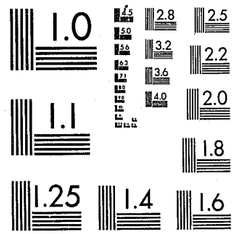
If you have issues viewing or accessing this file contact us at NCJRS.gov.

National Criminal Justice Reference Service

ncjrs

This microfiche was produced from documents received for inclusion in the NCJRS data base. Since NCJRS cannot exercise control over the physical condition of the documents submitted, the individual frame quality will vary. The resolution chart on this frame may be used to evaluate the document quality.



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Microfilming procedures used to create this fiche comply with the standards set forth in 41CFR 101-11.504.

Points of view or opinions stated in this document are those of the author(s) and do not represent the official position or policies of the U. S. Department of Justice.

National Institute of Justice
United States Department of Justice
Washington, D.C. 20531

DATE FILMED
9/11/81

U.S. Department of Justice
National Institute of Justice

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Justice.

Permission to reproduce this ~~copyrighted~~ material has been granted by

Public Domain
US Department of Justice

to the National Criminal Justice Reference Service (NCJRS).

Further reproduction outside of the NCJRS system requires permission of the ~~copyright~~ owner.

ESTIMATING THE SIZE OF DRUG USER POPULATIONS

Robert G. Demaree
Institute of Behavioral Research
Texas Christian University
Fort Worth, Texas 76129

Paper presented at the Special National Workshop on Research Methodology and Criminal Justice Program Evaluation, March 17-19, 1980, Baltimore, Maryland, sponsored by the Law Enforcement Assistance Administration, U.S. Department of Justice

NCJRS

MAY 6 1981

ACQUISITIONS

Support for the preparation of this paper was provided by Grant 2 R01 DA011765-03 and Contract 271-79-4710 with the National Institute on Drug Abuse. The points of view or opinions expressed do not necessarily represent official policy or positions of either the U.S. Department of Justice or the National Institute on Drug Abuse.

Abstract

Estimating the Size of Drug User Populations

Robert G. Demaree
Institute of Behavioral Research
Texas Christian University
Fort Worth, TX 76129

Following a definition of terms and discussion of some important aspects of the prevalence estimation task, two approaches -- sample surveys and records of known drug users -- are described. Data requirements, assumptions, and appraisals of these approaches are offered. Synthetic procedures and regression models for the indirect estimation of prevalence based on survey data, and log-linear models utilizing multiple samples of known cases are given particular attention. Examples are offered of prevalence estimates for drug abuse, as a means of illustrating different models and procedures.

ESTIMATING THE SIZE OF DRUG USER POPULATIONS[1]

Robert G. Demaree
Institute of Behavioral Research
Texas Christian University

In the late 1960's there was a groundswell of public concern and governmental response to drug use in this country. The history of the large scale efforts which were mounted to prevent or curtail drug use is well known. These efforts have continued over the years with little or no abatement, and have been accompanied by repeated requests for estimates of the size and nature of drug user populations. Such estimates are needed, and indeed are demanded, as a basis in part for policy and budgetary decisions, for gauging the social costs of drug abuse, and for evaluating the impact of law enforcement, prevention, and treatment programs in the drug field.

In the present paper methods currently employed (Richards & Blevens, 1977; Glenn & Hartwell, 1977) for estimating the size of drug user populations are discussed. The intent is to touch upon important, but sometimes neglected, features and limitations of these methods, rather than provide a comprehensive review. The paper begins with a definition of terms, such as prevalence, drug use, and population, and then touches upon some basic aspects of the prevalence estimation task. For estimating the size of drug user

[1]Appreciation is expressed to S. B. Sells for helpful comments and suggestions.

populations, the main data sources discussed are surveys based on questionnaires or interviews, and records of known users by law enforcement, medical, public health, and treatment agencies. Estimation procedures, assumptions, and problems associated with these data sources are treated only briefly. Log-linear models utilizing multiple samples of known users, and procedures for the indirect estimation of prevalence are given particular mention.

Background

Although it might seem that the definition of terms involving drug user populations would be relatively straightforward, such definitions are often quite fuzzy. For example, suppose that the user population includes all adults in a particular city who used heroin one or more times during 1979. The size of this population is what is meant by prevalence. Included in this figure is the incidence which refers to the number of adults who used heroin for the first time in 1979.

As just defined, the heroin user population would include those who had ceased using heroin by the end of the year by virtue of voluntary cessation, imprisonment, hospitalization, or death. Ostensibly, however, the population would not include a former heroin user who had turned to so-called heroin substitutes or had received methadone in a treatment program and did not use heroin during 1979. Thus, the use of heroin only once during 1979, and in however

slight a quantity, would qualify an individual as a member of the heroin user population. There is a question, however, as to whether the only adult heroin users to be counted are those who are residents of the city in which prevalence is being estimated.

The preceding question, while often disregarded, deserves careful consideration. In the example at hand of heroin users in a particular city, counts of users by various agencies (such as heroin-related arrests and admissions to treatment for heroin use) are unlikely to distinguish between residents and non-residents of the city. In a household survey, however, heroin users who are institutionalized (e.g., in a jail, hospital, or residential treatment facility) or have no regular place of residence are generally missed (Rittenhouse, 1977; Gould & Thompson, 1977). Thus, a household survey at best may provide an estimate of heroin use in the non-institutionalized household population, rather than the total adult population. As suggested by case studies of heroin addicts this is apt to be a serious source of bias.

In addition to the specification of the total population in which prevalence is to be estimated, it is important to be clear regarding what is meant by drug use, the drugs to be included, and the period of time covered. Depending on the purposes for which prevalence is being estimated and the sources of data available, the preceding specifications may differ in a host of ways.

For the purposes of one study it might be sufficient to estimate the use of legally obtained, prescription psychotherapeutic drugs, irrespective of whether the drugs were used only as prescribed. For another study, however, an estimate might be desired of the use of tranquilizers, or perhaps of a specific drug such as Valium (diazepam). Stipulation of the amount of drug usage is seldom possible because the dosage levels vary and are often unknown by the user. As a result, reliance is usually placed on reports of frequency of use; with respect to frequency of use, distinctions may be made between occasional or experimental use and regular use. Finally, the period of time covered may range, for example, over the past 7 days, 30 days, or year.

Beyond the matter of definitions, some basic aspects of prevalence estimation must also be considered. The first of these is that there is no single best way to estimate the prevalence of drug use. As discussed later, household surveys of self-reported drug use are a poor source of information about heroin use, but appear to be a good source of data concerning the use or misuse of prescription psychotherapeutic drugs. On the other hand, an individual's use of heroin or other opiates is more likely to become a matter of record by law enforcement agencies, hospital emergency rooms, drug abuse treatment programs, and medical examiner reports of overdose deaths than is generally the case for other drugs. That such a difference involves more than the

legality of the drugs is shown by marijuana which is rarely mentioned in reports of emergency room treatments or overdose deaths, and seldom requires treatment when it is the sole drug of abuse. In sum, the likelihood that drug use will become a matter of record, or conversely will be hidden from view, differs greatly from one drug to another. The message, it would seem, is that the estimation of prevalence for different drugs may call for different data sources.

To estimate prevalence, however, more is required than simple counts such as may be obtained from surveys or records of known users. The something else that is required, even though it may not be stated explicitly, is a statistical model containing assumptions about the observed data and the relationships between these data and prevalence. From a statistical point of view the model indicates how the data were generated, and explains how the prevalence estimates per se are produced. Either the data, the model, or both may cause the estimates to be wide of the mark. As a hypothetical example, suppose that data on admissions to drug abuse treatment programs in a city are to serve as a basis for estimating the prevalence of heroin use in the city. The model applied to these data may assume that all heroin users are equally likely to enter treatment during the time period under study. Now, let us imagine that the treatment programs attract clients only from a certain area of the city, but within that area all heroin users are

equally at risk of entering treatment. In this case the model would apply to one area of the city, but the prevalence citywide would be mis-estimated because of deficiencies in the data. Countless examples of this sort could be given in which the available data, irrespective of the model adopted, would not produce valid estimates.

One reason for distinguishing between data and models is that there is a temptation to pay more attention to details of models than to severe limitations or biases in the data. In this regard it is not suggested that a moratorium be placed on the testing of competing models, with their differing assumptions. Nor is it recommended that researchers who know that the data are contaminated, incomplete, or faulty in certain respects, but are unable to gauge the extent of the problems or do anything about them, give up in despair. The answer is twofold. First, there is a pressing need to refine the data we use for prevalence estimation purposes. And, second, we do need to test competing models for their fit to the data. Finally, whenever we can do so, it is important to use more than one data source and model to estimate prevalence, so that we may have a basis for comparison of the estimates and possibly more confidence if the results are essentially in agreement.

A further point is concerned with the admonition that confidence bands be placed upon estimates of prevalence in order to avoid misleading anyone as to the accuracy of the

estimates. The problem is that the calculation of the confidence range presumes that the data meet the assumptions of the model. Investigators who know that their data are faulty and do not conform to the model may be reluctant to report the conventional confidence range. Perhaps the best course of action to take in this situation is to let it be known that the prevalence estimate and confidence band are indicative of the order of magnitude, at best, and that there are shortcomings in the data.

Finally, it is important to recognize that prevalence estimates may be made at various levels of aggregation, starting with an individual and proceeding in turn to a neighborhood, city, metropolitan area, state, and national level. In an explanatory or conceptual sense the variables one would draw upon to account for variations in prevalence at these levels are apt to be quite different. The same is true for the prediction of prevalence. For actuarial prediction at the individual level, one is apt to think of age, sex, marital status, and socioeconomic level as being predictive of the likelihood of drug use. At the neighborhood level, one would be apt to employ sociodemographic characteristics as predictors. As one proceeds up the line the predictors become more global or diffuse, but at the same time the variations in prevalence are apt to differ. Thus, the character of the prevalence estimation task depends not only upon the drugs involved, but upon the populations.

Generally speaking, there are two sources of data for prevalence estimation. These are sample surveys and agency records of known cases. The following discussion touches briefly on surveys and devotes most attention to case records.

Sample Surveys

The methodology and sampling plans available for household surveys are generally quite impressive. Nevertheless, for a survey of drug use, particularly the use of illicit drugs, there are unfortunately a number of challenges and pitfalls. Perhaps the foremost of these are the refusal to participate and response bias.

Refusals to participate often occur at the time of initial contact with an adult of a selected household, where the interviewer attempts to explain the survey, enumerate the household members, and make a random selection of the adult with whom an interview is sought. Based on differing refusal rates by interviewers and associations between such rates and the sociodemographic characteristics of neighborhoods, it is evident that nonparticipation is by no means random. The scope of this problem is illustrated by a household survey of drug use which was conducted in 1977 in 45 census tracts in Dallas, Texas; the percentage of adults who declined to participate was computed by tract and ranged from 6% to over 35%. Further, even though more than one explanation could have been offered, the lowest refusal

rates occurred in tracts with populations predominantly black or Mexican-American.

The second problem mentioned is that of response bias, specifically under-reporting. In the Dallas survey, in order to encourage frank responses to what was believed to be a sensitive inquiry, a procedure was followed in which a drug use questionnaire was completed by the respondent without the field worker being privy to the responses made. The completed questionnaire was placed in a sealed envelope by the respondent and the envelope was then dropped into the slot of a locked box. In spite of these provisions for anonymity, only 21 of roughly 3400 adults reported any use of heroin during the past 12 months. This is not an atypical finding for household surveys in which inquiries are made about heroin use.

Case Records

Case records in the files of law enforcement, medical, and treatment agencies, have been a mainstay in the field of drug epidemiology; however, these records apply mainly to the use of addictive drugs. The institutional events or records relied upon most often are admissions to drug abuse treatment programs, hospital emergency room treatments, overdose deaths, serum hepatitis (an indicator, although fallible, of parenteral needle use), and drug-related arrests.

For purposes of discussion it is convenient to speak of the foregoing as capture or tagging mechanisms. In relation to heroin as an example, two assumptions are usually made about these mechanisms. The first is that during a given period of time all heroin users in the population are equally likely to be tagged by a given mechanism, such as admission to treatment. This assumption would be denied, for example, if the only heroin users at risk of being tagged were those who used heroin daily or nearly daily. The second asserts that the risk of being tagged may differ between mechanisms and time periods, but only if the differing risks apply uniformly to all heroin users in the population. A violation of this assumption would occur, for example, if one tagging mechanism operated only upon heroin users residing in one sector of a city, whereas another mechanism extended throughout the city.

Inasmuch as individuals, with some exceptions, may be tagged more than once during a given time period we can analyze the repeated taggings by a given mechanism or the common taggings by different mechanisms. This can also be extended to successive time periods. A number of mathematical models are potentially available for estimating the hidden or untagged population. The only ones discussed here, however, are the capture-recapture or tag-retag methods.

The method which has been used most often, with some variations, is the two-sample method. An example is Andima's

study (1973) in which he determined how many of the narcotic overdose deaths in a given year in New York City were accounted for by individuals appearing in the narcotics register. In this study Andima assumed that the probability that an addict would fall victim to an overdose was the same for those listed in the register and those not listed. Granting this assumption, he could estimate the number of heroin users who neither appeared in the register nor were victims of overdose death, and thus arrive at an estimate of the total number of heroin users in New York City.

With three or more capture-recapture samples, more powerful techniques become available in that one can test alternative models; more specifically, log-linear models (Bishop, Fienberg, & Holland, 1975). The simplest of these is based on the assumption that capture by any one mechansim in a given time period has no bearing on the probability of any other capture. This assumption involves the fewest parameters, and if it also yields estimates of prevalence with the smallest confidence band, one would proceed no further. What is often found, however, with the capture-recapture method is that the estimates are highly volatile. This is especially true for rare events, such as overdose deaths.

A modification of the estimation procedure that has sometimes been advised is to partition the estimates according to one or more demographic variables, such as sex or

age. The justification for this procedure is that the assumptions of the model may apply better to such subgroups. This requires, of course, that the data be adequate for such a subgroup analysis.

The main problems with the capture-recapture approach are with access to data and quality of the data. In this regard, access to individual records has become increasingly difficult as a result of stringencies on confidentiality. It is often not easy to persuade agencies to release data even when it is stipulated that the codes proposed will not enable individuals to be personally identified. As a consequence, matching across data sets can be a real problem.

Another problem is that there may be no easy way to purge the data file of unwanted cases. Consider, for example, persons admitted to treatment for drug abuse in successive time periods. Those who were in treatment continuously can be excluded, but this may not be possible for those who unbeknownst to the investigator were not available for treatment in one time period or the other due to such reasons as hospitalization or confinement.

Indirect Estimates

Another set of procedures relies upon prevalence estimates in a benchmark or calibration sample as a springboard for estimating prevalence beyond the sample. Included in these procedures is a method known as synthetic estimation. To gain an understanding of this approach, suppose that a

survey of drug use has been conducted in a national probability sample of adults and that each respondent is classified according to age, sex, marital status, and socioeconomic level. Suppose that age, marital status, and socioeconomic level have each been broken down into three groups. Along with sex, this would mean that in combination each respondent falls into one and only one of 54 groups.

In principle, at least, the household survey allows for prevalence to be estimated for each of the 54 groups. In a new population all that is needed is to estimate the proportion of the population falling into each of the 54 groups, and apply those estimates to previously obtained prevalence values. This approach is appealing, but it also is quite demanding and involves many assumptions. Finally, such an approach would not be workable in the case of a drug (e.g., heroin) whose use was rarely reported.

A related method involves linear regression equations. Suppose that heroin prevalence in 1979 has been estimated for a sample of cities or metropolitan areas. One or more prediction equations is then formed. The predictors may be indicators of heroin prevalence, such as heroin-related emergency treatments; indicators of heroin availability, such as heroin price and purity; or, sociodemographic and other characteristics of the cities. With such an equation one could proceed to estimate prevalence for cities not in the sample or could estimate prevalence for a later year.

The validity of the results yielded by this approach obviously depends upon how good the prevalence estimates are in the beginning.

The final approach included in this discussion does not preclude other estimation procedures; instead it emphasizes the importance of incorporating the prevalence of drug use within an explanatory framework or theoretical model. Whether one seeks to estimate prevalence at the neighborhood, city, state, or national level, it is urged that investigators be explicit as to underlying influences or factors which may account for variations in prevalence. When models are explicit, empirical efforts to obtain valid indicators of specified factors and to test the extent to which competing explanatory models fit the data offer promise of strong contributions to epidemiological understanding.

Examples of Prevalence Estimation

There have been so many studies of drug prevalence over the past decade that it would be impossible to review them in this paper. As an alternative, brief mention is made of three studies judged to be of methodological interest.

The first is a capture-recapture study of drug prevalence in five cities during 1975, 1976, and 1977 by Arthur Young & Co. This study revealed the demanding nature of the data collection and analysis and provided a number of insights into the sensitivity of estimates based on log-linear models.

The second is the Michigan Heroin Prevalence Study by Wenger and others (1979). This was a carefully planned effort involving capture-recapture studies, indicator equations, and ethnographically-oriented interviews. Again, the results were impressive in terms of both the data problems uncovered and the final estimates.

Finally, a study is currently in progress by J. Arthur Woodward at UCLA, which extends earlier developments by Person, Retka, & Woodward (1977) for estimating heroin prevalence, based on indicators. Woodward has incorporated capture-recapture estimates of heroin prevalence within an explanatory model that includes traditional indicators of heroin prevalence and heroin availability. Thus, he not only has produced prevalence estimates, but has placed them in the context of an explanatory model. To the extent that the model empirically fits the data, the prevalence estimates can be expected to have added meaning for agencies concerned with the reduction of heroin use.

References

Andima, H., et al. A prevalence estimation model of narcotics addiction in New York City. American Journal of Epidemiology, 1973, 98(1), 56-62.

Arthur Young & Company, Study to develop methods of estimating populations of urban drug abusers. Final Report. Prepared for: Office of Drug Abuse Policy, Washington, D.C., December 1, 1978. Contract No. DA7AC002.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. Discrete multivariate analysis: Theory and practice. Cambridge, Mass.: The MIT Press, 1975.

Glenn, W. A., & Hartwell, T. D. Review of methods of estimating number of narcotic addicts. Final Report. Prepared for National Institute on Drug Abuse under Contract No. 271-75-1016, Work Order No. 05. Research Triangle Institute, August 1975.

Gould, L. C., & Thompson, W. D. Biasing factors in the measurement of trends in heroin use. Addictive Diseases: An International Journal, 1977, 3(2), 151-176.

Person, P. H., Retka, R. L., & Woodward, J. A. A method for estimating heroin use prevalence. U.S. Department of Health, Education, and Welfare, Public Health Service. National Institute on Drug Abuse, Division of Resource Development, 1977.

Richards, L. G., & Blevens, L. B. (Eds.) The epidemiology of drug abuse: Current issues. NIDA Research Monograph 10, March 1977. Washington, D.C.: U.S. Government Printing Office, DHEW Publication No. (ADM) 77-432.

Rittenhouse, J. D. The epidemiology of heroin and other narcotics. NIDA Research Monograph 16, November, 1977. Washington, D.C.: U.S. Government Printing Office, DHEW Publication No. (ADM) 78-559.

Wenger, M. G., Newmeyer, J. A., N'Namdi, G. R., et al. Michigan heroin prevalence study. Conducted by New Detroit, Inc., Sponsored by the Michigan Office of Substance Abuse Services, August 27, 1979.

END