exchanges.

In such a study with multiple recapture data sources, drug injectors have to be matched across all four sources, so that the number of those who were found at each combination of sampling points can be calculated. In total, 3,444 records were collected: 508 from the police data source (S1); 1,179 from needle exchanges (S2); 507 from HIV test laboratories (S3); and 1,250 from specialist drug treatment agencies (S4). After matching for duplicates, 578 people were found to be on more than one data source (for example, 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 table 5 included interactions between three data sources: needle exchange (S2), HIV tests (S3) and drug treatment (S4). By contrast, police arrests for possession (S1) was included in no interaction terms in that best-fitting model, implying that 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 those circumstances where data sources are dependent, 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 the 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 per cent of adults aged 15 to 54 in Glasgow. Sufficient data were collected in the study to enable further analyses (that is, a further modelling exercise) of the data by sex and by age group. That is shown in table 6. It is called stratification and is one of the ways of dealing with the problem of heterogeneity (see "Assumptions", below).

Table 6. Multi-source capture-recapture data from a study of injectors in Glasgow, Scotland, 1990

Found in S1[a]	Found in S2[b]	Found in S3[c]	Found in S4[d]	Number of people identified
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	**Total in S2**	**Total in S3**	**Total in S4**	**Total population**
508	1 791	507	1 250	—

[a]S1 = those arrested for possession of illegal drug (excluding cannabis)
[b]S2 = those registered with needle exchange programme
[c]S3 = those tested for HIV with exposure as injecting drug users
[d]S4 = those in specialist drug treatment

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. That model produced the estimates given in table 7.

Table 7. Estimating the number of injecting drug users in Glasgow—numbers and prevalence estimates

Item	Observed/known	Estimate of number hidden	Estimated total	Estimated prevalence (percentage)
All	2 866	5 628	8 494	1.4
Males	1 977	3 567	5 544	1.8
Females	889	2 349	3 238	1.0
15-19 years	264	640	904	1.0
20-24 years	1 137	1 613	2 750	2.6
25-29 years	878	1 724	2 602	2.7
30-34 years	342	796	1 138	1.4
35+ years	245	1 273	1 518	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 and needle exchange: 2,508;

(b) Police arrests list: 508;

(c) Matches across police and combined lists: 150.

The 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 that it is equivalent to using two independent data sources. That 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 multiple-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.

The example illustrates a further point: only in rather special circumstances—such as those revealed in the study—is it a good idea to combine different data sources as though they were just one source. In general, although it is all right 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. That is probably not going to be the case if very different source lists, such as HIV registration and police arrests, are combined, and it will be a violation of the homogeneity assumption (see "Assumptions", below).

Source: Frischer and others, "Estimating the population prevalence of injection drug use ..." (1993).

What software to use

Two sample studies can be conducted using pen and paper, a calculator or a spreadsheet. In most cases, multi-source capture-recapture methods require statistical software. Bishop and others (1975) do give the formulae for calculating population estimates by hand. Guidelines from EMCDDA give examples of the use of the standard packages, GLIM and SPSS.

Data sources and how to match across them

Criteria for data sources to use

In capture-recapture studies, the best data sources ideally do the following:

(a) Identify clearly the target population to be estimated (for example, heroin users and injectors or problem drug users);

(b) Collect large amounts of data that can be used for matching;

(c) Collect potential stratifying variables;

(d) Provide the data in electronic form so that it need not be collated by hand.

Unfortunately that rarely happens, so the data sources needed must be good enough, in particular with regard to the following questions:

(a) Do they identify the target population to be estimated?

(b) What identifiers are collected, and what will the data owners release?

(c) How will the data be collected or provided?

(d) How many cases will be provided?

(e) 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 possible, data sources should therefore be selected from the criminal justice field and from the treatment field.

Routine data sources that may be available are shown above (see "Simple multiplier technique"). It is best to undertake an inventory in the study area to find out how many of them are available locally, and whether there are any other potential sources. In aggregate form (grouped), those sources could be used as benchmarks for multiplier methods, and, in disaggregate form (ungrouped or one line of data for each report), they could be used for capture-recapture studies. Remember that data sources for capture-recapture do not have to be complete (that is, include every possible case) but they do have to be accurate and reliable (that is, there must be correct identifiers and accurate drug information).

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

Matching across data sources

Unless thousands of reports are available, the best way to match is manually. But spreadsheets or databases can help match by sorting the data sources in different ways to identify people on two data sources. For example, two lists sorted by sex and date of birth can be used to see if someone has the same name or initials; or lists sorted by sex and name could be examined to see if they have the same date of birth.

It is sensible to decide what constitutes a match, since 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 26 April 1563 and in another it was 20 April 1563. The more data available for any one person the easier it is to match, but possible errors between data sources will still have to be contended with. 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 that as a match.

Whatever is chosen, it is important for it to be clear and specific. One way to do that is to adopt a rigid definition, do the analysis and then take a looser definition and compare the prevalence estimates. For instance, Mastro and others in the Bangkok study had six criteria based on six different pieces of information for each person, reflected in table 8.

Table 8. Matching criteria for prevalence estimates in Bangkok study

Match	Sex	First name	Surname	Age	Date of birth	Thai identification 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 table 8, 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 and others in the study in Glasgow used the following matching criteria based on five pieces of information, reflected in table 9.

Table 9.   Matching criteria for prevalence estimates in Glasgow study

Match	Sex	First character of surname	First character of first name	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 no local data sources are available, or none collect any identifying information, or confidentiality rules prevent their use? Should one give up, abandon the project, 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 has something in common that can be used to identify them: for example, street injectors or street sex workers that inject. Two examples of that method are given (case studies 7 and 8).

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 that it could use the calculations outlined above to estimate the total population. That 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 fieldwork, but used a different set of statistical methods to estimate the population size. The use of such a method should be considered only in consultation with a statistician who has experience of open capture-recapture models and estimation exercises.

Case study 7.   Bangladesh Dhaka study—estimating prevalence of sex workers
(Capture-recapture without routine data sources)

In the present example, the method was used for estimating the number of street-based sex workers in Dhaka. In the absence of any secondary data on the population of sex workers in the city, and of any routine data sources identifying that population, thorough ethnographic fieldwork was conducted. Several categories of key informants were interviewed, including 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.

With the help of sex workers themselves, coloured cards (the first time, red) were distributed in all known locations of the city starting from late evening

---

to midnight. The coloured card could be used for free health check-ups in one of the reproductive health and primary health-care clinics of the city. Since the cards were numbered, a sample of number of cards were distributed, that is, the capture sample was easily obtained. The same exercise was repeated after 7 days. That time the colour of the card was green, which could similarly be used for free health care. The second time, the sex workers were asked only one question—whether they had received a similar red card before. The estimated number of street-based sex workers at around 5,000 was derived from the data by the simple calculation in case study 5.

Source: SHAKTI Project, CARE Bangladesh (Dhaka).

Case study 8. 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 were working on each night and if they had been observed on previous nights. In total, over a period of seven months, 1,145 contacts were made with 206 women (147, or 71 per cent, 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 population changed at approximately 8 per cent per week, giving an annual total of about 1,150 prostitutes.

Source: Frischer and others, "Estimating the population prevalence of injection drug use ..." (1993).

Credibility of the estimate

How is it known whether the estimates are credible? The simple answer is that it is not known. Things can go wrong (see "Assumptions", below). But the researchers own local knowledge can be used to judge whether the estimates are credible, which is how animal ecologists deal with the uncertainty of the method. The following questions would therefore be asked:

(a) Do the estimates fit with what was expected?

(b) Do they fit with the evidence base, that is, estimates using other methods or estimates from different years?

(c) Are they ridiculously low or do they suggest that 1 in 5 of the population is an injecting drug user? If so, they are probably wrong;

(d) Is the lower 95 per cent confidence interval negative? If so, other methods can be used, but the accuracy of the data sources and the matching should be checked;

(e) In multiple-data-source capture-recapture, are the dependencies between data sources believable?

Case study 9.   Jersey study
(Heroin users giving false names)

The advantage of multiple-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, 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 2,000 injecting drug users, which would give Jersey a very high prevalence (that is, over 2 per cent overall, and over 4 per cent in those aged 15-54).

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

That 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.

Source: Fitch and others (2001). Responding to drug and alcohol use in Jersey: key findings report. Drug Strategy Unit, States of Jersey.

Assumptions

Finally, consideration will be given to the assumptions underpinning capture-recapture. The guidelines are intended to be mostly practical, rather than theoretical, so only a brief outline will be given. But it is important to have some understanding of the assumptions being made in order to interpret the findings of capture-recapture exercises. The implications of each assumption for the practice of capture-recapture will also be discussed.

Population closed—no deaths or new cases or migration into or out of the study area during the study

Such a condition is clearly impossible to guarantee. So the implication is that the length of the study is small in relation to the length of time that 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 that assumption. In comparison, a study using

five years or more of data risks including substantial numbers of problem drug users who were not using drugs over the full study period either because: they ceased using drugs; they have died or moved out of the area; they only recently started using drugs; or they recently moved into the area. In addition, the assumption suggests that when capture-recapture techniques are applied 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, which are usual in animal ecology but demand a different set of equations and statistical expertise. Only one study is known to have employed them in drug use (see case study 8 above)

No misclassification—individuals are correctly matched across all data sources

In human studies, the implication is that someone who is on more than one list is correctly identified. Every effort must be made to ensure that the data collected are accurate and reliable and the hope must be 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 the full names of people. Instead, information must be collected that allows matching across data sources without disclosing the identity of the person (see chapter II, "Capture-recapture methods", on ethical issues involved in data collection).

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

In the study in Jersey, misclassification bias led to an overestimate of the prevalence of heroin use and that 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, a number of true matches were 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 the fact of heterogeneity must be acknowledged. If there are enough

people and enough information on them, the population can be stratified in order to make an estimate. Separate estimates can be made for males and females, or by age group and ethnic group. If statistical support is available, more complex models can be fitted (see "Using advanced modelling techniques", below).

Representative—the data sources are representative of the population of problem drug users to be estimated

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 misclassification, then the study is designed to estimate how many diagnoses relevant to each of the specific data sources in the study have not been reported. Hence, that would be a study of "under-ascertainment".

Do the data sources to be used adequately define the target population of problem drug users? Defining problem drug use is different from defining, for example, diabetes. First, it is defined by a range of clinical, social and criminal problems, which means that it should be as inclusive as possible, 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. That 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 that 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 the direction and rough size of the bias are known, then two sample studies can provide very useful indications of the minimum or maximum size of the population. That is because "negative dependence"—the fact that if a person is on data source A, he or she is then less likely to be also on data source B—means that the calculation overestimates 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 underestimate of the true prevalence.

Three sample studies can drop that assumption, but that increases the computational complexity. Log-linear models are used to analyse the data, which in most cases means that a statistical software package and possibly statistical support are needed. Multi-source capture-recapture assumes that there is no interaction between all sources—that is, 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, that 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, seven models can be fitted, but with five, 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 the data sources are obtained, 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—that is, 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 those subgroups.

The key issues may be summarized as practical advice as follows:

(a) Be aware of potential problems;

(b) Try to find multiple data sources if at all possible;

(c) Select data sources carefully (try and get a range of different types of data source);

(d) Check that names are available and reliable;

(e) Seek corroborating evidence and ask whether the estimate is credible and whether it fits with other estimates using other methods;

(f) Improve the data sources for use next time.

Using advanced modelling techniques

There is a further set of modelling techniques that are of a more complex nature, requiring a more statistical analysis than the simple multiplier-benchmark and capture-recapture methods described above. For using those techniques, it is very strongly suggested that the help of an experienced statistician is required. In general,

the data demands are also greater, in terms either of routine data or collecting data specifically for the estimation exercise. In this section, therefore, some of those methods are summarized and some key references given in case there is a desire to examine the methods in more detail. A brief description will be given of covariate modelling in capture-recapture methods; enhanced event-based multiplier studies; truncated Poisson; back-calculation models; and dynamic modelling.

Covariate models in capture-recapture methods

Tilling and Sterne (1999) have developed a variation on the capture-recapture methodology that allows covariates (such as age, sex, area of residence and ethnicity) 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 and adjusts for heterogeneity. Traditionally, heterogeneity is dealt with by stratifying the data set into subgroups 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 subgroup models to be run. The covariate model is much more efficient. But the price for that 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.

Event-based and related models

Simeone and others (1997) proposed a modified and enhanced multiplier study, which they piloted in Chicago, and is being used in a border town in Mexico (information provided by Elena Medina-Mora). The method uses "events" as the multiplier and benchmark. Before briefly describing the method, consideration should be given to 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 (for example, deaths). Person-counts usually can be regarded as a simple form of event data, where work is done with an ever/never (in the time period) definition rather than a count. Sampling events is sometimes easier than sampling people if the aim is to collect personal or drug history information from the drug user. For example, if people are to be sampled 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 each person is counted only once, no matter how many attendances are made in the period, people who tend to turn up frequently are more likely to be interviewed 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 the 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-reports 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.

Truncated Poisson stochastic models of event counts

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

Note that that 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 (Hay and Smit, forthcoming). There is a need to:

(a)   Assume that repeat events are independent of the event history of an individual, that is, there is the same probability of being arrested at every point of time throughout the study, regardless of previous arrest history;

(b)   Assume that individuals are all equally likely to generate an event, although that assumption can be relaxed with some difficulty if there are enough individuals with common values (even then, there must be some assumption about the personal history values of the zero-attendance group, since if some individuals have an almost zero probability of being arrested, then the procedure must of necessity miss counting/estimating them).

Those unreasonable—and essentially uncheckable—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 capture-recapture study.

Back-calculation methods

Law and others (2001) 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 those three components allows estimation of the third. Typically, the distribution of the incubation time and the incidence of the end-point are assumed to be 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.

Dynamic models

Finally, one rapidly developing area of research into prevalence estimation methods is the field of dynamic models. Broadly speaking, they 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. The data are analysed jointly in a single dynamic model, in which the principal 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 that 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. That 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. That 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, 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 the use of treatment clinics is considered 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 on government 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 organization to another, with no overall structured uniformity nor consistency.

Finally, there is sometimes a question of the ownership of data that is difficult to organize at a national level. In particular, it is often the case that in more rural areas policing records are kept differently; but sometimes 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 those 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 nationwide 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 policymaking. 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, results are generalized 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 that, the 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

Consideration is first given to an example of the simplest kind—extrapolating from one local area to one other area, although in fact the area in the example is the rest of the whole country.

Case study 10. Extrapolation in the New South Wales study
(Simple single-point extrapolation with a drug-use indicator)

While a local study had produced an acceptable estimate for New South Wales of approximately 37,000 regular opioid users (see case studies 2 and 3), there was also considerable interest in extrapolating that finding to the whole of Australia. One very simple way to do that would have been just to multiply the estimate of the prevalence for New South Wales by the number of people in Australia. That would effectively assume that the prevalence per 100,000 of the population in New South Wales 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 New South Wales, 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 that 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 New South Wales 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 New South Wales, 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 to the size of the general population, or to the numbers of 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, the multiplier (of 2.0) is used to estimate the number of people in the whole of Australia who are heroin dependent. That gave an estimate of 2 x 37,000, or 74,000, people who were heroin-dependent in Australia as a whole. That 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.

Source: Hall and others (2000); and Lynskey and Hall (1998).

Even in the basic example given above, several fundamentally important points are demonstrated.

First and most obviously, the researchers chose not to simply apply the prevalence rate for drug abuse in New South Wales (termed the "anchor point") to the rest of the country (sometimes termed the "target" area), where it is unknown. When can a regional prevalence rate 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 urbanization and proximity to drug supply routes, can make the assumption that one local area is typical of another too simplistic to be useful. In the 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 10. How many people, overdose deaths and drug users there are in the anchor point area are known, and that makes it possible to calculate the prevalence rate— for example, per 100,000 of the adult population—for overdose deaths (the indicator) and for drug abuse. In the target area, the available information makes it possible to calculate the indicator—the overdose death rate per 100,000 of the adult population, the prevalence rate for drug abuse needs to be estimated.

Table 10. Data structure for extrapolation from known anchor-point drug prevalence levels to target areas where it is unknown

Item	Population (Number)	Drug abuse indicator (overdose deaths)		Drug abuse prevalence	
		(Number)	(Rate)	(Rate)	(Number)
Anchor point					
New South Wales	Known	Known	Calculated *(a)*	Calculated *(c)*	Known
Target					
Rest of country	Known	Known	Calculated *(b)*	Unknown *(d)*	Unknown
Overall	—	—	—	—	—

Technically, the steps involved are:

(a) To calculate the drug indicator prevalence rate for the anchor and for the target;

(b) To calculate the drug abuse prevalence rate for the anchor;

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

(d) To calculate, if desired, the actual numbers of drug abusers, as opposed to the abuse rate, in the target area;

(e) 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 that way will mimic the calculations of case study 10; they also show the assumption involved and make it possible to generalize the procedure in several ways—as discussed in the second and third points below.

Secondly, 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 table 10, provided the drug abuse indicator information for them is available, and to produce different estimates for 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. That raises a new element—how to combine the different pieces of information offered by the anchor points that have been added. In fact, that 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 the calculations is non-specific, that is, any indicator no matter how it is derived will do provided that: it is relevant to drug abuse and likely to have a stable or uniform relationship to it across anchor and target points; and information is available for it to be calculated in a uniform and parallel way across anchor and target points.

Finally, 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 10.

Before continuing, it should be noted that there are some technical, statistical considerations that apply when more than one anchor point is used in a regression analysis. Those considerations are as follows:

(a)   There is a question about whether to 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 to use the rates themselves as data or to use log-transformed data;

(b)   There is a question about whether to 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;

(c)   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

that might be checked. That 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;

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

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

Extrapolating using several drug indicators

One more generalization may be added to the method outlined in section 1 above 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 below 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 those 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, that 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, however, in that type of indirect estimation situation, 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. That is carried out using a principal components analysis, which is designed specifically to do 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 United Kingdom study using multiple indicators
(Multiple anchor-point extrapolation)

The data available for the study are shown in table 11, 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 the United Kingdom. Note that the anchor points that are used are themselves spread out over the country as a whole.

Aims

EMCDDA has produced methodological guidelines for national drug prevalence estimation. That paper pilots the methods to estimate prevalence for the United Kingdom and provides a commentary on the methods and resulting estimates. Three types of method were used to estimate prevalence: the multiple indicator method; multipliers applied to drug-treatment records, HIV estimates and mortality statistics; and the British/Scottish Crime Surveys. The present case study reports only to the multiple indicator.

Definitions

The definition of "problematic" drug use in the present study follows that of the EMCDDA working group, which defines it as "intravenous drug use or long duration/regular use of opiates, cocaine and/or amphetamines. Ecstasy and cannabis are not included". That definition is suitable for studies based on routine sources that 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 and others (1994)). The method was first used in the United States (Woodward and others (1984)) and has been described more fully elsewhere (Wickens (1993)).

The key assumption of the 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 that principal components analysis can be used to extract the main factor that explains the largest amount of variance in the indicators.

The steps below summarize the process used to analyse the United Kingdom data.

Step 1.   A range of indicators or predictors of the prevalence of problematic drug use, which are available for all geographical areas of the country, is obtained. The United Kingdom was divided into 11 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, reflecting the availability of the required data.

Step 2. General population figures for each of the 11 geographical areas are obtained from census information.

Step 3. The following drug abuse indicator variables were collected for all regions covered by the the United Kingdom pilot project: *(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 HIV related to injecting drug use; and *(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. They provide four anchor points for the extrapolation. The data and the sources of those estimates are shown in table 11

Table 11. Drug abuse indicator data for 11 regions of the United Kingdom and drug abuse figures for four anchor regions

Region	Population	Drug use indicators (see key)					Drug users (number observed)
		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	44 410
West Midlands[b]	5 150 246	7 125	5 398	4 322	26	193	13 130
Northern and Yorkshire	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	5 870	2 282	14	139	
All Wales[c]	2 835 073	6 110	5 870	2 282	14	139	8 357
Scotland	5 134 105	3 008	13 452	8 614	687	267	
Strathclyde[d]	2 283 671	943	7 989	4 331	97	127	18 000
Rest of Scotland	2 850 434	2 065	5 463	4 283	590	140	
United Kingdom	55 024 382	92 651	111 417	62 746	1 489	2 777	

Key:

A   Convictions for drug offences, 1996
B   Seizures of controlled drugs, 1996
C   People receiving treatment for drug abuse in England and Wales (regional drug abuse databases, October 1996-March 1997) and in Scotland (April 1995-March 1996)
D   Cases of HIV related to injecting drug use in England and Wales (1996) and in Scotland (1995)
E   Drug-related deaths in the United Kingdom (1995)

[a]Population survey: those having ever injected drugs (1991).
[b]General practitioners' computer records: extrapolation from diagnoses of drug abuse/dependence, 1996.
[c]Capture-recapture study of "serious drug users" (1994).
[d]Projection from computerized drug prevalence estimation program.

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. A drug abuse indicator for each geographical area is derived using principal component analysis 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 per cent 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 the analysis) is more strongly associated with convictions and HIV.

Step 6. Finally, regression analysis is used 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 the drug abuse prevalence rate in each of the seven target areas. The prevalence rate 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 table 12.

Table 12. United Kingdom data and multiple indicator prevalence estimates, 1996

Region	Population	Indicator index	Calculated observed drug rate (percentage)	Extrapolated drug rate (percentage)	Extrapolated number of drug users
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 Yorkshire	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)

Note: Figures within parentheses are totals.

[a]Population survey: those having ever injected drugs (1991).

[b]General practitioners' computer records: extrapolation from diagnoses of drug abuse/dependence, 1996.

[c]Capture-recapture study of "serious drug users" (1994).

[d]Projection from computerized drug prevalence estimation program.

Two technical points should be noted in the example. First, 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. That 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 principal components analysis to produce a single index of the remaining four, positively correlated indicators, and proceed as before. Table 12 would then be based on an indicator index derived from only four indicators.

The published paper, however, takes a more sophisticated route and identifies two summary indices from the set of five variables, using a principal components analysis that requires the extraction of two obliquely related components, one of which primarily covers convictions and the other treatment. Whether one or both of those indices is subsequently used in a regression analysis is a matter of judgement, 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 table 12. Again, in that data set, as chance would have it, there is no great difference in the resultant overall extrapolation, whichever of the alternatives is implemented. The more complex amendments to the basic procedure should, however, not be attempted without experienced statistical advice. That point cannot be too strongly stressed.

A second point is that a direct comparison can be made 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 those points. That 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 the case study the observed drug abuse prevalence rate value for Wales is not very well fitted at all by the model, and since that is one quarter of the anchor point data, it must give some cause for caution when interpreting the results. Examination of residuals 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 and may invalidate that assumption and the derived results. Those factors include the following:

(a)   The number of drug users in treatment may be restricted by the capacity of treatment services, or affected by the level of underreporting that can vary across the country;

(b)   The level of policing and attention given to drugs offences may vary across the country;

(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 (for example, police statistics) are reported by area of report.

Reliability and validity of estimates for the anchor points are of critical importance. While four anchor points were used for the United Kingdom study, they were obtained through various techniques that are subject to various assumptions that could not be evaluated. 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 1,000 population in the West Midlands to 75 per 1,000 population in Strathclyde (see table 12). In interpreting that 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 prevalence predictions of the multiple indicator model: 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

the United Kingdom. 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. Subnational estimates were made for regional health authorities because it was not possible (in the time frame of the study) to obtain data for smaller more meaningful populations within the United Kingdom. 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, for example, to separate inner London, and outer London from Thames Regional Health Authority, 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.

Source: EMCCDA project entitled "Study to obtain comparable national estimates of problem drug use prevalence in all European Union member States" (CT.97.EP.04); Frischer, M., and others, "A comparison of different methods for estimating the prevalence of problematic drug abuse in Great Britain ..." (2001); and Frank and others (1978).

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 those areas.

The published study uses all information transformed into rates per 100,000 of the population of each area (whether per 100,000 or per 1,000 or per 100 makes no difference to the analysis, provided it is a rate and not an absolute count that is used), and the analysis uses those rates directly. An alternative would be to use logarithmically transformed rates. Although, in the 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, that step ensures that estimates of problem drug use rates are greater than or equal to zero (that is, are technically plausible estimates).

Assumptions

In conclusion, it should be noted 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 those, if they are obtained without knowing the associated absolute counts. The following basic assumptions, however, are clear in all the potential variants:

(a)   Analysis should be based on prevalence rates of indicators and problem drug use;

(b)   It assumes that there is a stable relationship between prevalence and the indicator set, and that also implies that the indicators are defined and constructed in a similar manner for all data points;

(c)   The extrapolation is weaker in its validity and reliability if some data points do not conform to the presumed relationship;

(d)   The extrapolation is dependent on the form of the regression model fitted (Poisson, linear, log-transformed);

(e)   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);

(f)   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, that is, 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 it 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 that 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, however, the considerations are different. Study size is of little help in removing the distortion from the estimates produced by a method. For that reason, most methods in common use are developed because they produce unbiased 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 below in "Robustness of prevalence estimates". The use of formal statistical procedures for establishing confidence intervals around an estimate of prevalence are outlined below.

Confidence intervals from formal statistical theory

A confidence interval estimate simply gives a range of values within which it is likely that the true answer—the actual population prevalence—lies. That 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 confidence interval can be expressed in terms of absolute numbers of abusers (for example, 12,000 to 27,000) or as a factor applied to the estimate (for example, 18,000 within a factor of 1.5). Note that the validity of confidence interval calculations depends on the viability of the assumptions used in the analysis, and that it only reflects sampling variation, not bias. For that reason, given that indirect prevalence estimation methods are subject to many potential biases, confidence intervals are less important than using multi-methods or comparing any prevalence estimate with others.

Multiplier–benchmark confidence intervals

The calculation of a confidence interval for an estimate of prevalence derived by multiplier methods can be done easily if the estimation is considered as a Poisson (log-linear) regression procedure. Calculating the estimate through a computer software package will automatically give a confidence interval. Broadly speaking, when estimating the number of drug abusers in a country or region, the confidence interval (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 confidence interval is not often available and few multiplier-benchmark studies are able to quote one. That is not, however, a really serious drawback since there are more damaging forms of uncertainty in an estimate than sampling variation, as will be discussed in "Robustness of prevalence estimates", below.

Capture–recapture confidence intervals

For capture-recapture studies the confidence interval can be calculated most simply for models of any complexity by the computer software package used to fit the model. Bishop and others give equations for calculating the confidence interval 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 confidence interval are derived by trial and error based upon getting the nearest values for the "unobserved" population that give a difference of 3.84 (95 per cent) in the G^2, which is a measure of the fit of the model. See Hook and Regal (1995) for further discussion of that 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 confidence interval says nothing about the viability of the model, and therefore it 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 the present section, 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, that is seldom done in practice. General-purpose data-led methods such as bootstrap and jackknife estimates are plausible but not always satisfactory methods of assessing reliability. In general, those 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. Those failures tend to produce biased 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 and others (1985) used a mortality multiplier of 100—based on a presumed mortality rate of 1 per cent—and then repeated the calculation with a multiplier of 50—based on a presumed mortality rate of 2 per cent. In that 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 capture-recapture methods 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:

(a) The dangers of using only a two-capture capture-recapture design;

(b) 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 those 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 those 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 that the concordance of different methods probably gives the best indicator of a satisfactory estimate being derived, the following recommendations can be made:

(a) Use multi-methods:

   (i) Capture-recapture and multiplier-benchmark methods if possible;

   (ii) Multiple multiplier methods, ideally with the multipliers generated from more than one source (in case the sample used to derive the multiplier is biased);

(b) Use different models within the analysis of any one set of data to give a potential range of answers: in capture-recapture methods, comparison of different sub-optimal log-linear models;

(c) Look for plausibility and consistency of estimates of different behaviours or across different subpopulations:

  (i) 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);

  (ii) The model selected in capture-recapture estimation in terms of the proposed interactions between data sources should be credible.

Guidelines for producing research-based estimates

General guidelines

The data used in the indirect estimation methods described in the present manual are obtained by two means. The first is data collation from existing sources; the second is specialized 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 cooperation 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 research activities and other institutions, partly to fulfil the immediate research priorities and partly to develop networks of interested parties for research in the future. Using a research study in that way, taking the opportunity for 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 among otherwise isolated groups of workers. It is in that 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 checklist 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.

<div style="background:#8b2500;color:white;">

Checklist.   Minimum requirements in expertise and resources

</div>

1. Contact with the international scientific (epidemiology) community at large

2. Expertise in project management and planning

3. An advisory panel or group consisting of key scientists with previous experience

4. Statistical consultants available during the design stages

5. Questionnaire design skills

6. Fieldworkers with experience in drug research or similar work; training of fieldworkers

7. Computer access and computer skills in data entry and data file design

8. Statistical advice in later analysis

9. Web-based or other discussion groups with interested researchers

Many of the items on the checklist would be obvious to any researcher, and some 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 that source at the outset is vital. It is vital in helping to 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 unnecessary 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 generalized recommendations in a manual of this kind and discussion with people with prior experience cannot be underestimated in its importance. It is hoped that the numerous examples given in "Technical guidelines" 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 it 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 both in setting out the aims and

methods of the project and in assessing the progress of the project is very strongly recommended. That panel should be able to assist in making decisions to cope with difficulties that arise in 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 implementation began 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 is sought only when there is a set of data that has been collected and advice on some method of analysis is required. In fact, at the design stage of a project such 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 etc. 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, the burden of research falls upon the fieldworkers and the field organization. Good fieldwork is in the final analysis the cutting edge of data collection, be it collection from institutional records or through personal interviews. To that 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 that 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, cooperation should be of crucial importance 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 fieldworkers is a further aspect of the necessary training. Some degree of supervision and organization will be required in coordinating the fieldworkers. That 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 organization 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 those computer skills, then it is essential to have the cooperation of someone who does. A statistician can always advise on those matters as an addition to advice on statistical analysis. It is hoped that the present manual will give guidance on statistical analysis for those researchers with some expertise in such 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 indirect estimation methods.

Finally, when the analysis is under way and writing up the findings is the focus, it is important that it not take place in isolation. Statistical help is only one aspect of the benefits that come from teamwork at that stage: the advice of others—be it informally obtained through contact with peers or more formally obtained through 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 stigmatized 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 the life of a respondent. Different groups of people will need different and specialized assurances in many cases; for example, the issues relating to young people and to women are in some countries highly charged, requiring special attention.

Interview settings

The manner of contacting respondents in drug abuse inquiries is important: it should be discreet and as private as is possible. That consideration extends to the interview setting for questioning respondents: it is usually recommended that it be done in a private and isolated room or area. If that 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 special 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 that would apply, in particular, when help is sought from key socially prominent individuals. There can be no general guidelines that can apply across the many different social settings in different countries. The best the manual can do is to draw attention to the issue to alert the researcher to develop suitable approaches.

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 that 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, it should be restricted in its detail as far as is practical, and even that 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 records.

Technical guidelines

In the present 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 among those 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 in 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 they are kept as distinct as is practical and that they progress in a unified framework.

Target and reference population

The target population is, in the 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—among 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 also be considerations about nationality and residency. For example, are visiting or resident foreign nationals to be included or excluded from the prevalence figure? In some countries, that may make no tangible difference to the prevalence figure, in other places it might. When the studies are more locally based, 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 of age, for example, or with the entire age range? Does it cover both males and females in the population, both employed and unemployed? Most often the choice with respect to those 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. That will be part of the definition of the target behaviours (see "Key behavioural measures", below), along with the 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 subsets of interest. In the Introduction of this manual drug categories were given 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 those categories, be estimated within the same study. That is not always possible, particularly when some data sources relate to only opioid treatment, for example, 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. It should be noted, in that regard, that it is usually very difficult to operationalize the International Classification of Diseases or the Diagnostic and Statistical Manual of Mental Disorders criteria for those concepts in the field, or through a questionnaire administered by a fieldworker. Close—but always precise—definitions are often the best that can be achieved.

The Introduction lists the definitions for period of prevalence that might be used. Preferably, for compatibility with the annual reports questionnaire, a one-year period of 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 and others (1997); and Fischer and others, "A comparison of different methods for estimating the prevalence of problematic drug abuse in Great Britain ..." (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 (for example, Simeone and others (1997)), 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 that criterion threshold frequency and later rise above it again, effectively dropping out and back into drug using, that kind of detail is usually disregarded, or definitions are chosen in relation to the time period of the study that preclude that 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 those 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 those types of confounding issues are usually disregarded in analyses. If the study can use that type of information profitably, then expert statistical advice should be sought on how to do so (see Simeone and others (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 that 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. That 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 that status than would a heroin abuse treatment centre; and if data is collected on HIV tests among 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 those failures to match are discussed in several of the case studies.

Data possibilities and methods

Pre-assessment of situation

It is vital in the planning stages that effort is put into assessing the existing state of knowledge about drug use. Whether that 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 thinking in designing a study. Consideration should be given not

only to official statistics and publications when making that assessment, but the planners should also consider the use of key informants to help in the 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 those 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 those sources should be checked before any project design is finalized.

Since the availability of such data sources varies 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.

Checklist.   Possible data sources

1. Registers/records from treatment centres

2. Doctors and medical professionals generally

3. General hospitals

4. Psychiatric hospitals

5. Specialist drug user services

6. HIV or other health registers

7. Specialist addict registers

8. Deaths registers

9. Drug-related deaths registers

10. Police and judicial records

The EMCDDA guidelines to prevalence estimation (see Introduction, "Additional materials") should also be consulted on that 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 that they will use. The list is therefore a broad guide to some of the issues about the data source and its contents that may be relevant to a research study.

Checklist.   Relevant features of data and information bases
1. Who owns and maintains the database?
2. Who owns the data itself?
3. What type of drug(s) does the information describe?
4. Does it hold drug-injecting information?
5. What information on drug use is held?
6. Are the data based on persons or events/contacts?
7. Does it distinguish first-time contact from repeated contact?
8. Are individuals identifiable across events/contacts within the data?
9. Are individuals identifiable in terms of other data sources?
10. What is the geographical coverage?
11. What is the time-span covered?
12. In what physical format are they kept?
13. Are there potential obstacles to accessing the data?
14. Are there possibilities for future improvement in the data?
15. Are there networking possibilities arising from their use?

Again, the EMCDDA manual on guidelines to prevalence estimation (see Introduction, "Additional materials") 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—that is, from sources that are not routinely collected—for any of the procedures described in this manual, there may be a potential for that 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 bench-mark for the treatment multiplier estimation procedure that was used. It was a peripheral target in the study that the register could be kept updated and provide a resource in future for use either by further drug prevalence studies or for any other purpose. While that register was not an official data source, effort was put into leaving it in hands that would be able to maintain it in the future.

Issues in primary data collection methods

Role of piloting for data recording procedures and questionnaires

The process of data collection requires considerable organization, cooperation and planning. As part of that process, 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 record-ing 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. That is not often the case, but when it is, full advantage should be taken of that bonus and any transcribing to paper forms avoided. Usually though, when transcription or collection forms are required, there are general points about those 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. That will apply to both how the data are recorded and how they will be coded, as well as how the individuals and forms will be identified or numbered in the computer coding scheme. A second overarching design consideration relates to who it is that will fill in the forms—is it the researchers themselves or 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 estab-lished principles that aid good data collection. While 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 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 atti-tudes, 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.

Checklist.   Basic questionnaire design considerations

1. Ensuring that each question can be answered properly and sensibly by everyone

2. Making sure there is always room to answer "do not know the answer"

3. Allowing for people who refuse to answer some—or indeed all—of the questions

4. "Routing" the person filling in the questionnaire through sections that can be skipped in particular cases

5. Ensuring that no "blank-coded responses" occur—some mark or other should always be recorded for each question

6. Determining what categories of answer are of interest, and that they are exhaustive and mutually exclusive categories

7. Determining when multiple-response questions can be used

8. Deciding which questions can best be left "open-ended"

9. Ensuring that contingent questions (if "yes" then what ...?) are properly structured and routed

10. Ensuring that the relevant time frame or time window is specified

11. Ensuring that, where relevant, the appropriate geographical location is 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, those 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" 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 among those is site sampling, described immediately below, but a second frequently used procedure is nomination sampling. In essence that 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 that principle are known. The reader is referred to the publications by Intraval (Bieleman and de Bie (1992)) for detailed descriptions of procedures and safeguards in those 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 programmes and homeless shelters. That sort of sampling, that is, sampling that takes place physically at a site, rather than using a preconstructed sampling frame or list of drug users, is called "site" or "binomial" sampling (Bielemann and de Bie (1992); and 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. That consideration must be taken in relation to whether the aim is to sample people randomly or whether it is to sample events (attendances). For example, in site sampling, if the aim is to project the results for that sample to the entire population of hardcore drug users, then the sample event rates must be weighted by the inverse of the probability that the user was sampled (see, for example, Simeone and others (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. If 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 those institutions will be forth-

coming 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 their use. In practice, however, that 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 among 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 those types of cohort study (see, for example, Frischer (1998)) and to the EMCDDA publication on indirect estimation (see Introduction, "Additional materials").

Publication and reviewing by peers

Whether the proposed study is an internal report for a funding organization or for a government agency, or whether it is intended to provide a publication in the academic literature, it is vital that the report be submitted for review 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—that practice of critical reviewing is universal among serious and well-respected researchers, academic journals and funding bodies.

It is of paramount importance not only as a means of helping 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 recognized beyond the bounds of the immediate circle of colleagues and financial backers of the researchers.

When the primary aim is for the research to be published as a governmental report, it is often possible to take advantage of that 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 possibly because of required permissions, and 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 findings of the project, which enables those findings to have a wider circulation than they otherwise might. Permissions from the body that commissioned the report will need to be obtained at the point of publishing any follow-up paper, to ensure that there is no further administrative difficulty in accessing the findings.

A final report should have a detailed account of 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, namely, planning the research, collecting the relevant data, data analysis and writing up the research study, all carry equal weight and can each be as labour-intensive as the others. The drafting of both the data collection procedure and the results of the analyses is equally important. The easiest way to ensure accuracy in reporting is to draft the report on the research while it is in progress. Waiting until the end of the exercise before beginning to draft the report is likely to lead to a significant loss of detail and important reservations and potential trouble spots regarding the research. In addition, drafting the research procedures as the research proceeds not only ensures a more comprehensive report, but also lightens the workload in the final stages of the research project.

General resources for prevalence studies

Global Assessment Programme on Drug Abuse resources

GAP epidemiology web site—purpose and use

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

Useful web sites

Department of Mental Health and Substance Dependence, *Guide to Drug Abuse Epidemiology* (WHO/MSD/MSB 00.3) (Geneva, World Health Organization, 2000), available at www.who.int/substance_abuse/PDFfiles/EPI_GUIDE_A.pdf.

European Monitoring Centre for Drugs and Drug Addiction, *Estimating the Prevalence of Problem Drug Use in Europe*, Scientific Monograph No. 1 (Luxembourg, Office for Official Publications of the European Union, 1997). See www.emcdda.org/situation/themes/problem_drug_use.shtml.

General documentation

References cited in the text

Archibald, C. P., and others, "Estimating the size of hard-to-reach populations: a novel method using HIV testing data compared to other methods", *AIDS*, vol. 15 (supplement) (1992), S41-S48.

Bieleman, B., and de Bie, E., "Between the lines—a study of the nature and extent of cocaine use in Rotterdam" (Rotterdam, Intraval Foundation, 1992).

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W., "Estimating the size of a closed population", in *Discrete Multivariate Analysis: Theory and Practice* (Cambridge, Massachusetts, MIT Press, 1975), pp. 229-256.

European Monitoring Centre for Drugs and Drug Addiction, *Methodological Guidelines to Estimate the Prevalence of Problem Drug Use at Local Level* (Lisbon, 2000), available at www.emcdda.org/situation/themes/problem_drug_use. shtml.

European Monitoring Centre for Drugs and Drug Addiction, Scientific Monograph No. 1, *Estimating the Prevalence of Problem Drug Use in Europe* (Lisbon, 1997).

Fitch, C., and others, *Responding to Drug and Alcohol Use in Jersey: Key Findings Report* (Jersey, States of Jersey Drug Strategy Unit, 2001).

Frank, B., and others "Seeking truth in heroin indicators: the case of New York City", *Drug and Alcohol Dependence*, vol. 3 (1978), pp. 345-358.

Frischer, M., "Estimating the prevalence of drug abuse using the mortality multiplier method: an overview", in EMCDDA, *Estimating the Prevalence of Problem Drug Use ...*

Frischer, M., and others, "A comparison of different methods for estimating the prevalence of problematic drug abuse in Great Britain", *Addiction*, vol. 96, No. 10 (2001), pp. 1465-1476.

Frischer, M., and others, "Mortality among injecting drug users: a critical reappraisal", *Journal of Epidemiology and Community Health*, vol. 47, No. 1 (1993), pp. 59-63.

Frischer, M., and others, "Estimating the population prevalence of injection drug use and infection with human immunodeficiency virus among injection drug users in Glasgow, Scotland", *American Journal of Epidemiology*, vol. 138, No. 3 (1993), pp. 170-181.

Goodman, L., "Snowball sampling", *The Annals of Mathematical Statistics*, vol. 32 (1961), pp. 148-170.

Hall, W., and others, "How many opioid users are there in Australia?", *Medical Journal of Australia*, vol. 173 (2000), pp. 528-531.

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

Hickman, M., and others, "Estimating drug prevalence: review of methods with special reference to developing countries", *Bulletin on Narcotics* (2002).

Hook, E. B., and Regal, R. R., "Capture recapture methods in epidemiology: methods and limitations", *Epidemiologic Reviews*, vol. 17 (1995), pp. 243-264.

Hser, Y., and others, *Techniques for the Estimation of Illicit Drug User Prevalence: an Overview of Relevant Issues* (Washington, D.C., 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", *American Journal of Epidemiology*, vol. 142 (1995), pp. 1047-1057.

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

Mariani, F., Guaiana, R., and Di Fiands, T., "An epidemiological overview of the situation of illicit drug abuse in Italy", *The Journal of Drug Issues*, vol. 24 (1994), pp. 579-595.

Mastro, T. D., and others, "Estimating the number of HIV-infected injection drug users in Bangkok: a capture-recapture method", *American Journal of Public Health*, vol. 84 (1994), pp. 1094-1099.

McKetin, R., and others, *Drug Trends 1998. A Comparison of Drug Use and Trends in Three Australian States*, National Drug and Alcohol Research Centre Monograph No. 41 (Sydney, University of New South Wales, 1999).

Rehmann, Griffiths and Taylor, "Rapid assessment report on the Pakistan national assessment exercise" (UNODC, 2002).

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

SHAKTI Project, CARE Bangladesh (Dhaka).

Simeone, R., and others, "A plan for estimating the number of 'hardcore' drug users in the United States" (Washington, D.C., Drug Policy Research Group, Office of National Drug Control Policy, 1997).

Tilling, K., and Sterne, J.A.C., "Capture-recapture models including covariate effects", *American Journal of Epidemiology*, vol. 149, No. 4 (1999), pp. 392-400.

Wickens, T. D., "Quantitative methods for estimating the size of the drug using population", *Journal of Drug Issues*, vol. 23 (1993), pp. 185-216.

Woodward, J. A., Retka, R., and Nig, L., "Construct validity of heroin abuse estimators", *The International Journal of the Addictions*, vol. 19, No. 1 (1984), pp. 93-117.

Other useful references

Capture–recapture references

Hook, E. B., and Regal, R. R., "Effect of variation in probability of ascertainment by sources upon 'capture-recapture' estimates of prevalence", *American Journal of Epidemiology*, vol. 137, No. 10 (1992).

Doscher, M. L., and Woodward, J. L., "Estimating the size of subpopulations of heroin users: applications of log-linear models to capture-recapture sampling", *The International Journal of the Addictions*, vol. 18, No. 2 (1983), pp. 167-182.

Wittes, J. T., "Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources", *Journal of Chronic Diseases*, vol. 27 (1974), pp. 25-36.

Brownie, C., "Reader reaction—recent models for mark-recapture and mark-resighting data", *Biometrics*, vol. 43 (December 1987), pp. 1017-1022.

Seber, G.A.F., "A review of estimating animal abundance", *Biometrics*, vol. 42 (June 1986), pp. 267-292.

Wolter, K. M., "Capture-recapture estimation in the presence of a known sex ratio", *Biometrics*, vol. 46 (March 1990), pp. 157-162.

Cormack, R. M., "Log-linear models for capture-recapture", *Biometrics*, vol. 45 (June 1989), pp. 395-413.

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

Cowan, C. D., and Malec, D., "Capture-recapture models when both sources have clustered observations", *Journal of the American Statistical Association*, vol. 81, No. 394 (June 1986).

Chao, A., "Estimating the population size for capture-recapture data with unequal catchability", *Biometrics*, vol. 43 (December 1987), pp. 783-791.

Haber, M., "Testing for pairwise independence", *Biometrics*, vol. 42 (June 1986), pp. 429-435.

Community and snowball sampling references

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 (1987).

Biernacki, P., "Snowball sampling. Problems and techniques of chain referral sampling", *Sociological Methods and Research*, vol. 10, No. 2 (November 1981), pp. 141-163.

van Meter, K. M., "Methodological and design issues: techniques for assessing the representatives of snowball samples" (source not stated).

Bieleman, B., and de Bie, E., *Between the Lines—a Study of the Nature and Extent of Cocaine Use in Rotterdam* (Rotterdam, Intraval Foundation, 1992).

Goodman, L., "Snowball sampling" *The Annals of Mathematical Statistics*, vol. 32 (1961), pp. 148-170.

Griffiths, P., and others, "Reaching populations of drug users by the use of privileged access interviewers: methodological and practical issues", *Addiction*, vol. 88 (1993), pp. 617-1626.

Mortality and deaths multiplier references

Darke, S., Ross, J., and Hall, W., "Overdose among heroin users in Sydney, Australia: II. Responses to overdose", *Addiction*, vol. 91, No. 3 (1996), pp. 413-417.

Davoli, M., and others, "Risk factors for overdose mortality: a case-control study within a cohort of intravenous drug users", *International Journal of Epidemiology*, vol. 22, No. 2 (1993), pp. 273-277.

Farrell, M., and others, "Suicide and overdose among opiate addicts", *Addiction*, vol. 91, No. 3 (1996), pp. 321-323.

Frisher, M., and others, "Mortality among injecting drug users: a critical reappraisal", *Journal of Epidemiology and Community Health*, vol. 4 (1993), pp. 59-63.

Gossop, M., and others, "Frequency of non-fatal overdose", *British Medical Journal*, vol. 313 (1996).

Hammersley, R., Cassidy, M., and Oliver, J., "Drugs associated with drug-related deaths in Edinburgh and Glasgow, November 1990 to October 1992", *Addiction*, vol. 90, pp. 959-965.

Home Office, "Statistics of drug addicts notified to the Home Office of the United Kingdom of Great Britain and Northern Ireland, 1994", *Home Office Statistical Bulletin 17/95* (London, 1995).

Joe, G. W., Lehman, W., and Simpson, D., "Addict death rates during a four-year post-treatment follow-up", *American Journal of Public Health*, vol. 2, No. 7 (1982), pp. 703-709.

Ruttenber, A., and Luke, J., "Heroin-related deaths: new epidemiological insights", *Science*, vol. 226, pp. 14-20.

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

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

General references

EMCDDA, *Methodological Guidelines to Estimate the Prevalence of Problem Drug Use at Local Level* (Lisbon, 2000), available at: www.emcdda.org/situation/themes/problem_drug_use.shtml

EMCDDA, *Recommended Draft Technical Tools and Guidelines—Key Epidemiological Indicator: Prevalence of problem drug use* (Lisbon, 2000), available at: www.emcdda.org/situation/themes/problem_drug_use.shtml.

Hickman, M., and others, "Estimating drug prevalence: review of methods with special reference to developing countries. *Bulletin on Narcotics* ...

Hook, E. B., and Regal, R. R., "Capture-recapture methods in epidemiology: methods and limitations", *Epidemiologic Reviews*, vol. 17 (1995), pp. 243-264.

International Working Group for Disease Monitoring and Forecasting, "Capture-recapture and multiple record systems estimation I: history and theoretical development", *American Journal of Epidemiology*, vol. 142 (1995), pp. 1047-1057.

Bishop, Y.M.M., Fienberg, S. E., and Holland, P. W., "Estimating the size of a closed population", in *Discrete Multivariate Analysis: Theory and Practice* (Cambridge, Massachusetts, MIT Press, 1975), pp. 229-256.

Mastro, T. D., and others, "Estimating the number of HIV-infected injection drug users in Bangkok: a capture-recapture method", *American Journal of Public Health*, vol. 84 (1994), pp. 1094-1099.

Fitch., C., and others, *Responding to Drug and Alcohol Use in Jersey: Key Findings Report* (Jersey, States of Jersey Drug Strategy Unit, 2001).

SHAKTI Project, CARE Bangladesh (Dhaka).

McKeganey, N., and others, "Female streetworking prostitution and HIV infection in Glasgow", *British Medical Journal*, vol. 305 (October 1992), pp. 801-804.

Ghodse, H., "Casualty departments and the monitoring of drug dependence", *British Medical Journal*, vol. 1 (1977), pp. 1381-1382.

Frank, O., "Estimation of population totals", in *Perspectives on Social Network Research*, Holland, P. W., and Leinhardt, S. L., eds. (New York, Academic, 1979).

Picklands III, J., and Raghavachari, M., "Exact and asymptotic inference for the size of a population", *Biometrika*, vol. 74, No. 2 (1987), pp. 355-363.

Further references

Hickman, M., and others, "Surveillance of problem drug use in the UK: a review of a regional drug abuse database", *Journal of Public Health*, vol. 21 (1999), pp. 271-277.

Crabbe, T., Donmall, M. C., and Millar, T., "Validation of the University of Manchester drug abuse database", *Journal of Epidemiology and Community Health*, vol. 53, No. 3 (1999), pp. 159-164.

Prevalence of HIV in the United Kingdom: Report of the Unlinked Anonymous Prevalence Monitoring Programme in the United Kingdom (London, Department of Health, 1999).

Frischer, M., and others, "A comparison of trends in problematic drug abuse from two reporting systems", *Journal of Public Health Medicine*, vol. 22, No. 3 (2000), pp. 362-367.

Bloor, M., Wood, F., and Palmer, S., *Estimating the Prevalence of Injecting Drug Use and Serious Drug Use in Wales* (Cardiff, Social Research Unit, 1998).

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