Document Title:	Estimating Drug Use Prevalence Among Arrestees Using ADAM Data: An Application of a Logistic Regression Synthetic Estimation Procedure
Author(s):	Mary-Lynn Brecht ; M. Douglas Anglin ; Tzu-Hui Lu
Document No.:	198829
Date Received:	January 2003
Award Number:	2000-IJ-CX-0017

198829

Estimating Drug Use Prevalence Among Arrestees Using ADAM Data: An Application of a Logistic Regression Synthetic Estimation Procedure

Mary-Lynn Brecht, M. Douglas Anglin, Tzu-Hui Lu

UCLA Integrated Substance Abuse Programs
1640 South Sepulveda Boulevard, Suite 200
Los Angeles, CA 90025
Office Number: (310) 445-0874
Fax Number: (310) 473-7885
www.uclaisap.com

Project funded by National Institute of Justice, 2000-IJ-CX-0017

January 6, 2003

FINAL REPORT Approve
Approved By: [signature]
Date: 1/22/03

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

The authors and project staff at UCLA Integrated Substance Abuse Programs (ISAP) conducted this study under National Institute of Justice grant number 2000-IJ-CX-0017. We wish to thank NIJ project officer C. Crossland for her oversight of this project; H. Shen, Y. Cao, and C. Peng for analytical support; ISAP administrative staff; D. Gerstein, P. Reuter, and anonymous reviewers for their helpful comments.

ABSTRACT

Society continues to suffer the immense costs and consequences associated with drug use and crime. Rates of drug use among arrestees in selected sites typically runs 10 times as high as for the general population. But rates alone cannot give policymakers a magnitude of drug use prevalence among arrestees upon which to base rational policy development. Lacking complete enumeration of the arrestee population, decisions must be based on estimates. This study estimates the prevalence of drug-using arrestees in the U. S. by using available ADAM data for calendar year 2000 as a calibration sample and projecting to the national level. Prevalence estimates are presented for any illicit drug use (of 10 tested by urinalysis) and specifically for cocaine, for gender by age group by offense category subgroups. Estimation has also been done for state and county level data (California and its largest and smallest counties, Los Angeles and Alpine), for any illicit drug use.

The study used a logistic regression synthetic estimation approach, in which prevalence rates in a calibration sample (ADAM) are used to estimate the equivalent rates in a target population (national, state, or county) where the prevalence rates are unknown. The approach has the advantages of being low cost, relatively simple to implement and understand, and using available data.

Using this method, estimated prevalence of U. S. arrestees with recent use of any illicit drug is 6.4 million or approximately 65% of the arrestee population for use of any of 10 illicit drugs. We see substantially higher estimated prevalence for males (4.8 million) than for females (1.5 million), approximately in the same ratio as are numbers of arrests for males and females. For cocaine, the overall U.S. estimate is 3.8 million with 2.7 million males and 1.0 million females. Estimates were also calculated for California at 780 thousand, approximately 61% of the arrestee population for use of any of the 10 illicit drugs.

Evaluation of the methodology supported the acceptability of most results, in terms of reasonability, replicability, and reliability. However, the method did not perform as well as expected for estimating opiate and methamphetamine prevalence, and results for some small subgroups were less reliable than desired. Assessment of potential bias suggested that estimates are likely conservative.

The study recommends continued application and refinement of this ADAM-based method for prevalence estimation.

EXECUTIVE SUMMARY

Purpose

This report presents an application of a logistic regression synthetic estimation approach using available ADAM data for estimating the prevalence of illicit drug-using arrestees. Prevalence estimates are presented for any illicit drug use (of 10 tested by urinalysis) and for cocaine use for the U.S. Estimation has also been done for state and county level data (California and its largest and smallest counties, Los Angeles and Alpine) for any illicit drug use. The report describes the methodology and evaluates results to allow appropriate interpretation, considering reasonability of results, replicability of observed data, reliability of results, and potential bias. Recommendations are made for data and method improvements. The methodology is relatively simple and uses available data, facilitating its application to support policy development.

Background

Studies of arrestees in selected locations have shown rates of illicit drug use and potential need for treatment that far exceed those of other populations. But obtaining accurate national prevalence figures for this high risk, high social cost population has been challenging. Large-scale population surveys such as the National Household Survey on Drug Abuse (NHDSA) often under represent arrestees, causing potential inaccuracy in prevalence projections for the arrestee population and for general prevalence into which arrestees are aggregated. However, data on arrestee drug use are available for selected U.S. county areas through the Arrestee Drug Abuse Monitoring (ADAM) Program funded by the National Institute of Justice (NIJ). These data have provided critical information on the potential magnitude of the drug use problem among arrestees, trends in use, and changing patterns of types of drugs used. But the lack of complete geographic representation by the available arrestee data has hampered the translation of reported drug use prevalence into nationally representative numbers, thus limiting the wider applicability of prevalence results. The current study uses ADAM data as a basis for mathematical estimation to national, state, and county populations of arrestees. The estimation method partially adjusts for ADAM site selection bias by including population characteristics in the estimation via logistic regression.

Several developmental studies on regression modeling synthetic estimation approaches have been conducted by the project team. Synthetic estimation refers to a family of methods in which prevalence rates in one (calibration) sample are used to estimate the equivalent rates in a target population in which the prevalence rates are unknown (e.g., for drug use or treatment need). This type of estimation has been widely used by drug abuse treatment planners and researchers. The approach has the advantage of being a low-cost, relatively straightforward application that utilizes readily available data on drug use and related indicators. The study team's development of the logistic regression approach to synthetic estimation began in 1994 with estimation using socioeconomic indicators to calculate levels of drug use prevalence among arrestees at the national level (Hser et al., 1998). Reliability and validity studies of the method were completed in 1999, and the application of the logistic regression estimation procedure was further expanded to the local (county) level (Anglin et al., 1999). The method was also used at the state and within-state regional levels to estimate need for drug treatment (Shen, 2002).

The method has been shown to be applicable to ADAM data, providing a cost-effective method of utilizing these available data along with readily available related geographic area population characteristics and numbers of arrestees to estimate drug use prevalence among arrestees at national, state, and local levels. The current study has continued developmental application of the methodology, re-estimating at the national and state levels with more current data, and also expanding the application to a large and small county. Accurate estimates of prevalence of drug use and need for treatment lay the foundation for policy formation and services planning, particularly important for the high risk, high cost population of arrestees that appear to have rates of substance use that far exceed those of other population groups.

Methodology

Four data sources for the year 2000 were used for the estimation: ADAM data provided the calibration sample, including an objective measure of drug use; the 2000 Census and related sources provided socioeconomic data for use in the estimation; California Monthly Arrest and Citation Register (MACR) provided arrest data from which to calculate an adjustment rate for translating numbers of arrests into numbers of arrestees; data from the FBI Uniform Crime Reports (UCR) provided national, state, and county numbers of arrests

An objective measure of drug use was used: results from urinalysis for 10 illicit drugs (cocaine, marijuana, methamphetamine, opiates, phencyclidine (PCP), methadone, benzodiazepines, methaqualone, propoxyphene, and barbiturates). Estimation was done for use of any illicit drug and for cocaine; estimation was also explored for opiates and methamphetamine. Because drug use prevalence often differs by basic demographics including gender and age and by type of criminal offense, estimation was done within each of the 24 combinations of these variables. Age was grouped into three categories: ages 18-24, 25-34, and 35 or older. Offenses were grouped into four categories: violent, drug-related, property, and other. Estimates were calculated at the national, state (California), and county (largest, Los Angeles, and smallest, Alpine, of California's counties).

The estimation procedure is a multi-step calibration and projection process. 1) In the calibration phase for the sample of arrestees for ADAM sites, rates of drug use are related using logistic regression to socioeconomic indicators for the ADAM sites. 2) These estimated relationships are then used to project drug use rates for national, state, or county levels using population socioeconomic characteristics for the relevant geographic areas. 3) An adjustment factor is calculated using MACR data for translating numbers of arrests into numbers of arrestees. 4) Prevalence estimates are calculated for national, state, and county levels using the projected rates of drug use and the numbers of arrests from UCR data adjusted with multiple capture rates. Estimates reflect the objective drug-use definition and the procedures, thus representing numbers of arrestees with recent drug use (as would be indicated by a positive urinalysis if tests were universal for all arrests) on at least one arrest occasion during the year 2000.

Results

Selected results are summarized below:

National (any illicit drug)

- Estimated prevalence of recent users of any illicit drug among arrestees in the U.S. for the year 2000 is 6,395,927 or approximately 65% of the UCR arrestee population for use of any of the 10 illicit drugs.
- We see substantially higher estimated prevalence for males (4,851,247) than for females (1,544,680), approximately in the same ratio as are numbers of arrests for males and females.

- In terms of age groups, the highest estimate is for the oldest age group (2,485,311), which is also the broadest category in terms of age range, with 2,106,503 in the 18-24 year group and 1,804,114 in the 25-34 year group.
- By offense category, the highest estimated prevalence is for the "other" offense category (2,880,259), followed by drug-related (1,985,164), property (864,593), and violent crime (665,910), in that order, which is in the same order as for numbers of arrestees.

National (cocaine)

- For cocaine, the overall U.S. estimate for year 2000 is 3,779,263 with 2,748,461 males and 1,030,802 females.
- Similar to patterns described above for any illicit drug, the highest cocaine prevalence by subgroup is for the oldest age group and for the "other" offense category.

California (any illicit drug)

- For California estimated prevalence for year 2000 for users of any (of 10) illicit drug among arrestees is 782,774, approximately 61% of the arrestee population.
- Prevalence patterns by gender and age groups are similar to those for the U.S. California, however, differs from the U.S. in that the largest prevalence by offense category is for drug-related offenses, which also constitute the largest proportion of arrestees.

County (any illicit drug)

- For Los Angeles County, prevalence was estimated at 198,580 (approximately 64% of the arrestee population).
- For Alpine County, prevalence was estimated at 69 (approximately 66% of the county's arrestee population).

Evaluation of the methodology supported the acceptability of most results, in terms of reasonability, replicability, and reliability. However, the method did not perform as well as expected for estimating opiate and methamphetamine prevalence, possibly caused by substantial numbers of areas with negligible penetration of these drugs. Results for some subgroups were less reliable than desired, suggesting caution in interpreting results for specific small subgroups of arrestees. Several sources of potential bias in the estimates have been assessed, with the conclusion that estimates are likely conservative (that is, slightly lower than actual prevalence).

Conclusions and Recommendations

Based on the use of relevant data and through examination of the estimation process and results, we recommend the continued use of this type of prevalence estimation method as a way to provide an information basis for policy development and resources planning and allocation. We also recommend continued refinement of the estimation process, both to data and method, including:

- Continued collection of ADAM data, with expansions of sites and sizes of selected subsamples (e.g. females), and continued use of ADAM for prevalence estimation
- Use of an event-based national arrest system (rather than having analysis rely on aggregate tables)
- Improvements in timely availability and accessibility of data sources.
- Continued exploration of aspects of the calibration model to improve reliability, including model formulation (e.g. generalized models that account for non-independence of observed individuals), specification (e.g. additional socioeconomic variables or supply indicators), and application to additional subgroups (e.g. ethnicity).

- Comprehensive simulation to assess and illustrate the sensitivity and robustness of results across several dimensions (e.g. calibration sample size, number and size of subgroups, number of geographic subcomponents for projection, variability of observed prevalence rates and rates of penetration) and to assess possible improvement in estimation compared to the cost of additional data requirements and/or application complexity.
- Automation and streamlining of the estimation procedure method to facilitate its broader use.

Estimation of prevalence of drug use among arrestees remains a necessity since large-scale population surveys are prohibitively expensive and unlikely to be implemented in the absence of a congressional mandate for this information. However the magnitude of drug use in this high consequence, high cost group is critical to informed policy development and resource planning and allocation, especially where diversion programs are impacting the local treatment system. Therefore, we urge continuing efforts to refine and apply a scientifically rigorous estimation approach that uses the valuable data collected through the national ADAM program.

INTRODUCTION

Studies of arrestees in selected locations have shown rates of illicit drug use and potential need for treatment that far exceed those of other populations. For example, in 2000, the National Household Survey on Drug Abuse estimated that 14 million Americans had used illicit drugs in the past month, about 6.3% of the population 12 years old or older (SAMHSA, 2001); in 2001, numbers had risen to 15.9 million recent illicit drug users, 7.1% of the population age 12 and older (SAMHSA, 2002a). By contrast, urine test data on arrestees showed that 65% had recent use of one or more of five illicit drugs in the year 2000 (cocaine, marijuana, methamphetamine, opiates, PCP) (Taylor et al., 2001). However, the arrestee population is commonly underrepresented in general surveys of substance use, and thus often underrepresented in prevalence studies and in state and national treatment needs assessment. [Note that to facilitate use of this report, a glossary of abbreviations and terms appears in *Appendix 1*.]

Data on arrestee drug use are available for selected U.S. county areas through the Arrestee Drug Abuse Monitoring (ADAM) Program funded by the National Institute of Justice (NIJ). These data have provided critical information on the potential magnitude of the drug use problem among arrestees, trends in use, and changing patterns of types of drugs used. But the lack of complete geographic representation by the available arrestee data has hampered the translation of reported drug use prevalence into numbers representative at the national or state level, thus limiting the wider applicability of prevalence results.

Accurate estimates of prevalence of drug use and need for treatment lay the foundation for policy formation and services planning, particularly important for the high risk, high cost population of arrestees that appear to have rates of substance use that far exceed those of other population groups. For example, in California, prior work by the authors has estimated that about 770,000 out of the 1.27 million arrestees in 1996 and 789,000 of 1.29 million arrestees in 1997 were recent drug users (Anglin et al., 1999; Shen et al., 2002). Other estimates suggest substantial prevalence of chronic cocaine users among arrestees (e.g. estimated at 2.5 million nationwide for 1997 by Abt Associates, 2001). However, despite the large numbers of drug-using arrestees, studies suggest that many have never been in treatment, and may not be in treatment even when acknowledging need for treatment (Anglin & Hser, 1992; National Institute of Justice [NIJ], 2001). Moreover, in-depth needs assessment studies of arrestees indicate that those abusing drugs need assistance in numerous areas, including health, mental health, and employment (Danila et al., 1997; Kerber, 1998; Kerber & Harris, 1998).

The current capacity of most drug treatment programs, however, is not adequate in terms of either capacity or range of services. For example, a statewide needs assessment in California using 1997 data from various sources showed that a possible gap of 200,000 unserved individuals existed between a realistic demand for publicly funded treatment and actual capacity for that year; a sizeable proportion were arrestees (Brecht et al., 2002). Planning for a possible increase in the gap between realistic demand and capacity has become a critical imperative in California where the treatment system must respond quickly to a major perturbation caused by the Substance Abuse and Crime Prevention Act of 2000, which mandates into treatment many non-violent substance-using offenders who would otherwise have been incarcerated. This scenario may play out in other states which enact similar reforms in coming years.

Unfortunately, despite research showing high rates of drug use in high-risk groups such as arrestees, "concrete" figures on drug use prevalence and drug treatment needs, expressed as actual numbers of drug users, are difficult to obtain for many reasons. Because drug use and related activities are illegal and invite social disapproval and legal consequences, obtaining direct measures of such behavior is not a simple undertaking. Arrestees appear to be underrepresented in the National Household Survey on Drug Abuse (NHSDA), and drug use may be underreported by many individuals (SAMHSA, 1999; Schmidt & Weisner, 2000). For example, if the rate for the Pacific area from the 1997 NHSDA indicating proportion of adults with an arrest in the past year is projected to the California population for that year, the resulting estimate is only 213,300 arrestees (see *Technical Note 1* in *Appendix 2*); yet the actual number of arrestees in California was almost 1.3 million (Brecht, et al., 2002). Therefore, various proxy measures and indirect estimation methods have been used to estimate drug use prevalence in many populations (see, e.g., Hser & Anglin (Eds), 1993).

This report summarizes a study in the continuing development and refinement of a prevalence estimation method that is relatively simple to apply and that uses available data including ADAM. The report presents an application of a logistic regression synthetic estimation approach, by describing the methodology, presenting the resulting estimates, and discussing issues related to the reliability and applicability of the approach. This application estimates the prevalence in the U. S. of arrestees with recent drug use, using available ADAM data for calendar year 2000 as a calibration sample and projecting to the national level. Estimation has also been done for state and county level data (California and its largest and smallest counties, Los Angeles and Alpine) to illustrate the robustness and sensitivity of the method. The method partially adjusts for ADAM site selection bias by including population characteristics in the estimation. The prevalence estimates should aid in assessing the scope of the drug problem among arrestees as a basis for developing intervention efforts in prevention and treatment and planning for provision of sufficient and appropriate treatment services nationally and regionally. Evaluation of the methodology should facilitate its appropriate use and understanding of strengths and constraints and provides a basis for recommendations for further methodological refinement.

This report includes a *Background* section, which describes the context for prevalence estimation, synthetic estimation approaches, and previous applications of the methodology used in this study. The *Methods* section describes the databases used for estimation, the measure of drug use, categories used to subdivide ADAM data and arrest data, and the estimation methodology in a step-by-step manner. The *Results* section presents relevant intermediate results and the prevalence estimates for the U.S., California, and Los Angeles and Alpine Counties. Assessment of the estimation procedure and estimates appears in the section *Evaluation of Estimation.* The *Discussion* section includes constraints of the method, suggestions for further refinement, and summary.

BACKGROUND

Prevalence Estimation

Complete enumeration of drug use prevalence is rarely feasible in large populations such as arrestees because of the prohibitive cost. Other direct methods such as random probability sampling are typically not applicable for assessing drug use prevalence because the total drug-using population is unknown and accurate lists are not available as a frame for sampling. In addition, legal and social reasons exist for nondisclosure of this type of information so underreporting is frequently substantial in self report surveys not corroborated by objective measures such as urine and hair testing (e.g. Harrison, 1997; Magura & Kang, 1996; Wish et al., 2000). Thus, estimates of the numbers of drug users among arrestees must realistically be based on less than completely representative samples where objective measures are available (such as ADAM data) and on mathematical estimation procedures.

A simple approach to the calculation of numbers of drug users is direct projection, by which we infer the value of a population statistic from a small survey sample. For example, ADAM data were collected in four California sites in 2000; the percentage of drug use for the four sites combined could be calculated from the sample data. This percentage could then be used to calculate the number of drug users in the statewide arrestee population by multiplying it by the number of total arrestees in the state. However, the California ADAM data were not a random sample of California arrestees and the ADAM sites themselves did not provide a representative sample of the state's arrestee population. Thus the validity of results of a direct projection estimation method in this context would be questionable.

Another method commonly employed to estimate prevalence involves synthetic estimation, an approach that uses prevalence rates derived from one sample (often a smaller sample) for sociodemographic subcategories (such as age group, gender, and other variables potentially related to drug use prevalence) and projects these rates to other similarly categorized (often larger) populations for which the prevalence rates are not known. The current study uses this approach, adapted to utilize the advantages of increased availability of arrestee survey data (i.e. a larger sample of ADAM sites than in previous applications) and recent enhancements to estimation methodology techniques.

Synthetic estimation refers to a family of methods in which prevalence rates in one (calibration) sample are used to estimate the equivalent rates in a target population in which the prevalence rates are unknown (e.g., for drug use or treatment need) (Rhodes, 1993). This type of estimation has been widely used by drug abuse treatment planners and researchers (Hser et al., 1992; Schmeidler, 1991; Simeone et al., 1991; Wilson & McAuliffe, 2000). The approach has the advantage of being a low-cost, relatively straightforward application that can utilize readily available data on drug use and related indicators (e.g., Household Survey on Drug Abuse, Arrestee Drug Abuse Monitoring data, and Census data). One approach to synthetic estimation, often termed the social indicator approach, applies regression analysis to the projection method. This adjustment allows differential weighting of drug use indicators depending on their empirical relationship to some index of prevalence hypothesized to be related to the underlying prevalence of drug use in the target area. Such indicators to be used in the regression adjustment may be environmental/socioeconomic characteristics of a community (e.g., population density, unemployment, poverty rates) thought to be indirectly related to levels of drug use.

Overall, a variety of approaches have been used to apply survey results from a relatively small sample of cities/counties in order to obtain estimates of drug use prevalence or need for drug treatment for larger populations and to ascertain the utility and validity of the derived estimates (cf.. Maxwell, 2000; Simeone et al., 1995a,b). However, these strategies have proved only partially successful in obtaining acceptable results for supporting policy decisions. Difficulties have included, for example, bias resulting from use of self-report data, inadequate representativeness of the calibration sample, variability in drug use patterns across geographic areas, and reliability and validity of social indicators. The current study uses a specialized synthetic estimation approach in which ADAM sites are used as the calibration sample, for which an objective measure of drug use (urine test results) is related to population characteristics using logistic regression; these results are then applied to a population of arrestees (from FBI Uniform Crime Reports data) to produce prevalence estimates. This method was developed and previously applied by the UCLA team (Hser et al., 1998; Anglin et al., 1999; Shen et al., 2002) and has been updated and used in the current study.

Developmental Applications of the Method

Several developmental studies on regression modeling synthetic estimation approaches have been conducted by the UCLA Drug Abuse Research Center (DARC, now called the Integrated Substance Abuse Programs). This development began in 1994 with estimation using socioeconomic indicators to calculate levels of drug use prevalence among arrestees at the national level (Hser et al., 1998). Reliability and validity studies of the method were completed in 1999, and the application of the logistic regression estimation procedure was further expanded to the local (county) level (Anglin et al., 1999).

The approach was initially developed to estimate the number of drug users among arrestees in the U.S. based on information collected from 23 U.S. cities surveyed in the national Drug Use Forecasting (DUF). In this first application, logistic regression models were estimated relating each of five drug use variables, (cocaine, opiates, amphetamines, drug injection, and any drug) to social indicator variables from census data (total population, poverty, unemployment, high school graduates, and youth population). Using these empirical relationships, the method allowed projection from the few DUF surveyed cities to a large number of non-DUF (non-surveyed but relatively similar) cities with arrest data from the FBI Uniform Crime Reports (UCR). The method performed well, and resulting national estimates proved to be satisfactory. One inherent limitation to this approach was that the national UCR contains the aggregate number of arrests within the various crime categories; however, some of the arrests are accounted for by repeated arrests of the same individuals over the course of the reporting year.

For the UCR data, exact rates of re-arrest are not available. In order to convert arrests to arrestees, the initial UCLA study applied a multiple capture rate of 0.13 as a correction parameter, a value that had been used in other studies (e.g. Rhodes, 1993).

In 1999, the UCLA-DARC researchers completed a validity and reliability study using the previously developed logistic regression synthetic estimation method for smaller regions (Anglin et al., 1999). The study demonstrated that improved logistic regression analysis models could be fairly robust to the size of the targeted population at the state, regional, and large county levels. Prevalence estimates were calculated for 1996 for the U.S. and for California, Texas, and Los Angeles County. The estimation was conducted for subgroups of arrestees defined by gender, age group, and type of offense; race/ethnicity was included for California. The study included assessment of multiple capture (rearrest) rates in order to improve the accuracy of the drug use prevalence estimation among the arrestee populations studied. Findings revealed that in 1996 about 1.27 million adult Californians were arrested or cited 1.64 million times. The observed multiple capture rate was approximately .22, nearly double that used in prior research. Among the 1.27 million arrestees, about 770, 000 arrestees statewide were estimated as likely to be urinalysis positive for at least one of ten illicit drugs examined (at least once during the year, if urine testing were applied to all arrests).

The method was also used to estimate need for drug use treatment for arrestees in California and by region of the state for 1997. The calibration data came from a one-time study (CAL-DUF) in 13 California counties extending the DUF data collection to additional counties and expanding the survey protocol. Estimation based on urine test results in the DUF calibration sample showed that approximately 789,000 or 61% of the 1.29 million adult arrestees could be considered in need of treatment. Geographical regions were differentially represented in the need for treatment estimates compared to their population numbers and composition.

Summary

The methodology used in the current study has been applied successfully in several contexts. It has been shown to be applicable to ADAM data, providing a cost-effective method of utilizing these available data along with readily available data on related geographic area population characteristics and numbers of arrestees to estimate drug use prevalence among arrestees at national, state, and local levels. However, the early studies of national drug use prevalence among arrestees relied on a smaller ADAM sample (n=23); the larger sample available for the estimation for the year 2000 should improve nationwide representation. The current study applies this methodology to estimate illicit drug use prevalence among arrestees at the national and state levels and also expands the application to a large and small county.

METHOD

Data Sources

Data for developing the mathematical models and estimating prevalence come from several sources primarily for the year 2000: ADAM, the Census, the California Monthly Arrest and Citation Register data from the State Department of Justice, and Uniform Crime Reports. These are described in more detail below.

Arrestee Drug Abuse Monitoring (ADAM) Data

The National Institute of Justice ADAM data provided the basis for the calibration phase of the synthetic estimation. ADAM collects data on drug use in selected county-level sites across the U. S. This system began as the Drug Use Forecasting program in 1987 in 23 sites; in 1996 it expanded into ADAM with additional sites and updated methodology. In 2002 there are a total of 40 participating data collection and affiliate sites. Recent improvements have included a redesigned sampling strategy and new interview questions (Hunt & Rhodes, 2001a,b; Taylor et al., 2001). The probability sampling implemented with the 2000 survey allows improved generalization of ADAM site drug abuse rates to the county populations of arrestees.

Data on drug use come from interviews and urinalyses conducted at booking facilities in the participating counties during quarterly data collection periods. The participating ADAM sites for which any data were available for 2000 (n=36), including partial year participation, constitute the calibration sites for the prevalence estimation; the calibration aspect of the estimation process included data for n=24,387 *booked* arrestees who provided a urine specimen. Note that for simplicity this report uses the general term *arrestees*; when used in conjunction with ADAM data, it refers specifically to booked arrestees. Additional reference to this distinction and its implication appears in the *Evaluation of Estimation* section.

Unweighted data were used in this estimation for several reasons. This study chose to maintain consistency of procedures across subgroups; weights were available for men but not for women. The current study estimates calibration models within gender by age by offense categories, which may partially overcome ADAM sample representation bias. Preliminary analysis showed that for males, weighted and unweighted site-specific urine positive rates for any illicit drug were highly correlated at .97. The overall male urine positive rate for use of any illicit drug was .66 for either weighted or unweighted data, and rates were very similar for the subgroups to be used in the estimation (discrepant by at most 1.1%). The evaluation phase of the study included additional comparison of prevalence estimates with weighted vs. unweighted data; further discussion and detail appears in the section on *Evaluation of Estimation.*

Population Characteristic Data

The demographic and socioeconomic information necessary to compute regression estimates for the synthetic estimation model was abstracted from the 2000 Census data and Bureau of Labor Statistics data for 2000. Further detail is given in *Appendix 3*. Indicators included overall population size and percentages below poverty level, unemployed, population between 18 and 24, and high school graduates. In the calibration phase of the estimation procedure, these data were required for each of the counties (or relevant geographical unit) for which ADAM data were available. For the projection phase of the estimation procedure, census data were required for the geographical subcomponents of the level to which projections were being made; that is, for the national estimate, census data were required for each state, and for the California estimate, census data were required for each county. See the section *Socioeconomic Indicators* below for additional description of specific variables.

The California Monthly Arrest and Citation Register (MACR) Data

The arrest record database (MACR) was used to generate a multiple capture rate necessary in the current study to convert arrest data into the numbers of arrestees. This database is maintained by the California Department of Justice based on monthly reports by law enforcement agencies in California. The data are available as annual files. This database is event-based, with each record representing a single arrest for one individual; thus one individual could have multiple records if arrested more than once in a given year. The database carries personal identifiers that allow the calculation of multiple capture rates. Arrest data for 2000 were used; the file contained 1,460,705 adult arrest records.

FBI Uniform Crime Reporting (UCR) Data

Numbers of arrests were from the FBI Uniform Crime Reporting system, a national data system of arrest data. The 2000 UCR data report consists of reported arrests for the UCR index crimes (murder and non-negligent manslaughter, rape, robbery, aggravated assault, burglary, larceny, auto theft, and arson) and arrests for additional crimes such as forgery, fraud, vice offenses, and drug possession or sales. Approximately 17,000 city, county, and state agencies contribute data. In 2000, these contributing agencies represented 94% of the total U.S. population (FBI, 2001). The current study used unimputed data totaling 12,627,856 arrests nationally; data for 49 states came from the national UCR database, while data for Florida came from that state's UCR report. Data for Washington D.C. were not available; thus the estimates represent 50 states. The database available to the current project did not contain indicators for imputation to account for non-contributing agencies.

The UCR data are available in aggregate form by demographic groupings, offense type, selected characteristics of law enforcement agencies, and geographic location. An individual could be counted multiple times if arrested multiple times in the same year. Since UCR data do not carry personal identifiers that can be used to distinguish arrestees from arrests, the multiple capture rates generated from the California MACR were used as the best available empirical results to convert the UCR arrest statistics into numbers of unique arrestees.

Measures

Drug Use

Urine testing for illicit drug use offers an objective assessment of current or recent drug use. It has been frequently used as a calibrator to test the validity of other measures of drug abuse, such as self-report (e.g. Anglin, et al., 1993; Hser et al., 1999; Mieczkowski et al., 1998). ADAM urinalysis tests for 10 drugs: cocaine, marijuana, methamphetamine, opiates, phencyclidine (PCP), methadone, benzodiazepines, methaqualone, propoxyphene, and barbiturates. The first five are the "NIDA-5" panel of commonly used illegal drugs designated by the National Institute on Drug Abuse (NIDA). A positive urine test result indicates that the arrestee has used at least one illicit drug within approximately two to three days prior to arrest. More specifically, cocaine, PCP, methamphetamine, and opiates can be detected in the urine for 2-3 days after ingestion; marijuana remains in the body for up to 30 days after use.

Urinalysis completion rates in ADAM are high, for example above 85% in most sites reporting in 2000 (Taylor et al., 2001). Urinalysis provides higher rates of substance use than does self-report in the ADAM sample: among those who test positive for a specific drug, self-reported drug use typically ranges about 40-70%, differing by site and by drug. Typically congruence is somewhat higher for marijuana than for cocaine, opiates, or methamphetamine (Taylor et al., 2001). Discrepancies between urinalysis and self-report are also reviewed in Harrison (1997) and Magura & Kang (1996) and for ADAM in Rhodes (2002). While there is variability in consistency of self-report and bio-assays such as urinalysis that may vary by sample characteristics and specific drug, many studies show considerable underreporting of drug use by high-risk populations and especially for cocaine use (e.g. Hser et al., 1999; Mieczkowski et al., 1998; Morral et al., 2000; Rhodes & Kling, 2002; Wish et al., 1997).

The measure of drug use partially delimits the prevalence estimate. Thus in our application, the estimates of drug use prevalence refer to "recent" users of illicit drugs, as would be determined by a urine test. This is one of many possible measures of drug use and was selected for this study because of its objective nature, which precludes the necessity of using various adjustments for truthfulness in self-report (e.g. Abt Associates, 2001; Rhodes & Kling, 2002). Additional discussion of this issue appears in the *Evaluation of Estimation section* below.

Stratification Variables

Drug use prevalence often differs by basic demographics including gender and age and by type of criminal offense. To increase the accuracy of estimation, the calibration data and arrest data were stratified by these three variables in order to represent potential differences that might also vary across geographic units. Age was grouped into three categories: ages 18-24, 25-34, and 35 or older. These age categories are approximately similar to those used for ADAM reporting (except for the category designations for age 25 and for age 35). The slight discrepancy from ADAM categories is because of the limitations in available categories in the already aggregated national UCR arrest data. Offenses were categorized into violent, drug-related, property, and other offenses. A list of offense coding for both the MACR and UCR data appear in *Appendix 4*. Ethnicity was not included in this analysis because major subcategories of Hispanic and non-Hispanic white could not be distinguished within available UCR data. Moreover, a further subcategorization (across four variables instead of three) frequently produces cell sizes too small within the ADAM calibration sample for accurate estimation. As the ADAM sample expands in future years, additional stratification variables may be able to be included for prevalence estimation if these variables are also available in the UCR data.

Socioeconomic Indicators

Five socioeconomic indicators, commonly linked to levels of drug use and used in previous applications of our methodology, were included as predictors in the logistic regression analysis that related these indicators to drug use across the ADAM calibration sample. The five county-level socioeconomic indicators were overall population size, poverty (percentage below poverty level), unemployment (percentage unemployed), education (percentage with high school diploma or higher), and youth (percentage population 18-24). These indicators represent some of the many environmental/socioeconomic characteristics of a community that may be indirectly related to levels of drug use, and several studies have shown these indicators to contribute to stable and reasonable estimates of drug use prevalence in synthetic estimation approaches (e.g. Hser, 1993; Hser et al., 1998; Levy, 1979; McAuliffe et al., 2000; Person et al., 1977, 1978; Rhodes, 1993). The current study continued the use of these five indicators from the calibration development in earlier studies by the project team (Hser et al., 1998; Anglin et al., 1999; Shen et al., 2002). A major advantage of these indicators is their ready availability for national, state, and county levels. *Appendix 3* lists values of the five socioeconomic indicators for each of the calibration counties (ADAM sites) that were used in the prevalence estimation.

Prevalence for Specific Drugs

Overall illicit substance use was estimated based on urinanalysis for any of 10 substances: cocaine, marijuana, methamphetamine, opiates, phencyclidine (PCP), methadone, benzodiazepines, methaqualone, propoxyphene, and barbiturates. For this study, estimation was also done for each of three primary illicit substances: cocaine, opiates, and methamphetamine. These substances were selected for illustrative purposes because of their prevalence, increasing use, and/or deleterious effects.

For ADAM sites in 2000, urinalysis results showed that a median rate of 30% of arrestees tested positive for cocaine (either crack or powder); site-specific rates ranged from 8-49%, with five site rates above 40%. Cocaine also was the most frequently occurring drug among deaths reported by medical examiners participating in the Drug Abuse Warning Network (DAWN) (Ducharme & Ball, 2001).

While generally less prevalent than cocaine, opiate use and problems remain at relatively high levels, and heroin is often perceived as the drug associated with the most serious consequences (ONDCP, 2002). For example, in most Community Epidemiology Work Group (CEWG) areas, rates of Emergency Department heroin/morphine mentions reached or exceeded their highest levels in more than 10 years in 2000 (CEWG, 2001). Moreover, heroin use is increasing among younger people, and among females in some areas; and use of synthetic opioids has increased (ONDCP, 2002).

Methamphetamine use is particularly high in several western states, and has also been increasing its geographic penetration within the U.S. Dramatic increases were seen in methamphetamine use indicators in 1992-1997; a slight reduction in this upward trend occurred in 1998-99, but many indicators again showed increases in 2000 (CEWG, 2001).

The scope of this study did not allow estimation for additional specific drugs. While marijuana is a commonly used drug, we did not include it within this study for several reasons. Its use maintains an ambiguous position; for example, marijuana use for medical purposes is legal in California, and prevalence estimation based on urine test results has no way to differentiate illegal use from legal. In addition, law enforcement agencies often treat marijuana possession differently from other hard-core drugs. Finally, marijuana use is not commonly perceived by the public as a drug related to serious consequences. However, prevalence estimates for marijuana, as well as for other specific drugs not covered in this study, may be needed for certain policy development; while such estimation was outside the scope of the current study, future studies should include estimation of additional drugs.

National, State, Local Area Estimates

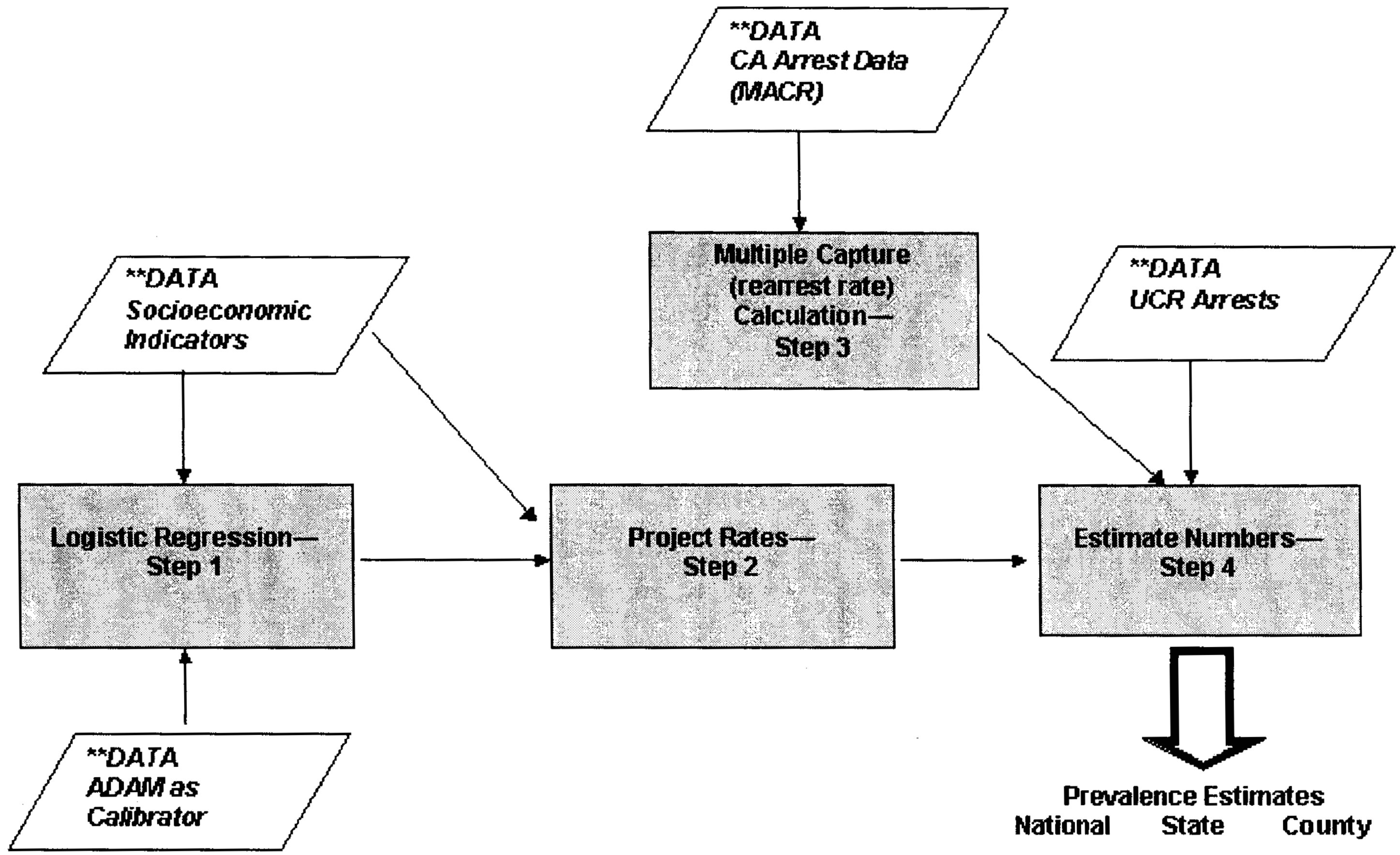
Estimation was done at the national, state, and local level in order to assess the sensitivity of the model to gross differences in numbers of arrests and arrestees. California was selected as the illustrative state, and the local level was represented by the largest and smallest California counties (Los Angeles and Alpine, respectively). Total numbers of UCR arrests for adults for these four geographic levels were 12,627,856 (U.S), 1,640,480 (California), 395,984 (Los Angeles County), and 133 (Alpine County). At the national level, estimates were calculated for use of any illicit drug and separately for cocaine, opiates, and methamphetamine. Note, however, that the national estimates are presented in this report only for any illicit drug use and cocaine, since estimates for opiates and methamphetamine were clearly biased for reasons discussed later in this report. At the state and county level, estimates were calculated (and are presented) only for use of any illicit drug.

Within national and state levels of estimation, prevalence was computed for geographic subcomponents and then aggregated to the appropriate level. For national prevalence, estimates for each gender by age group by offense category were projected for each state and summed across 50 states. For California prevalence, estimates were computed and summed across 58 counties. The smaller area estimation allows a greater sensitivity of the models to socioeconomic diversity across the geographic subcomponents. The grosser level of estimation masks such variability, and prior studies by the project team showed the grosser estimates to be less accurate.

Estimation Procedures

The estimation procedure involved four phases: 1) estimate logistic regression models relating drug use to population demographic characteristics for the ADAM calibration sample; 2) project drug use rates for county, state, and national prevalence using regression coefficients from Step 1 and population demographic characteristics; 3) calculate multiple capture (or rearrest) rates for translating numbers of arrests into numbers of arrestees; 4) calculate numbers of drug-using arrestees at national, state, and county levels using rates from Step 2 and UCR data adjusted with multiple capture rates from Step 3. Figure 1 shows a schematic diagram of the procedures, and procedures are described briefly below. In the description of the procedures below, we do not include extensive justification of, or issues in, the development of these procedures, since these have been covered in previous reports of development (e.g. Hser et al., 1998; Anglin et al., 1999). However, in this report we have considered many issues related to the appropriateness, accuracy, and utility of the methodology as used in the current application; these are summarized in the *Evaluation of Estimation* Section.

Figure 1: Schematic for Logistic Regression Synthetic Estimation of Drug Prevalence



Step 1: Estimate logistic regression models for ADAM calibration sample

For each gender by age group by offense category subgroup (totaling 24), a logistic regression model was estimated relating the probability of drug use as indicated by ADAM urinalysis results to local population demographic characteristics (see *Technical Note* 2). Further discussion of model validity appears in the *Evaluation of Estimation* on calibration. SAS PROC LOGISTIC was used for the logistic regression, with drug use data for each ADAM site input in terms of counts (the number of positive urine test outcomes and number of urine tests for the site, representing the rate). Predictors were five social indicators from census data for specific site areas: total population, percentage unemployed, percentage living under the poverty level, percentage of high school graduates, and percentage of youth 18-24 years old. This model estimation process was repeated to predict use of any illicit drug (from among 10: cocaine/crack, opiates, marijuana, methamphetamine, PCP, methadone, diazepam, methaqualone, propoxyphene, barbiturates) and then separately for use of each of three drugs (cocaine/crack, opiates, and methamphetamine). Thus, a total of 96 logistic models were estimates (24 subgroups times four types of drugs).

Step 2: Calculate predicted drug use rates for geographic areas beyond ADAM sites

For each gender by age group by offense category subgroup, predicted drug use rates were calculated for relevant geographic subcomponents within the national and state levels of estimation. The coefficients for each subgroup (e.g. male, 18-24 years, offense category 1) from Step 1 above were used in a logit model along with values of socioeconomic indicators (total population, unemployment, poverty level, education, youth population) of the geographical subcomponents to give predicted rates of drug use. In the current study, individual states were used as subcomponents for the national estimates; thus, a projection was done for each state using the relevant subgroup coefficients and the state's socioeconomic data. Counties within California were used as the geographic subcomponents for estimation for that state. This procedure was replicated for each of the gender by age group by offense category subgroups and for each type of drug considered. The current study replicated the procedure for use of any illicit drug, use of cocaine, opiates, or methamphetamine for national estimates. But the study projected the rates only for any illicit drug use for California and for Los Angeles and Alpine counties.

Step 3: Calculate multiple capture (rearrest) rates

Prevalence, as reported by this project, was intended to reflect number of arrestees in a given year with recent drug use (that is, who would test positive for illicit drugs on at least one arrest occasion during the year if urine tests were imposed for all arrests). However, available data sources use "arrest" in the specified year as the unit of measurement rather than "arrestee." Thus, an individual arrestee may have multiple records if arrested more than once if the specified year. The study calculated a multiple capture (or rearrest) rate for conversion of the number of arrests into number of arrestees. If this adjustment is ignored, then serious overestimation bias may occur in prevalence estimates meant to represent unique individuals rather than arrest occasions. This multiple capture rate is calculated for a specified time period as

(number of arrests – number of arrestees) / (number of arrests)

Early prevalence estimation studies used a general multiple capture rate of .13 (or 13%) (Hser et al., 1992, 1998; Rhodes, 1993). That is, .13 (or 13%) of the total arrests represented the second, third, or subsequent arrests for individuals. But more recent studies have shown substantially higher rates of multiple captures, and differential rates for various subgroups of arrestees. For example in 1997, 1.29 million California adults were arrested 1.65 million times. Thus, .36 million of the arrests were rearrests, and the multiple capture rate was .36/1.65 = .22 (or 22%) (Shen et al., 2002). Ignoring this factor in prevalence estimation may lead to inflated estimates. The current study (like its developmental predecessors) used a one-year period for assessment of multiple capture rates, corresponding to the time period of calibrator data.

In the current study, these rates were calculated using arrest data from California (MACR), which contained identifiers that could be used for counting multiple records for each individual. Multiple capture rates were calculated for the 24 combinations of gender, age group, and offense category, as well as by selected marginals (e.g. for males and females) and for the entire arrestee population. Since at present the UCR data do not support the intrinsic calculation of these rates and consistently defined data were not available for gender-by-age-by-offense subgroups for each state, the calculated California rates were applied to the national data. These multiple capture rates were then multiplied (in Step 4 described below) by the numbers of arrests in the relevant category from the national arrest database (UCR) to obtain an adjustment number; this adjustment number of multiple arrests is then subtracted from the total to produce a number of arrestees for the year 2000 (that is, one count per arrestee). Note that the extent to which the California rates are generalizable affects the final prevalence estimates. If the national rates are, in fact, lower than those for California, then the final prevalence results for the U.S. will be an underestimate; and if the national rates are higher than those for California, then the final prevalence results will be an overestimate. Further discussion of multiple capture rates appears in *Evaluation of Estimation*.

In the matching procedure, the first arrest for an individual determined that individual's subgroup classifications. Identifiers used for matching of arrest records were last name, first name, middle initial, and date of birth. An exact match on the four identifiers was required. This computer matching procedure is conservative and may miss some possible matches because of arrestee use of aliases or differences in presentation of names. For example, John Doe, John M. Doe would be recognized as different arrestees. However, the use of standard identification (e.g. drivers license) in the arrest process minimizes the chance of differential name presentation, but does not preclude the possibility of aliases. Previous work comparing the computer procedure and examination by hand for a sample of arrest records showed slightly higher possible multiple capture rates (by 4-8%) using the manual matching procedure with researcher judgement of match (Anglin et al., 1999). Other probabilistic matching procedures are available (see for example, Committee on Applied and Theoretical Statistics, National Research Council, 1999). But this study elected to use the simple, straightforward, and less costly deterministic matching procedure. A further benefit is that it is the more conservative estimate, as well as the most practical.

Step 4: Calculate prevalence (numbers of drug-using arrestees)

First, numbers of arrests in the UCR data were adjusted by the relevant multiple capture rates calculated in Step 3 to give numbers of arrestees; this is done for each subcomponent geographic area (i.e. individual states within the U.S., and counties within California) for each gender by age group by offense category subgroup. Then the predicted drug prevalence rates calculated in Step 2 were multiplied by these arrestee numbers to give prevalence estimates. Within each gender by age group by offense category, prevalence estimates for subcomponent geographic areas (each of 50 states) were summed to give the national prevalence figures; and 58 county prevalence estimates were summed to give the California estimate. Subgroup estimates were summed to give marginal totals for gender, age group, gender by age group, and gender by offense category as well as the estimated overall total. Standard errors for the prevalence estimates were calculated with SAS PROC IML; description of the statistical basis for these algorithms is given in *Appendix 5*.

RESULTS

Positive Urine Rates

Calculation of positive urine rates for the gender by age group by offense category subgroups was an intermediate result of the first step of analysis. The rates for the year 2000 ADAM calibration sample are presented in Table 1. Rates are shown for any illicit drug, for cocaine, opiates, and methamphetamine. Within the ADAM sample, rates of positive urines for any illicit drug, for methamphetamine, and for opiates are roughly similar overall for males and female arrestees (66% vs. 68%, 10% vs. 12%, and 8% vs. 11%, respectively); for cocaine use, rates are higher overall for females than for males (39% vs. 30%). However, rates vary considerably across the gender by age by offense categories, as well as across types of drugs. For any illicit drug, rates vary from a low of 48% for 18-24 year-old females arrested for violent offenses to 90% for older (both 25-34 and 35 and older) females arrested for drug-related offenses; variability is only slightly less among males, from 52% among 35-and-older males arrested for violent offenses to 89% for older males arrested for drug-related offenses. With the notable exception of the youngest age category for any illicit drug use, rates are more often higher for females than for males. When considering offense categories, rates are highest among arrestees charged with drug-related crimes as would be expected; this holds across all age groups, genders, and types of drugs.

Note also (not shown in Table 1) that the pattern of rates across specific ADAM sites is somewhat unusual for methamphetamine, where there are some sites with extremely high rates of methamphetamine use, and many sites with none or very little. This pattern, anchored at the low end with a large group of zeroes, appears to weaken the validity of estimation results for methamphetamine; this issue is discussed in more detail in the section on *Evaluation of Estimation*. The ADAM site-specific urine-positive data noted above have not been included in this report, but are available from the ADAM website or from ADAM annual reports (NIJ, 2001).

Table 1. Percentages of Positive Urinalysis in 2000 ADAM Calibration Data for Any Illicit Drug, Cocaine, Opiates, and Methamphetamine by Gender, Age, and Offense Category

	Any Drug (%)		Cocaine (%)		Methamphetamine (%)		Opiates (%)	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Age 18-24	**69.1**	**60.0**	**17.7**	**18.3**	**8.3**	**11.3**	**4.8**	**5.8**
Violent	61.9	47.7	13.4	9.5	6.0	5.1	3.2	1.2
Drug-related	86.3	84.7	25.5	31.8	12.6	19.3	6.9	13.7
Property	68.4	53.0	19.3	11.2	9.7	14.0	7.2	5.4
Other	63.6	58.4	14.8	21.1	6.5	8.2	3.3	4.2
Age 25-34	**64.7**	**69.3**	**29.5**	**42.9**	**12.1**	**14.4**	**7.5**	**11.7**
Violent	53.0	47.9	21.7	23.6	9.1	10.4	4.5	4.5
Drug-related	86.5	89.8	42.4	60.9	19.8	19.3	13.4	19.5
Property	69.7	65.8	40.4	37.6	11.9	16.8	9.8	12.3
Other	59.5	68.2	23.8	43.5	10.4	11.5	5.5	9.7
Age 35+	**63.9**	**71.8**	**40.0**	**50.6**	**9.5**	**9.9**	**10.3**	**14.7**
Violent	51.6	54.8	28.8	29.9	6.3	9.2	4.3	8.1
Drug-related	89.1	89.9	61.1	68.0	16.1	11.8	19.7	24.5
Property	71.6	67.0	51.1	44.3	8.7	8.6	13.8	15.0
Other	55.4	66.5	31.3	48.0	8.5	9.4	7.5	9.5
Total	**65.8**	**67.7**	**29.8**	**39.2**	**9.9**	**11.8**	**7.7**	**11.3**

Logistic Regression

For any illicit drug use, the calibration phase produced logistic regression models with acceptable fit ($p<=.15$) for 11 of the 12 male age by offense subgroups and 10 of the female subgroups (*Technical Note 3*). Examining the relationships of any illicit drug use to the five socio-demographic indicators across the 24 subgroup models, the most frequently significant predictors of any illicit drug use were percent below poverty level and percent with high school education. (Logistic regression results for any illicit drug use are given in *Appendix* 6.) There is good consistency of relationships among the male age by offense subgroups, with poverty a strong ($p<=.10$) predictor of drug use rate in 11 of the 12 subgroups (positively related) and with education a strong predictor in 10 of the 12 subgroup models (positively related); other variables were rarely significant. That is, controlling for the other four predictors, higher rates of poverty were associated with higher rates of illicit substance use, and higher rates of high school graduates were associated with higher rates of illicit substance use (*Technical Note 4*). For females, education was a strong predictor in eight of the 12 subgroups (all but one of those positively related). Note that socioeconomic indicators are frequently at least somewhat intercorrelated, limiting the interpretation of individual regression coefficients, but not decreasing the predictive capability of the models.

For cocaine, the fit for all 12 estimated models for male age by offense subgroups and eight of the 12 for females were adequate. As with any illicit drug use, poverty was a strong and consistent predictor of cocaine use, in eight of the male subgroups and nine of the female subgroups (positively related). Other results for cocaine differed somewhat from those for any illicit drug use. Percentage of young was a strong predictor of cocaine use (negatively related) in seven of the male models and six of the female models. Education was a strong positive predictor of cocaine use for females (in seven of the models); but for males, education was a negative predictor in all four models for the youngest age group (18-24 years) while a positive predictor in three of the other older age group models.

Further assessment of these logistic regression models in terms of replicability appears in the *Evaluation of Estimation* section.

Numbers of UCR Arrests

For the UCR data, Table 2 shows numbers of arrests by subgroup. These numbers form the basis for calculations and projections in Step 4 of the estimation procedure. They are displayed here for descriptive purposes only. Approximately four times more arrests in the U.S. were for males than for females (10,119,565 vs. 2,508,291); the male to female ratio was approximately the same for California and for Alpine County, but was slightly higher for Los Angeles County. For the U.S. the numbers of arrests were approximately equal in the youngest (18-24) and oldest (35+) age categories. Because the oldest category has a broader age range and thus represents a larger segment of the population, there are fewer arrests per capita than in the younger age categories.

Table 2. Frequencies of 2000 FBI UCR Arrests by Age, Gender and Offense Category*

	U. S.			California			Los Angeles County			Alpine County		
	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total
Age 18-24	**3,674,954**	**791,052**	**4,466,006**	**410,791**	**83,334**	**494,125**	**96,117**	**17,270**	**113,387**	**50**	**10**	**60**
Violent	405,209	94,301	**499,510**	49,339	8,801	**58,140**	13,346	2,229	**15,575**	4	0	**4**
Drug-related	1,229,080	187,798	**1,416,878**	173,132	34,599	**207,731**	41,628	6,447	**48,075**	21	6	**27**
Property	432,220	165,217	**597,437**	47,457	15,486	**62,943**	13,543	4,144	**17,687**	14	2	**16**
Other	1,608,445	343,736	**1,952,181**	140,863	24,448	**165,311**	27,600	4,450	**32,050**	11	2	**13**
Age 25-34	**2,838,418**	**773,816**	**3,612,234**	**388,425**	**99,508**	**487,933**	**99,527**	**22,419**	**121,946**	**19**	**2**	**21**
Violent	413,888	102,016	**515,904**	48,547	10,644	**59,191**	13,327	2,739	**16,066**	0	0	**0**
Drug-related	792,003	194,744	**986,747**	161,708	47,666	**209,374**	40,930	9,792	**50,722**	6	0	**6**
Property	301,544	151,370	**452,914**	29,315	13,243	**42,558**	8,577	3,509	**12,086**	5	1	**6**
Other	1,330,983	325,686	**1,656,669**	148,855	27,955	**176,810**	36,693	6,379	**43,072**	8	1	**9**
Age 35+	**3,606,193**	**943,423**	**4,549,616**	**520,049**	**138,373**	**658,422**	**129,190**	**31,461**	**160,651**	**37**	**15**	**52**
Violent	491,656	111,860	**603,516**	57,465	12,832	**70,297**	14,981	3,162	**18,143**	4	0	**4**
Drug-related	862,179	245,939	**1,108,118**	203,258	64,006	**267,264**	49,839	13,995	**63,834**	3	6	**9**
Property	355,816	161.807	**517,623**	33,343	15,494	**48,837**	10,253	4,257	**14,510**	1	1	**2**
Other	1,896,542	428,817	**2,320,359**	225,983	46,041	**272,024**	54,117	10,047	**64,164**	29	8	**37**
Total	**10,119,565**	**2,508,291**	**12,627,856**	**1,319,265**	**321,215**	**1,640,480**	**324,834**	**71,150**	**395,984**	**106**	**27**	**133**

*See Appendix 4 for types of offenses included in UCR and categorization.

It must be remembered that while UCR participating sites represent almost all (94%) of the U.S. population, there are still arrests not counted in UCR; prevalence results based on an undercount of arrests will likely be an underestimate of drug use prevalence. Further discussion of this issue appears in the *Evaluation of Estimation* section.

Multiple Capture Rates

The MACR contains records of arrests in California; data records represent individual events. In the estimation procedures, the numbers of arrests were transformed into numbers of persons (arrestees) to ensure that estimates of drug use prevalence accurately reflect numbers of unique drug users. The multiple capture rates were calculated for 24 subgroups by combinations of gender, age group, and offense category and by overall and selected marginal totals.

Based on computer matching by personal identifiers, an overall multiple capture rate of 22% (or .22) was calculated for the statewide adult arrestee population. That is, 22% of the arrest records represented the second, third, or later subsequent arrest for individuals (multiple captures or rearrests for individuals). Table 3 shows the multiple capture rates among California adult arrestees in 2000 for gender, age groups, and offense category. While the overall multiple capture rate among California adult arrestees was .22, the rate for males was .23 vs. .18 for females. Rates for age groups varied only slightly, ranging from .23 for the oldest age group (35 and older) to .20 for the middle age group (25-34). The highest multiple capture rates are seen for drug-related crimes (as high as .47 for 18-24 year-old males). These recapture rates are generally in line with those from previous studies using 1996 and 1997 arrest data (Anglin et al., 1999; Shen et al., 2002).

Use of arrests (a substantially larger number) rather than arrestees would have produced an inflated prevalence estimate, representing numbers of arrests during 2000 for which we would expect a positive urinalysis if testing were universal. Ignoring the differential rates among subgroups could bias the subgroup estimates in either direction. For example, using the overall rate of .22 would be satisfactory for estimation of total numbers of arrestees with recent illicit drug use; however, if that overall rate were applied to female arrests, the calculated number of female arrestees would be too low and prevalence estimates would also be too low.

Table 3. Multiple Capture Rates among California Adult Arrestees 2000 by Gender, Age, and Offense Category*

	Male	Female	Total
Age 18-24	**0.23**	**0.18**	**0.22**
Violent	0.20	0.11	**0.18**
Drug-related	0.47	0.35	**0.45**
Property	0.21	0.12	**0.18**
Other	0.19	0.16	**0.18**
Age 25-34	**0.20**	**0.18**	**0.20**
Violent	0.16	0.11	**0.15**
Drug-related	0.33	0.29	**0.32**
Property	0.21	0.16	**0.19**
Other	0.18	0.17	**0.18**
Age 35+	**0.24**	**0.19**	**0.23**
Violent	0.16	0.12	**0.15**
Drug-related	0.30	0.24	**0.28**
Property	0.25	0.15	**0.22**
Other	0.25	0.19	**0.24**
Total	**0.23**	**0.18**	**0.22**

* Calculation was based on the 2000 California monthly Arrests and Citation Registers database from the California Department of Justice.

Estimated Prevalence

Prevalence estimates are described below for the U.S. for any illicit drug and separately for cocaine. Then estimates are presented for the prevalence of use of any illicit drug for California and its largest and smallest counties. Because of greater empirical stability in estimates, the most detail is given in describing results for the prevalence of use of any illicit drug for the U.S. Results have not been included in this section for opiates and methamphetamine because the variability in the penetration of these drugs across geographic areas appears to reduce the validity of the estimates without adjustment; this issue is discussed in more detail in the section on *Evaluation of Estimation* and summary tables appear in *Appendix 8*.

National Estimates

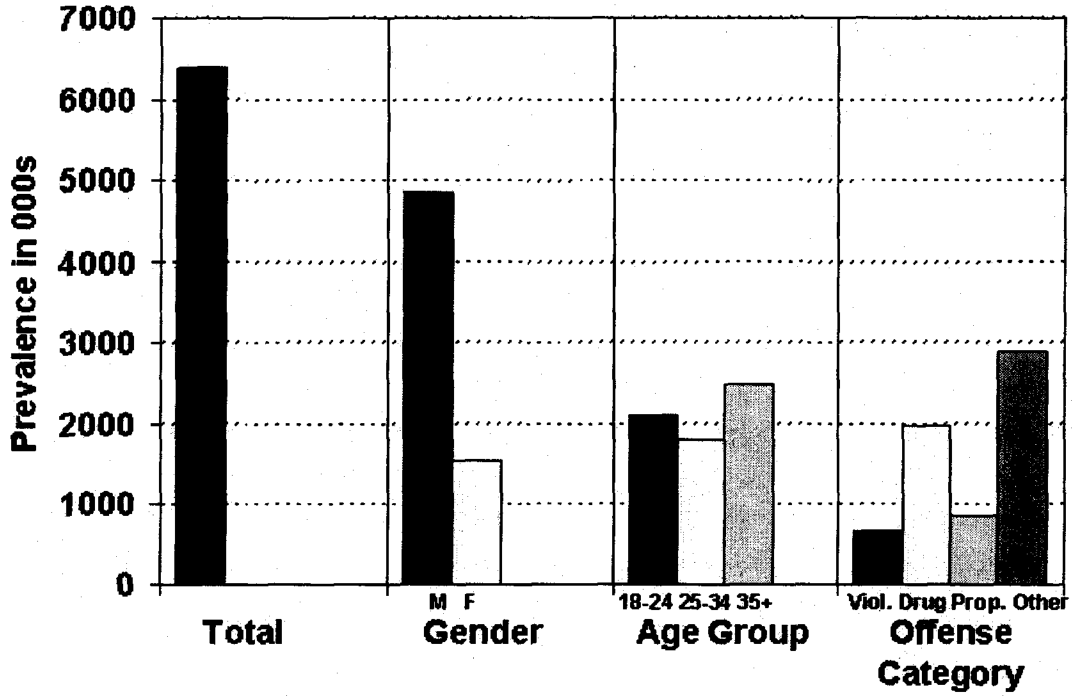
Table 4 shows estimated prevalence for year 2000 of recent users among arrestees of any of 10 illicit drugs for the U.S. by gender, age group, and offense categories. The total is large, 6,395,927 or approximately 65% of the arrestee population for use of any of the 10 illicit drugs as would be measured by urinalysis (see *Technical Note 5*). This number represents the number of arrestees in the U.S. who would test positive for illicit drug use on at least one arrest occasion during 2000 if urine tests were done for every arrest. For males the estimate is 4,851,247 and for females 1,544,680. Figure 1 shows the prevalence estimates in graphical form for gender, age group, and offense category subtotals. As already described, we see substantially higher estimated prevalence for males than females. In terms of age groups, the highest estimate is for the oldest age group (2,485,311), which is also the broadest category in terms of age range, with 2,106,503 in the 18-24 year group and 1,804,114 in the 25-34 year group. By offense category, the highest estimated prevalence is for the "other" offense category (2,880,259), followed by drug-related (1,985,164), property (864,593), and violent crime (665,910), in that order.

Table 4: Prevalence of Recent Users of Any Illicit Drug Among United States Adult Arrestees in 2000, Stratified by Gender, Age, & Offense Category[1]

Age Group/ Offense Category	95% CI[2] Lower Limit	**Male Prevalence Estimate**	95% CI Upper Limit	95% CI Lower Limit	**Female Prevalence Estimate**	95% CI Upper Limit	95% CI Lower Limit	**Total Prevalence Estimate**	95% CI Upper Limit
Age 18-24	1,213,316	**1,640,464**	2,067,613	401,027	**466,038**	531,049	1,674,435	**2,106,503**	2,538,570
Violent	172,914	**233,088**	293,262	14,943	**44,454**	73,965	210,521	**277,542**	344,563
Drug	232,848	**462,385**	691,922	69,262	**104,570**	139,879	334,718	**566,955**	799,192
Property	83,641	**203,546**	323,451	53,142	**89,792**	126,443	167,957	**293,338**	418,719
Other	407,124	**741,445**	1,075,766	199,552	**227,222**	254,891	633,203	**968,667**	1,304,131
Age 25-34	1,059,554	**1,332,825**	1,606,097	424,493	**471,288**	518,083	1,526,864	**1,804,114**	2,081,363
Violent	63,457	**162,184**	260,911	18,617	**49,487**	80,357	108,230	**211,670**	315,111
Drug	508,332	**516,143**	523,953	115,469	**129,915**	144,361	629,636	**646,058**	662,480
Property	136,573	**185,752**	234,932	33,967	**60,146**	86,325	190,185	**245,898**	301,611
Other	218,845	**468,747**	718,649	213,226	**231,740**	250,255	449,900	**700,487**	951,074
Age 35+	1,605,513	**1,877,958**	2,150,402	536,795	**607,353**	677,911	2,203,878	**2,485,311**	2,766,744
Violent	58,151	**128,820**	199,490	7,536	**47,878**	88,220	95,325	**176,698**	258,071
Drug	580,770	**588,892**	597,013	181,476	**183,259**	185,042	763,836	**772,151**	780,466
Property	180,081	**216,475**	252,868	93,755	**108,882**	124,010	285,945	**325,357**	364,769
Other	683,307	**943,771**	1,204,235	211,487	**267,334**	323,181	944,721	**1,211,105**	1,477,489
Total	4,275,609	**4,851,247**	5,426,886	1,437,934	**1,544,680**	1,651,425	5,810,475	**6,395,927**	6,981,379

[1] Note that subcategory cells may not add exactly to totals because of rounding
[2] CI=confidence interval

Figure 2: Estimated Prevalence of Users of Any Illicit Drug Among United States Arrestees (2000)



Estimated numbers of arrestees in the U.S. with recent cocaine use are given in Table 5 for the gender, age, and offense subgroups. The overall U.S. total is 3,779,263 (about 38% of the total number of UCR arrestees) with 2,748,461 males and 1,030,802 females. As with any illicit drug use, we see considerably higher estimates for males than for females. Again, we see the highest cocaine prevalence estimates in the oldest age group (2,004,041) but with substantially smaller numbers with decreasing age (1,133,798 for age 25-34 and 641,424 for age 18-24). By offense category, the highest estimated cocaine prevalence is for the "other" offense category (1,594,489), followed by drug-related (1,326,526), property (580,837), and violent crime (277,411), in that order.

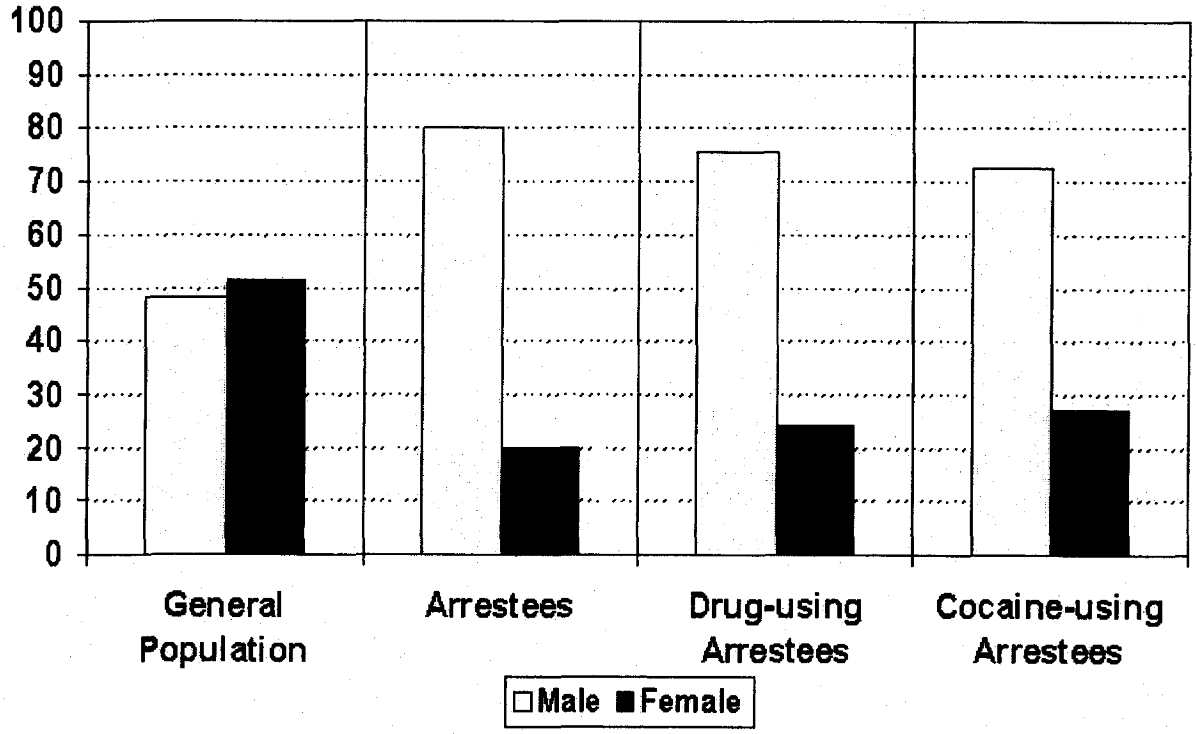
Table 5: Prevalence of Recent Users of Cocaine Among United States Adult Arrestees in 2000, Stratified by Gender, Age, & Offense Category[1]

Age Group/ Offense Category	95% CI[2] Lower Limit	Male Prevalence Estimate	95% CI Upper Limit	95% CI Lower Limit	Female Prevalence Estimate	95% CI Upper Limit	95% CI Lower Limit	Total Prevalence Estimate	95% CI Upper Limit
Age 18-24	141,799	**485,517**	829,236	81,423	**155,907**	230,390	289,728	**641,424**	993,121
Violent	0	**56,321**	149,293	0	**7,868**	28,202	0	**64,189**	159,359
Drug	8,718	**118,006**	227,295	0	**36,205**	81,676	35,841	**154,211**	272,582
Property	0	**88,169**	212,163	41,074	**68,453**	95,832	29,640	**156,621**	283,603
Other	0	**223,022**	509,693	0	**43,381**	91,517	0	**266,403**	557,087
Age 25-34	501,605	**816,009**	1,130,413	249,375	**317,789**	386,203	812,036	**1,133,798**	1,455,559
Violent	1599	**50,272**	98,944	15,571	**40,416**	65,261	36,041	**90,688**	145,335
Drug	288,238	**371,810**	455,383	105,273	**115,376**	125,478	403,005	**487,186**	571,368
Property	29,880	**116,046**	202,211	8,969	**17,636**	26,303	47,082	**133,682**	220,282
Other	0	**277,881**	564,363	82,023	**144,361**	206,699	129,056	**422,242**	715,428
Age 35+	1,088,815	**1,446,935**	1,805,054	503,995	**557,107**	610,218	1,642,004	**2,004,041**	2,366,078
Violent	13,431	**76,876**	140,322	10,123	**45,658**	81,192	49,815	**122,534**	195,253
Drug	473,272	**515,375**	557,478	164,042	**169,754**	175,465	642,640	**685,129**	727,618
Property	170,772	**200,196**	229,619	72,342	**90,338**	108,335	256,044	**290,534**	325,024
Other	305,796	**654,488**	1,003,179	216,692	**251,357**	286,021	555,434	**905,844**	1,256,255
Total	2,160,888	**2,748,461**	3,336,034	916,570	**1,030,802**	1,145,035	3,180,689	**3,779,263**	4,377,838

[1] Note that subcategory cells may not add exactly to totals because of rounding
[2] CI=confidence interval

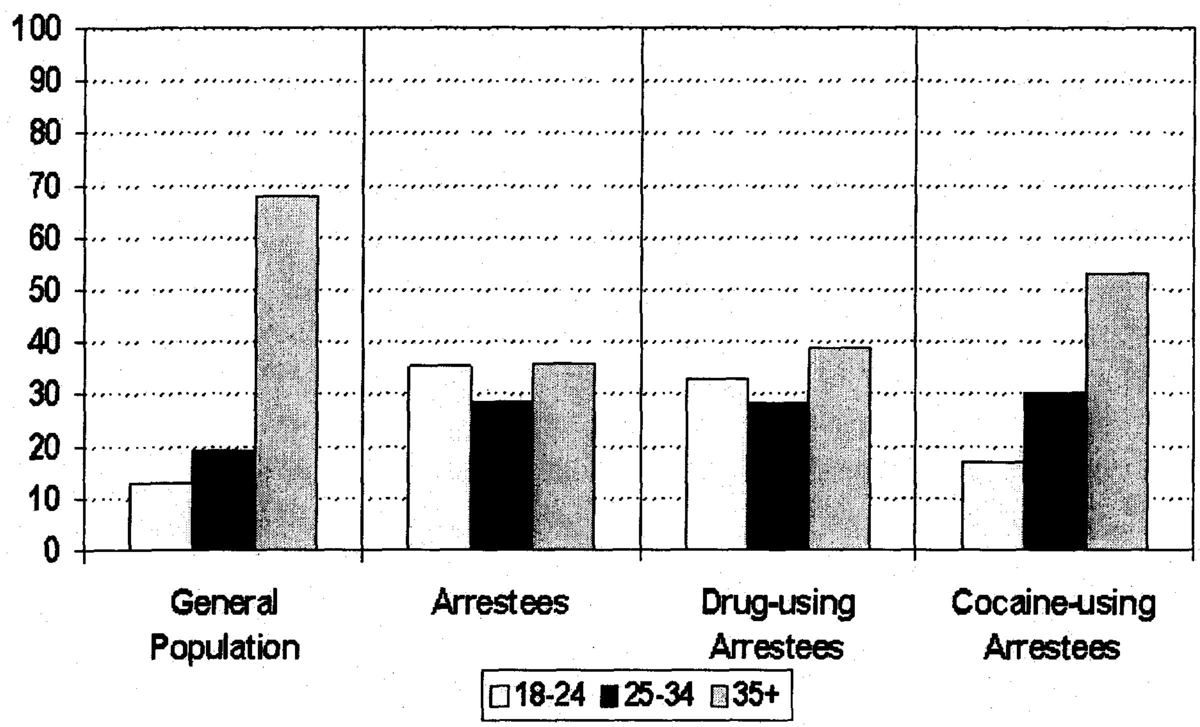
As a way of placing these numbers in context, we present the percentage composition of these prevalence estimates (for any illicit drug and cocaine) by gender, by age groups, and by offense categories (Figures 3-5). The composition of these estimates are placed along side the percentage composition for the general U.S. population and for all U.S. arrestees. Differences in heights of the bars in the figures indicate discrepancies in proportional composition across the populations. In Figure 3, we see that males are over represented in the all-arrestee population compared to their proportion of the general population; and we see that males are over represented (compared to the general population) among arrestees with recent drug use any illicit drug or cocaine). Females constitute slightly larger proportions of drug using arrestees (any illicit or cocaine) than they do of the general arrestee population.

Figure 3: Percent by Gender of the U.S. General Population, Arrestees, and Estimated Drug-Using Arrestees (Any Drug, Cocaine)



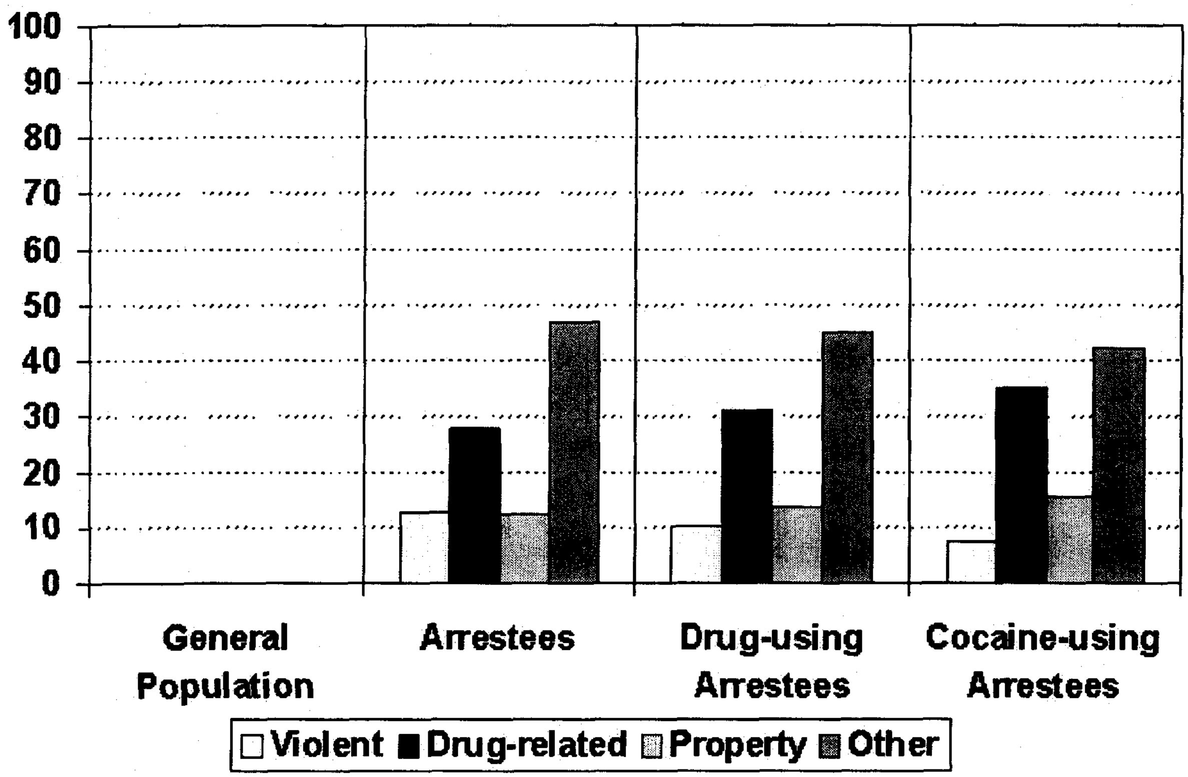
When considering population proportion by age group (Figure 4), we see patterns for the general arrestee and any illicit drug-using arrestees similar to each other and quite different from the general population and with patterns for cocaine-using arrestees different from the other arrestee populations. The largest segment of the general population falls in the 35-year and older category; this follows logically from the broader age range included in this category. However, the arrestee population falls almost equally into the three age categories, indicating that younger age groups are over represented in the arrestee population. The estimated numbers of drug-using arrestees also fall approximately equally across the three age categories, with only a slightly greater percent in the oldest category. The pattern for cocaine-using arrestees appears to more closely resemble that of the general population, with the majority in the 35-year and older age category.

Figure 4: Percent by Age Group of the U.S. General Population, Arrestees, and Estimated Drug-Using Arrestees (Any Drug, Cocaine)



When considering the percentages by offense category (Figure 5), the patterns for arrestees, for drug-using arrestees, and for cocaine-using arrestees are similar. The largest segment of each of these arrestee groups is in the "other" crime category, the next smallest proportion in the drug-related category, followed by property crimes, and last, violent crimes. In spite of the overall similarity of crime proportions for the three arrestee groups, as one might expect the proportion of drug-related arrests in higher for drug-using and for cocaine-using arrestees than for the general arrestee population, and higher for cocaine-using than for arrestee users of any illicit drug.

Figure 5: Percent by Offense Category of the U.S. General Population, Arrestees, and Estimated Drug-Using Arrestees (Any Drug, Cocaine)



State Estimates (California)

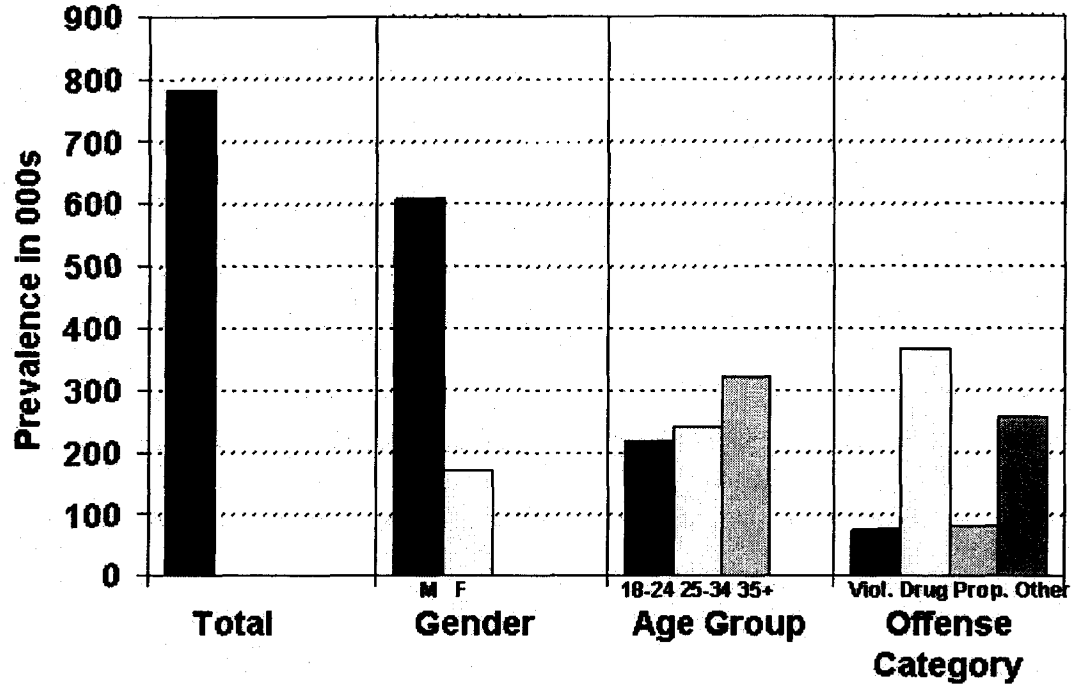
Table 6 shows prevalence estimates of illicit drug users among California arrestees totaling 782,774, approximately 61% of the state's UCR arrestee population for use of any of the 10 illicit drugs. For male arrestees the estimate is 609,678 and for females 173,095. Figure 6 shows the prevalence estimates in graphical form for gender, age group, and offense category subtotals. As already described, we see substantially larger estimated prevalence for males than females. In terms of age groups, the largest estimate is for the oldest age group (321,991), which is also the broadest category in terms of age range, with smaller and roughly similar numbers for the other two groups (218,656 for the 18-24 year group and 242,127 for the 25-34 year group). By offense category, the largest estimated prevalence is for the drug offense category (368,432), followed by "other" offenses (259,012), property crime (80,573), and violent crime (74,758), in that order. The pattern for offense category is quite different from that of the nation as a whole, with a higher proportion of arrests in the drug-related category.

Table 6: Prevalence of Recent Users of Any Illicit Drug Among California Adult Arrestees in 2000, Stratified by Gender, Age, & Offense Category[1]

Age Group/ Offense Category	95% CI[2] Lower Limit	**Male Prevalence Estimate**	95% CI Upper Limit	95% CI Lower Limit	**Female Prevalence Estimate**	95% CI Upper Limit	95% CI Lower Limit	**Total Prevalence Estimate**	95% CI Upper Limit
Age 18-24	171,607	**183,192**	194,778	32,856	**35,464**	38,072	206,781	**218,656**	230,531
Violent	20,584	**23,820**	27,055	2,438	**3,375**	4,312	23,826	**27,195**	30,563
Drug	63,326	**71,388**	79,450	13,864	**15,654**	17,443	78,784	**87,042**	95,300
Property	21,307	**24,826**	28,345	4,730	**5,785**	6,840	26,937	**30,611**	34,285
Other	56,349	**63,159**	69,969	9,382	**10,650**	11,918	66,882	**73,809**	80,736
Age 25-34	171,725	**183,547**	195,370	55,816	**58,580**	61,343	229,986	**242,127**	254,268
Violent	14,922	**17,953**	20,984	2,983	**3,906**	4,828	18,691	**21,859**	25,027
Drug	82,242	**91,019**	99,797	29,848	**31,012**	32,176	113,177	**122,031**	130,886
Property	13,968	**16,073**	18,178	6,338	**7,114**	7,890	20,943	**23,187**	25,431
Other	51,494	**58,502**	65,510	14,351	**16,548**	18,745	67,706	**75,050**	82,394
Age 35+	229,841	**242,939**	256,037	70,792	**79,052**	87,312	306,506	**321,991**	337,476
Violent	17,838	**21,267**	24,696	3,497	**4,436**	5,375	22,148	**25,704**	29,259
Drug	109,227	**117,995**	126,764	33,909	**41,364**	48,818	147,850	**159,359**	170,868
Property	16,625	**18,219**	19,813	7,322	**8,556**	9,790	24,759	**26,775**	28,791
Other	76,492	**85,457**	94,422	21,495	**24,696**	27,898	100,634	**110,153**	119,673
Total	588,570	**609,678**	630,787	164,003	**173,095**	182,187	759,791	**782,774**	805,757

[1] Note that subcategory cells may not add exactly to totals because of rounding
[2] CI=confidence interval

Figure 6: Estimated Prevalence of Users of Any Illicit Drug Among California Arrestees (2000)



County Estimates (Los Angeles and Alpine)

Table 7 shows illicit drug use prevalence estimates for recent users (among arrestees) of any of the 10 illicit drugs, totaling 198,580 (approximately 64% of the arrestee population) for Los Angeles County, the largest county in California. A graph has not been given for county estimates, since patterns are similar to those of the state of California. We see substantially higher estimates for males (156,649) than for females (41,930). In terms of age groups, the highest estimate is for the oldest age group (88,266), followed by the age 25-34 category (62,090) and age 18-24 (48,223). By offense category, the highest estimated prevalence is for the drug offense category (95,625), followed by "other" offenses (60,306), property crime (23,328), and violent (19,320), in that order.

Table 7: Prevalence of Recent Users of Any Drug Among Los Angeles County Adult Arrestees in 2000, Stratified by Gender, Age, & Offense Category[1]

Age Group/ Offense Category	95% CI[2] Lower Limit	Male Prevalence Estimate	95% CI Upper Limit	95% CI Lower Limit	Female Prevalence Estimate	95% CI Upper Limit	95% CI Lower Limit	Total Prevalence Estimate	95% CI Upper Limit
Age 18-24	33,058	**40,482**	47,905	6,075	**7,742**	9,408	40,615	**48,223**	55,832
Violent	5,101	**7,180**	9,259	496	**1,007**	1,519	6,046	**8,187**	10,328
Drug	10,756	**15,605**	20,455	1,481	**2,596**	3,711	13,225	**18,201**	23,178
Property	3,790	**6,555**	9,319	683	**1,485**	2,287	5,162	**8,040**	10,918
Other	6,712	**11,142**	15,572	1,860	**2,653**	3,446	9,295	**13,795**	18,296
Age 25-34	42,105	**48,596**	55,086	12,153	**13,495**	14,837	55,463	**62,090**	68,718
Violent	2,623	**4,970**	7,317	676	**1,365**	2,054	3,889	**6,335**	8,781
Drug	25,531	**26,678**	27,824	5,842	**6,464**	7,085	31,837	**33,141**	34,446
Property	3,363	**5,011**	6,659	653	**1,303**	1,953	4,542	**6,314**	8,086
Other	6,229	**11,937**	17,646	3,644	**4,363**	5,082	10,547	**16,301**	22,054
Age 35+	59,993	**67,572**	75,151	19,155	**20,694**	22,233	80,532	**88,266**	96,000
Violent	1,497	**3,998**	6,500	164	**800**	1,436	2,218	**4,798**	7,379
Drug	32,602	**33,957**	35,313	9,860	**10,326**	10,791	42,850	**44,283**	45,716
Property	4,968	**6,225**	7,481	2,008	**2,750**	3,492	7,515	**8,974**	10,434
Other	16,480	**23,392**	30,304	5,725	**6,818**	7,912	23,212	**30,210**	37,208
Total	144,212	**156,649**	169,087	39,295	**41,930**	44,566	185,866	**198,580**	211,293

[1] Note that subcategory cells may not add exactly to totals because of rounding
[2] CI=confidence interval

Table 8 shows estimates for California's smallest county, Alpine County, totaling 69, approximately 66% of the county's arrestee population. We see substantially higher estimates for males (56) than for females (13). In contrast to Los Angeles County, the youngest group age 18-24 has the highest prevalence (32), followed by 25 in the oldest age group and 12 for ages 25-34. By offense category, the highest estimated prevalence is for "other" offenses (27), followed by drug (23), property (14), and violent offenses (5).

Table 8: Prevalence of Recent Users of Any Drug Among Alpine County Adult Arrestees in 2000, Stratified by Gender, Age, & Offense Category[1]

Age Group/ Offense Category	95% CI[2] Lower Limit	Male Prevalence Estimate	95% CI Upper Limit	95% CI Lower Limit	Female Prevalence Estimate	95% CI Upper Limit	95% CI Lower Limit	Total Prevalence Estimate	95% CI Upper Limit
Age 18-24	25	**27**	29	3	**5**	7	29	**32**	35
Violent	2	**3**	3	0	**0**	0	2	**3**	3
Drug	9	**10**	11	1	**3**	5	11	**13**	15
Property	7	**8**	10	0	**1**	1	7	**9**	11
Other	5	**6**	7	0	**1**	1	6	**7**	8
Age 25-34	10	**11**	12	1	**1**	2	11	**12**	13
Violent	0	**0**	0	0	**0**	0	0	**0**	0
Drug	3	**4**	4	0	**0**	0	3	**4**	4
Property	3	**3**	4	1	**1**	1	4	**4**	5
Other	3	**4**	5	0	**1**	1	4	**4**	5
Age 35+	16	**18**	20	6	**7**	8	22	**25**	28
Violent	2	**2**	2	0	**0**	0	2	**2**	2
Drug	2	**2**	2	4	**4**	5	6	**6**	7
Property	1	**1**	1	1	**1**	1	1	**1**	2
Other	11	**14**	16	0	**2**	3	13	**16**	18
Total	53	**56**	60	11	**13**	16	65	**69**	73

[1] Note that subcategory cells may not add exactly to totals because of rounding
[2] CI=confidence interval

Evaluation of Estimation

To provide a context for considering results, we briefly address several selected issues related to reliability and validity. As already described, synthetic estimation approaches have a history of use in estimating drug prevalence. The specific logistic regression approach of the current study has been used several times in the past decade with various degrees of success. But besides the conceptual and statistical appropriateness of a method, evaluation of the estimation also requires additional support of reliability and validity. We touch briefly on background issues relating to data and definitions and then consider selected methodology issues in somewhat more detail, especially as applicable for interpretation.

Calibration Sample Representativeness

It is important to note that ADAM is a sample of *booked* arrestees, and thus may not represent exactly the same types of arrests as in the UCR data used for projection. Thus, we estimate calibration models for booked arrestees and must assume that similar relationships of sociodemographic variables to prevalence rates hold among all arrestees within the gender by age by offense subgroups. Other exploratory studies suggest that non-booked arrests often are for less serious offenses (many falling into our "other" offense category) with possibly lower rates of drug use (e.g. Abt Associates, 2001). Because our analysis includes categorization by type of offense, we are minimizing potential bias due to differential arrest representation; however, we must still assume that calibration relationships for the "other" offense categories among ADAM booked arrestees hold for "other" offense arrests among all (including non-booked) arrests. Further discussion appears below in the paragraphs on UCR data.

While ADAM is not a random sample of sites, it does contain geographic and socioeconomic diversity important for calibration. Sites represent a broad range of county population totals, from 193,000 to over 9,000,000. A frequent criticism of use of ADAM data for estimating national drug use prevalence among arrestees is the possible under representation of rural areas of the U.S. and the often-suggested resulting overestimation of prevalence. The use of a calibration sample that does not cover the entire range of possible attributes (e.g. from the smallest to largest geographic unit population) requires that we assume a continuity of *relationships* that extend to values outside the range of the calibration sample. That is, if population size is related within the ADAM sample range to drug use, then the *relationship* would allow us to predict drug use in a smaller county. This does not preclude the possibility of different levels of prevalence for less populous counties (see also *Technical Note 5*); population size is included as a predictor in the calibration models and thus the models should allow appropriate estimation of prevalence for small as well as large population counties. In addition, the ADAM site values for predictors other than population cover approximately the same range as do the states used for projection (see also *Technical Note 6*). However, expansion, both in terms of numbers of sites and diversity, could improve ADAM utility for estimation.

Another possible limitation in ADAM data representation comes from less than 100% completion rate for urine tests (approximately 12% overall refusal rate). No adjustment was done in the current study for refusals; thus we are assuming the same prevalence rates and relationships among those who would refuse urine tests as for the observed data. If prevalence rates were higher among those who would refuse, then our estimates will be an underestimate of overall prevalence. We recommend that future development should consider adjustments.

The current study has used unweighted ADAM data for reasons already listed. However, we explored the use of weighted data in estimation for any illicit drug use. ADAM rates of positive urine results from weighted data were similar to those from unweighted data among the gender by age by offense categories, differing by less than 1.3%. Using the same analysis procedures as for unweighted data, the total prevalence estimate for any illicit drug use was remarkably similar (6.3 million) to that based on unweighted data. Since using weighted data may also affect the standard errors in the calibration models, we also produced an estimate adjusting for sampling weights in the logistic regression (see *Technical Note 7*); this also produced an estimate of 6.3 million but with slightly larger confidence intervals (± .77 million) than with unweighted data (± .58 million). While using analyses that adjust for sampling weights in standard errors adds complexity to the estimation procedure, we recommend further exploration of the utility of this approach.

Arrest Population for Projection

UCR data provides consistent data across the U.S. While quite comprehensive, UCR nevertheless is an undercount of arrests for several reasons. Approximately 6% of the U.S. population was not represented by agencies contributing data in 2000 (FBI, 2001). While unimputed data were not used in the current study, exploration of adjustment for incomplete coverage is recommended for future studies to assess the complexity of calculation and the impact on prevalence estimates. Not all crimes are reported; but a majority of and serious crimes are included (see *Appendix 4*). Because UCR data are available only in aggregate form, there is some limitation in terms of flexibility in defining subgroups (e.g. ethnicity, age) for estimation. In addition, if it were event-based with identifiers that would allow matching multiple records for individual arrestees, then multiple capture rates could be calculated for designated geographic subcomponents.

Other Data Issues

Census data used in this estimation were for the most part concurrent with the arrest data (for the year 2000). However, census data are comprehensively updated only periodically and available data for prevalence estimation may be out of date or may represent projections from a prior census. Rapid availability of all data would improve the validity of prevalence estimation results as an indicator of current conditions; this, in turn, could facilitate quicker policy response in changing environments. We have intentionally kept data requirements for this study to a minimum, using only readily available data. Many suggested improvements to the methodology require additional data, which could decrease the practical utility of such improvements.

Multiple Capture Rates

California data (MACR) were used to estimate the multiple capture rates for deduplicating UCR numbers of arrests into numbers of arrestees. Prevalence estimates may be biased up or down depending on the extent to which the California rates represent all arrests. We suggest further exploration on this issue. One approach would be to use geographic unit-specific multiple capture rates, when such consistent data can be identified. Another approach would be to model rearrests based on self-reported arrest patterns in ADAM (e.g. Rhodes & Kling, 2002).

Measure of Drug Use

The estimation has used urinalysis results as an objective measure of drug use. Prevalence estimates thus indicate number of arrestees with recent drug use (typically within two to three days). This measure is more reliable than self-reported drug use, which as described earlier in the report is often underreported, and gives a global indicator of drug use among arrestees. It should capture most chronic users but may also include some occasional users (c.f. Rhodes & Kling, 2002). The measure does not, however, give any direct indication of severity of use nor of less proximal drug use. Nevertheless, the measure has been used as a proxy indicator for need for drug use treatment; see, for example, Shen et al. (2002) for discussion and justification. It is important to consider the match between policy purpose of prevalence estimation and the measure of drug use used in estimation. For some policy purposes, we should be concerned about any (even occasional) illicit substance use among arrestees, because of the relationships between drug and crime. Other measures of drug use, e.g. frequency (to indicate chronic use), severity or dependence scales, may be more appropriate for other contexts.

Calibration Models

As described in the *Results* section, logistic regression models for some gender by age group by offense category subgroups were not statistically significant. These were primarily for certain female subgroups. One reason for this lack of statistical significance may be the smaller ADAM sample sizes for women; for example, model fit improves if the female sample size is doubled from that observed (assuming that rates of positive urine tests remained consistent with those observed). This suggests the utility of including a larger sample of female arrestees in ADAM. Not, however, that lack of statistical significance in some subgroup calibration models does not necessarily invalidate estimation results, since additional analysis of predicted values (described below) shows very good replicability of observed rates by the estimated regression models. But these results do suggest the utility of reconsidering and expanding the set of socioeconomic indicators to improve prediction.

Another assessment of the logistic regression models is how well their results can replicate the observed prevalence rates (similar to an analysis of residuals for simple regression analysis). Predicted prevalence for each site was computed, summing across predicted prevalence from the gender by age by offense calibration models, and compared to the observed prevalence using the mean absolute percentage error (MAPE) (see *Technical Note 9*). For any illicit drug use the MAPE was 5.4%. The calibration models for cocaine prevalence replicated observed values more poorly, with an MAPE of 26%; however, much of the discrepancy appears due to four specific sites (three in California) for which the calibration model overestimated cocaine use. Further study of unique characteristics of these sites may suggest additional predictors to add to the calibration model.

The choice of both the form and specification of the calibration model can affect estimation results. The current study use a simple logistic regression model, developed and applied in earlier related work (Hser et al., 1998; Anglin et al., 1999). However, other models are possible; and while extensive exploration was not within the scope of this study, a limited number of alternatives were explored (see *Appendix 7*). Two other forms of the model were applied: a linear model of rates on ADAM site-level data and generalized estimating equations (GEE) with a logit link function on individual arrestee-level data. The first approach produced poorly fitting calibration models, decreased replicability (MAPE of 5.7%), and an apparent overestimate (7.6 million) of illicit drug user among U.S. arrestees. The GEE approach produced acceptable fitting calibration models and replicability (MAPE of 5.4%) and estimate (6.4 million) similar to the simple logistic model.

Model specification was also based on previous development work by the project team. While some predictors were not significant in specific subgroup models, the prediction purpose of the calibration did not require parsimony; so non-significant predictors were not dropped from the models. Exploration suggested that dropping predictors decreased the replicability of observed values (see *Appendix 7*). However, results suggest that prediction was not optimal, and we recommend further exploration of additional predictors. These might include other county sociodemographic characteristics and drug supply indicators. Individual arrestee characteristics could also be considered (if used in conjunction with a GEE approach that can adjust for potential non-independence of arrestees within sites). However, any expansion of calibration predictors also requires similar data to be available for the projection population. While we recommend further study of the calibration model, we argue for keeping models as simple as possible, assessing whether assumptions are met, and evaluating the impact of violations on resulting estimates. (See also Brecht & Wickens (1993) and Wickens (1993) for additional discussion of general issues in choice of models for prevalence estimation). We would continue to argue for consistency of calibration model across subgroups (for a specific drug) to maintain ease of application; but results suggest that different drugs may require different models.

Confidence Intervals

The size of the confidence intervals gives some indication of the reliability (or stability) of the estimates, with narrower confidence intervals desired. Using this criterion, we see that the estimates for any illicit drug use for the total U.S. population and for males and females in the U.S. are very reasonable with 95% confidence intervals of ± 9% of the estimated value for the total U.S. arrestees, ± 12% for male arrestees, and ± 7% for females. As subgroups become more specific, variability in size of confidence intervals increases. Confidence intervals are also relatively small for overall illicit drug use prevalence for state and county levels (±3%, ±6%, and ±6% for California, Los Angeles County and Alpine County, respectively), but with less reliability at the subgroup level. For cocaine users, the confidence interval is acceptable for the overall prevalence estimate (± 9%), but there are larger subgroup confidence intervals and greater variability than for any illicit drug.

Reasonability of Estimates

While there is no "gold standard" against which to check estimates of prevalence, we can apply some reasonability checks. For example, estimates of drug use prevalence should obviously be a number smaller than the number of arrestees. In addition, one would expect estimates to be of similar order of magnitude to results of simpler, less refined calculations. For example, if we apply a simple calculation multiplying the overall ADAM sample urine positive rate (66.2%) for any illicit drug by the UCR number of arrestees, we see a result of 6,520,520, a similar order of magnitude to the estimate of 6,395,927; but we would not expect an exact match since the ADAM sample does not directly represent the entire U.S. arrestee population. Note that this similarity supports the reasonability of the estimate, but does not detract from the utility of the more complicated estimation used in this report: while the discrepancy of 124,593 may be a small proportion of the total number of arrestees, its magnitude has social cost implications. For California the simple projection of ADAM urine positive rates from the three California sites to state arrestees gives an estimate of 849,637, similar in magnitude to our estimate of 782,744.

Estimates, however, for opiates and methamphetamine show substantially greater discrepancy from simple ADAM rate projections and cannot pass the reasonability test without further analysis, which is beyond the scope of the current study. Additional discussion appears below under *Drug Penetration.*

The method appears sensitive to gross differences in frequencies, both in the subgroups and at the different levels of estimation (national vs. state vs. county). The estimates are of appropriate magnitude for the different levels.

We can compare current estimates with applications of the methodology to previous years' data. The estimate for California (782,774 illicit drug users among the 1.28 million arrestees) is consistent with those for earlier years: about 770,000 of the 1.27 million arrestees in 1996 and 789,000 of 1.29 million arrestees in 1997 were current drug users as determined by urine tests (Anglin et al., 1999; Shen et al., 2002). These earlier studies used the same methodology, but based calibration on a broader sample of California booking facilities (CAL-DUF) than the ADAM representation for 2000, and based projections on the state arrest database

(MACR) instead of UCR. The 1996 estimates were based on gender by age group by ethnicity by offense category subgroupings, and the 1997 estimates used gender by age group by ethnicity subgroups. The pattern is consistent with other indicators and estimation studies showing fairly stable illicit drug use among adults over this period nationwide (Abt Associates, 2001; SAMHSA, 2000a,b).

Our estimate for the U.S. prevalence of cocaine-using arrestees is higher than the 2.4 million chronic cocaine users among booked arrestees proposed by Abt Associates (2001). However, there are several differences in assumptions and methods that could properly place the two estimates in perspective. There are differences in who is being estimated. The Abt estimate adjusts UCR arrests (our projection population) downward to represent *booked* arrests for consistency with ADAM representation. Alternatively, our estimate is based on *all* UCR arrests, assuming that relationships calculated from ADAM data can be applied to all arrests in UCR and thus should be greater than for only booked arrestees. There are also differences in the definition of drug use behavior being estimated. The Abt estimate is for "chronic cocaine users," defined as self-reported cocaine use 10 or more times per month, with an adjustment to account for underreporting. Self-report is typically an underestimate of use, and consensus has not been reached on the accuracy of adjustments for truthfulness (e.g. Abt Associates, 2001; Rhodes & Kling, 2002). Our calculations are based on urine test results, thus giving an estimate of number of arrestees with "recent illicit drug use" (as would produce a positive urine test). The Abt estimates were calculated for males and adjusted to include females using the female percent of total arrests, assuming similar drug use behaviors. On the other hand, our analysis suggests that while the overall rate of illicit drug use among ADAM female arrestees is only slightly higher than that of males, many subgroups of female arrestees have considerably higher rates of illicit drug use than do males. This difference could contribute to the lower estimate by the Abt study.

Drug Penetration

The synthetic estimation appears to produce biased estimates when there is substantial lack of penetration of specific drugs in many geographic areas. This is especially true for methamphetamine, which remains a regional problem even as the nationwide penetration is increasing from the earlier designation of a "West Coast" phenomenon. This situation illustrates a problem common in small area estimation, in which geographic units with similar values of socioeconomic indicators may differ widely in drug abuse. It also illustrates a statistical problem in which outcomes for certain units are not representative of random variables but are fixed at zero (certain sites can have no arrestees with methamphetamine use because there is virtually no methamphetamine available in that area).

Several potential adjustments may warrant consideration in future work. One approach would be to adjust the estimation process by inclusion of an indicator variable in the logistic regression that represents geographic penetration (e.g. a binary measure indicating negligible availability using law enforcement or problem indicator data such as Emergency Department data) or other supply indicator. Another approach is to adjust estimation results, using some fraction representing proportion of calibration sites or population in areas with greater than negligible penetration.

For illustrative purposes, we calculated a very simple gross adjustment as the proportion of ADAM sites with observed prevalence of more than 5%. For any illicit drug and for cocaine, this adjustment factor was 1.0 (penetration in all sites). For opiates, 16 of 36 sites had negligible penetration, giving an adjustment factor of .56. For methamphetamine, 20 sites had negligible penetration, giving an adjustment factor of .44. When the national prevalence estimates are multiplied by these adjustment factors, we see no change for any illicit drug use or cocaine use. For opiates, the adjusted prevalence is 1,804,875 (20% of UCR arrestees) and for methamphetamine, 1,484,630 (16% of UCR arrestees). These adjusted figures for opiates and methamphetamine are still higher than overall ADAM rates but now fall into the plausible range. A more sensitive approach might be to base the adjustment on the percentage of states with negligible penetration; but this would require additional state-relevant indicators of penetration. Further work is recommended in this area.

Possible Bias

As with all statistical estimates, the method used here can either under- or over-estimate drug use prevalence depending on a variety of factors including data characteristics, assumptions, and statistical artifacts. Unfortunately, in most situations, the size of bias is usually not easily assessable. One must thoroughly examine factors potentially causing the bias in order to interpret and utilize the prevalence estimates appropriately. In decision-making at each stage of the current analysis, we have usually adopted conservative strategies leading to a likely overall minimum estimate.

Factors that could cause under-estimation of prevalence:

- UCR underreporting. While most jurisdictions report to UCR, the coverage of the U.S. population is still slightly less than complete at 94%. One might account for this underreporting by adjusting our prevalence estimates upward (dividing by .94). Doing so for our national prevalence estimate gives an adjusted total estimate of users of any illicit drug among arrestees in the U.S. of 6,804,178.

- Urine refusal rate. The median rate of refusal for urine tests in the ADAM sites in 2000 was 12% (Taylor et al., 2001). If most or all of these refusals were drug users, a correction for this underreporting would increase the drug use rates in the ADAM calibration sample and the estimated number of drug users.

Factors that could cause over-estimation of drug prevalence:

- Inaccurate multiple capture (rearrest) rate calculation. Because of inaccurate identifiers in the MACR arrest data, multiple captures could still be undercounted, leading to artificially low multiple capture rates, which would have the effect of producing a too-high number of arrestees from which to project prevalence. Based on validation procedures in previous studies, this undercount of multiple captures is small (likely range 4-8%), thus minimizing its overestimation impact on estimated prevalence.

Factors that could cause either under- or over-estimation of prevalence:

- Non-representativeness of ADAM sites. The ADAM data do not include rural or particularly low population counties. If relationships between drug use rates and sociodemographic calibration predictors are qualitatively different in the less populated areas not represented by ADAM than in the ADAM sites, the estimated prevalence could somewhat biased depending on the nature of the difference. Non-uniform penetration of specific drugs across the geographic areas may exacerbate the effects of non-representativeness.
- Inadequacies in socioeconomic indicator data. For example, even with improvements in the 2000 Census, data may undercount certain at-risk populations, thus potentially affecting indicator data used in the estimation.

- Inappropriate multiple capture rates used in projections. This study has used multiple capture rates based on California data to adjust the national numbers of arrests to represent arrestees. If the national rates are higher than those calculated for California, then the final prevalence results for the U.S. will be an overestimate; if the national rates are lower than those of California, then the final prevalence results for the U.S. will be an underestimate.
- Inconsistent drug availability or penetration rate. When availability of specific drugs (such as methamphetamine) has not penetrated certain geographic areas, this appears to bias the estimation. In the current study, this bias was an overestimate; but insufficient study has been done to assess the likely direction of bias under varied circumstances.

In summary, we have noted several factors that may cause bias in the prevalence estimates. While the overall degree of bias is not known in the current study and can never be completely known in a real world situation, the aggregate of several sources of bias where some information is available (e.g. UCR underreporting, urine test refusal, inaccurate multiple capture) suggest our prevalence estimate for any illicit drug use may be a slight underestimate. Further developmental work in these identified areas is encouraged to assess their impact on prevalence estimates.

DISCUSSION

This study makes several major contributions in estimating drug use prevalence among arrestees: a) it provides an example of a relatively simple and cost-effective method of estimating prevalence of illicit drug use among the high risk and high social cost population of arrestees using data available from the ADAM system; b) it provides not only aggregate totals, but also detailed estimates of drug use prevalence for subgroups of arrestees at the national level and for the most populous state and two of its counties; and c) it continues development of the method, identifying issues and suggesting future refinements.

Prevalence Estimates

The study has estimated that 6,395,927 arrestees in the U.S. were users of illicit drugs in the year 2000 (and would have produced a positive urine test on at least one arrest occasion during the year if such tests were given for all arrests); this constitutes approximately 65% of the total number of arrestees. This estimate was based on ADAM as a calibration sample and the use of urinalysis results from the ADAM along with census data and FBI UCR arrest data to project to the national level. In California, an estimated 782,774 (61% of the total number of state arrestees) were recent illicit drug users in 2000. In Los Angeles County, an estimated 198,580 arrestees were recent illicit drug users. In Alpine County, an estimated 69 arrestees were recent illicit drug users. Based on an assessment of possible bias due primarily to data and definitional aspects of the estimation, these estimates for any illicit drug use are likely conservative (that is, too low). Even so, the magnitude of these estimates indicates a high level of potential social cost and emphasizes the challenge to provide adequate prevention and treatment services to this population. The estimated number of cocaine user was more than half of the estimated number of illicit drug users (3,779,263 or 38% of the total number of UCR arrestees).

The estimates provide a rational basis for policy and resource planning. Moreover, they provide a level of detail to support planning for specific geographic areas and major subgroups of arrestees that may have differential needs and thus receive specific services. Such estimates can be particularly useful in a changing social, economic, and political environment. For example, they provide a perspective on the potential impact of legislation that would divert certain types of arrestees to drug treatment instead of prison. Such a program was legislated in California in 2000 and has now been in effect for over one year (SACPA); planning for its impact has been difficult. Similar programs are being considered in other states and advance planning could benefit from prevalence estimates. Because only certain types of offenses are eligible for diversion to treatment, the estimation method could be applied to finer categories of offenses to refine the estimates for specific types of diversion programs.

Use of ADAM and Other Data

ADAM provides a cost-effective and useful infrastructure for obtaining high quality data on the arrestee population. ADAM is the only national survey with objective data (from urinalysis) on drug use and its cost ($8.2 million in 2002) is considerably less than that of the NHSDA (approximately $44 million), which may also underrepresent arrestees. As shown in this report, the ADAM data provide a basis for prevalence estimation. Expansion of ADAM (e.g. additional sites, broader geographic and demographic representation) and more immediate availability of data would enhance its effectiveness in prevalence estimation.

This application has intentionally kept data requirements to a minimum, in order to facilitate the use of the estimation procedure in policy contexts. Many of the potential improvements to estimation could be implemented only by increasing data requirements; and thus the impact of estimation methodology modifications on the reasonability and reliability of estimates must be weighed against the cost and/or difficulty of obtaining appropriate data.

Estimation Method

The continuing development and application of the logistic regression synthetic estimation method has supported its utility in prevalence estimation. The method is applicable with ADAM data as the calibrator; and the estimation across designated subgroups and with relevant socioeconomic indicators should overcome, for policy purposes, the potential limitation of non-random ADAM site selection. Results of the calibration phase of the estimation achieve acceptable replicability of observed data for estimating prevalence any illicit drug use. Overall prevalence estimates for any illicit drug use and for cocaine appear reasonable and reliable, and subgroup totals and many of the more detailed subcategories have acceptable confidence intervals. Results, however, for opiates and methamphetamine suggest a need for further refinement of the method for specific drugs for which availability and use is non-existent or negligible in many geographic areas. A potential correction for specific drug penetration levels appears to be a promising approach that warrants further study.

For some subgroups, while prevalence estimates appear reasonable from a relative perspective, the regression method has resulted in some wide 95% confidence intervals, indicating uncertainty of the estimation. Since the estimation was based on urine positive rates for gender by age group by offense category across the ADAM sites, the variability in confidence intervals may reflect both sample size variability and variation in these rates across ADAM sites. This is particularly true of rates of opiate and methamphetamine use. An increased arrestee sample size for small subgroups may improve the stability of calibration, which in turn will improve the projections to national, state, or county levels. In addition, combining ADAM data across contiguous years may partially resolve this problem, for subgroups with relatively stable prevalence across the combined years.

Results have suggested directions of further exploration of alternative calibration models (both in form and specification) and of adjustments in the procedures to account for inadequacies in data representation. However, modifications to the methodology will likely add complexity for application. While estimation methods are not useful if they do not produce reliable and valid results, they are not likely to be utilized for many policy purposes if their implementation is too costly or complex. We think the current approach, with some selected modifications to improve calibration and estimation for specific drugs, holds promise of balancing these criteria.

An important advantage of the regression method, as demonstrated in this study, is that it can be applied at different levels, including demographic subgroups, counties, and state populations. The method is cost effective, in that it uses available data and is relatively simple to apply. For any illicit drug use total estimates, category subtotals, and some gender by age by offense categories, estimation is sufficiently reliably to allow detection of relatively small changes over time; the application suggests that this holds at national, state, and county level.

Recommendations

We recommend the continued use of this type of prevalence estimation method as a way to provide an information basis for policy development and resources planning and allocation. We also recommend continued refinement of the estimation process, both to data and methodology.

ADAM and Other Data Resources

- Continue data collection through ADAM, expanding ADAM wherever possible. An increase in the number of ADAM sites (both to increase the number of sites and their representation of the nation) would improve its utility for prevalence estimation. An increase in the number of interviewed female arrestees would improve the calibration phase of the prevalence estimation. Increased sample size of arrestees may allow estimation of more refined cross-classifications of arrestee characteristics (e.g. inclusion of ethnicity as well as gender, age, offense). Expanded geographical and socioeconomic diversity will improve ADAM for calibration purposes.

- Encourage improvements in other data resources used in the estimation. For example, use of event-based UCR reporting (with unique arrestee identifiers of some kind) would allow calculation of multiple capture rates for geographic subcomponents, and the use of different arrestee subgrouping categories.
- Continue to improve the timely availability of data needed for the estimation. Implementation of federated data systems could facilitate data availability (Li et al., 1998). With federated data systems, software interfaces allow analysts to access data elements from data systems at their source, thus not requiring physical centralized data archives or transfer of databases.

Estimation Methodology

- Continue prevalence estimation using ADAM. The method used in this study provided acceptable prevalence estimates at the national level for any illicit drug and for cocaine, and at state and county levels for any illicit drug.
- Support the automation and streamlining of this method to facilitate its broader use. Parts of the procedure can be set up for easier use, e.g. through a set of linked SAS macros, to allow design flexibility for specific contexts.

- Continue to study and develop the methodology, comparing modified approaches in order to assess whether improvements in estimation are substantial enough to warrant the considerable increase in complexity, potential cost, and accompanying difficulties in their use:
    - o Include other arrestee subgrouping characteristics (for example, race/ethnicity) in the calibration. However, note that adding subgroups requires a larger calibration sample.
    - o Consider a GEE (generalized estimating equation) approach (rather than simple logistic regression) in calibration, allowing adjustment for within-site correlation of errors and inclusion of individual-level characteristics. Note, however, to include individual-level predictors, these data must be available both in calibration and in projection databases.
    - o Reevaluate social indicators for use in estimation, e.g. other sociodemographic variables, indicators of drug supply and/or penetration (e.g. price, seizures, satellite data allowing identification of methamphetamine labs).
    - o Further study on penetration rate and possible adjustments
    - o Further exploration of multiple capture rates and adjustments, e.g. modeling recapture such that adjustments are linked to geographic unit characteristics
    - o Modeling urine refusal rates

Replication/Validation

- Conduct comprehensive simulation to assess the sensitivity and robustness of results across several dimensions, including calibration sample size (both sites and arrestees), number of subgroups and subgroup size, number of geographic subcomponents for projection, variability of observed prevalence rates and rates of penetration, predictive capability of logistic regression models.

- Apply procedures to 2001 data
- Use procedures to estimate prevalence of other drugs or drug groupings, e.g. any illicit drug excluding marijuana, marijuana, other specific drugs.

Summary

Estimation of prevalence of drug use among arrestees remains a necessity since large-scale population surveys are prohibitively expensive and unlikely to be implemented in the absence of a congressional mandate for this information. However the magnitude of drug use in this high consequence, high cost group is critical to informed policy development and resource planning and allocation, especially where diversion programs are impacting the local treatment system. Therefore, we urge continuing efforts to refine and apply scientifically rigorous estimation approaches that use the valuable data collected through the national ADAM program; the availability of different estimation approaches is advantageous, in order to allow flexibility in definitions, assumptions, and data requirements in order to meet different policy purposes. The logistic regression synthetic estimation approach used in this study appears to provide estimates adequate for describing prevalence and trends over time for any illicit drug use and cocaine at national, state, and county levels. We have made recommendations for further study of possible modifications related to several data and methodological issues that may improve this estimation approach.

REFERENCES

Abt Associates (2001) What America's Users Spend on Illegal Drugs. Report prepared for the Office of National Drug Control Policy. Available at http://www.whitehousedrugpolicy.gov/publications/pdf/spending_drugs_1988_1998.pdf

Anglin, M.D., & Hser, Y.-I. (1992). Drug abuse treatment. In R. Ross Watson (Ed.), Drug and Alcohol Abuse Reviews. (Vol. 4, pp. 1-36) Totowa, NJ: Humana Press.

Anglin, M.D., Hser, Y., & Chou, C. P. (1993). Reliability and validity of retrospective behavioral self-report by narcotics addicts. Evaluation Review, 17(1), 90-107.

Anglin, M.D., Shen, H., Hser, Y.-I. & Brecht, M.-L. (1999). Drug Use Prevalence Estimates among Adult Arrestees in California, Texas, and the U.S.: A special study report submitted to the National Evaluation and Data Technical Assistance Center (NEDTAC) of the Center for Substance Abuse Treatment.

Brecht, M.-L., Anglin, M. D., Shen, H., Urada, D., & Hser, Y.-I. (2002) Integration of Results—California State Treatment Needs Assessment Program. In Department of Alcohol and Drug Programs, Office of Applied Research and Analysis, A State Treatment and Demand Needs Assessment: Alcohol and Other Drugs. Sacramento, CA: State of California Department of Alcohol and Drug Programs, Office of Applied Research and Analysis.

Brecht, M.-L. & Wickens, T. (1993) Application of multiple-capture methods for estimating drug use prevalence. Journal of Drug Issues, 23(2), 229-250.

Committee on Applied and Theoretical Statistics, National Research Council. (1999) Record Linkage Techniques, 1997. Proceedings of an International Workshop and Exposition. Washington, DC: National Academy Press.

CEWG (Community Epidemiology Work Group), (2001) Epidemiologic Trends in Drug Abuse, Advance Report. Washington, DC: NIDA.

Danila, B., Annon, J., & Anglin, M.D. (1997). State demand and treatment needs assessment study: Dependence and abuse of alcohol and other drugs among California arrestees. Final report submitted to the California Department of Alcohol and Drug Programs, March.

Dewit, D. & Rush, B. (1996) Assessing the need for substance abuse services: A critical review of needs assessment models. Evaluation and Program Planning, 19(1), 41-64.

Ducharme, L., & Ball, J. (2001) Major drugs of abuse in ED visits, 2000. The DAWN Report 2001. Washington, DC: Substance Abuse and Mental Health Services Administration (SAMHSA).

Ebener, P., Paddock, S., Khosla, S., McCaffrey, D. & Chien, S. (2002) Assessing need for alcohol and other drug treatment among the California household population. In Department of Alcohol and Drug Programs, Office of Applied Research and Analysis, A State Treatment and Demand Needs Assessment: Alcohol and Other Drugs. Sacramento, CA: State of California Department of Alcohol and Drug Programs, Office of Applied Research and Analysis.

Federal Bureau of Investigation (2001) Crime in the United States, 2000: Uniform Crime Reports. Washington, DC: US Government Printing Office. Available at http://www.fbi.gov/ucr/cius_00/contents.pdf

Harrison, L. (1997) The validity of self-reported drug use in survey research: an overview and critique of research methods. In L. Harrison & A. Hughes, The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates, NIDA Research Monograph, Number 167. Rockville, MD: National Institute on Drug Abuse. Available at http://165.112.78.61/pdf/monographs/monograph167/download167.html

Hser, Y.-I. (1993). Population estimation of illicit drug users in Los Angeles County. Journal of Drug Issues, 23(2), 323-334.

Hser, Y., & Anglin, M.D., (Eds.) (1993). Prevalence estimation techniques for drug-using populations (Special Issue). Journal of Drug Issues, 23(2).

Hser, Y.-I., Longshore, D., & Anglin, M.D. (1992). Drug use prevalence among criminal offender populations: Implications for control and treatment. In D.L. Mackenzie & C.D. Uchida (Eds.), Drug and criminal justice system: Evaluating public policy initiative. Beverly Hills, CA: Sage.

Hser, Y.-I., Maglione, M. & Boyle, K. (1999) Validity of self-report of drug use among STD patients, ER patients, and arrestees. American Journal of Drug and Alcohol Abuse, 25(1), 81-91.

Hser, Y.-I., Prendergast, M., Anglin, M.D., Chen, J.K., & Hsieh, S-C. (1998). A regression analysis estimating the number of drug-using arrestees in 185 US Cities. American Journal of Public Health, 88(3): 487-490.

Hunt, D., & Rhodes, W. (2001a) Methodology Guide for ADAM. Washington, DC: National Institute of Justice.

Hunt, D., & Rhodes, W. (2001b) Sampling Guide for ADAM. Washington, DC: National Institute of Justice.

Kerber, L. & Harris, R. (1998) Substance Use Among Female Inmates. Austin, TX: Texas Commission on Alcohol and Drug Abuse.

Kerber, L. (1998) Substance Use Among Male Inmates. Austin, TX: Texas Commission on Alcohol and Drug Abuse.

Levy, P.S. (1979). Small area estimation—Synthetic and other procedures, 1968-1978. In J. Steinberg (Ed.), Synthetic estimates for small area (NIDA Research Monograph, No. 24). Washington, DC: HEW.

Li, C.S., Bergman, L., Castelli, V., Smith, J., Thomasian, A., Lele, S., Patz, A. & Glass, G. (1998) Model-based mining of remotely sensed data for environmental and public health applications. In S. Wong (Ed.), Medical Image Database. Dordrecht, The Netherlands: Kluwer Academic.

McAuliffe, W., LaBrie, R., Lomuto, N., Pollock, N., Betjemann, R., & Fournier, E. (2000) Measuring interstate variations in problems related to alcohol use disorders. In Wilson, R. & Dufour, M., The Epidemiology of Alcohol Problems in Small Geographic Areas (NIAAA Research Monograph No. 36), Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.

Magura, S. & Kang, S. (1996) Validity of self-reported drug use in high risk populations: A meta-analytical review. Substance Use and Misuse, 31(9), 1131-1153.

Maxwell, J. (2000) Methods for estimating the number of "hard-core" drug users. Substance Use & Misuse. 35(3), 399-420.

Mieczkowski, T., Newel, R., & Wraight, B. (1998) Using hair analysis, urinalysis, and self-reports to estimate drug use in a sample of detained juveniles. Substance Use and Misuse, 33(7), 1547-1567.

Morral, A., McCaffrey, D., & Iguchi, M. (2000) Hardcore drug users claim to be occasional users : drug use frequency underreporting. Drug & Alcohol Dependence, 57(3), 193-202.

NIJ (National Institute of Justice) Office of Justice Programs (2001) 2000 Annualized Site Reports, ADAM. Washington, DC: National Institute of Justice. Available at http://www.adam-nij.net/files/2000AnnualReports.pdf

ONDCP (Office of National Drug Control Policy) (2002) PulseCheck, Trends in Drug Abuse, April 2002. Washington, DC: ONDCP.

Person, P.H., Jr., Retka, R.L., & Woodward, A.W. (1978). Toward a heroin problem index—An analytical model for drug abuse indicators (NIDA Technical Paper). Washington, DC: U.S. Government Printing Office.

Person, P.H., Jr., Retka, R.L., & Woodward, A.W. (1977). A method for estimating heroin use prevalence (NIDA Technical Paper). Washington, DC: U.S. Government Printing Office

Rhodes, W. (1993). Synthetic estimation applied to the prevalence of drug use. Journal of Drug Issues, 23(2), 297-321.

Rhodes, W. & Kling, R. (2002) Estimating the Prevalence of Hardcore Drug Use Using ADAM Data. Preliminary report. Abt Associates.

Rhodes, W. (2002) Estimating Trends Using Drug Use Forecasting Data. Abt Associates.

(SAMHSA) Substance Abuse and Mental Health Services Administration, Office of Applied Studies (2001) Summary of Findings from the 2000 National Household Survey on Drug Abuse. NHSDA Series H-13, DHHS Publication No. SMA 01-3549. Rockville, MD. Available at http://www.samhsa.gov/oas/nhsda.htm

(SAMHSA) Substance Abuse and Mental Health Services Administration, Office of Applied Studies. (2002a). Results from the 2001 National Household Survey on Drug Abuse: Volume I. Summary of National Findings. NHSDA Series H-17, DHHS Publication No. SMA 02-3758. Rockville, MD. Available at http://www.samhsa.gov/oas/nhsda.htm

(SAMHSA) Substance Abuse and Mental Health Services Administration, Office of Applied Studies. (2002b) Emergency Department Trends from the Drug Abuse Warning Network Final Estimates 1994-2001. DAWN Series D-21, DHHS Publication No. SMA 02-3635. Rockville, MD. Available at http://www.samhsa.gov/oas/DAWN/Final2k1EDtrends/text/EDtrend2001v6.pdf

Schmeidler, J. (1991). Projecting sample results to substate areas with known population characteristics. In S. Wallack (Chair), Resource materials for state needs assessment studies. Washington, DC: Office of Treatment Improvements.

Schmidt, L. & Weisner, C. (2000) Estimating alcohol problems in small areas: Contrasting data sources, definitions, and measures. In Wilson, R. & Dufour, M., The Epidemiology of Alcohol Problems in Small Geographic Areas (NIAAA Research Monograph No. 36), Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.

Shen, H., Brecht, M.-L., Anglin, M. D., Hser, Y.-I., & Urada, D. (2002) Prevalence of Drug Treatment Need among Adult Arrestees in California, 1997. In Department of Alcohol and Drug Programs, Office of Applied Research and Analysis, A State Treatment and Demand Needs Assessment: Alcohol and Other Drugs. Sacramento, CA: State of California Department of Alcohol and Drug Programs, Office of Applied Research and Analysis.

Simeone, R., Frank, B., & Aryan, Z. (1991). Social indicator models. In S. Wallack (Chair), Resource materials for state needs assessment studies. Washington, DC: Office of Treatment Improvement.

Simeone, R., Rhodes, W., & Hunt, D. (1995a). A plan for estimating the number of "hardcore" drug users in the United States. International Journal of the Addictions, 30(6): 637-657.

Simeone, R., Rhodes, W., & Hunt, D. (1995b). Social indicator models. In S. Wallack (Chair), Resource materials for state needs assessment studies. Washington, D.C.: Office of Treatment Improvement.

Suciu, G., Hoshaw-Woodard, S., Elliott, M., & Doss, H. (2001) Uninsured Estimates by County: A Review of Options and Issues. Report to the Ohio Dept. of Health.

Taylor, B., Fitzgerald, N., Hunt, D., Reardon, J., & Brownstein, H. (2001) ADAM Preliminary 2000 Findings on Drug Use and Drug Markets--Adult Male Arrestees. Washington, DC: National Institute of Justice.

Wilson, R. (2000) A social indicator methodology for estimating the prevalence of alcohol problems in small geographic areas. . In Wilson, R. & Dufour, M., The Epidemiology of Alcohol Problems in Small Geographic Areas (NIAAA Research Monograph No. 36), Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.

Tweed, D. L. & Ciarlo, J. (1992) Social-indicator models for indirectly assessing mental health service needs. Evaluation and Program Planning, 15(2), 165-179.

Wilson, R. & McAuliffe, W. (2000) Synthetic estimates of the prevalence of DSM-III-R alcohol abuse and dependence: Issues of validity and methodology. In Wilson, R. & Dufour, M., The Epidemiology of Alcohol Problems in Small Geographic Areas (NIAAA Research Monograph No. 36), Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.

Wickens, T. (1993) Quantitative methods for estimating the size of a drug-using population. Journal of Drug Issues, 23(2), 185-216.

Wish, E., Gray, T., Sushinsky, J., Yacoubian, G., & Fitzgerald, N. (2000) An experiment to enhance the reporting of drug use by arrestees. Journal of Drug Issues, 30(1), 55-76.

Wish, E., Hoffman, J., & Nemes, S. (1997) The validity of self-reports of drug use at treatment admission and at followup: comparisons with urinalysis and hair assays. In L. Harrison & A. Hughes, The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates, NIDA Research Monograph, Number 167. Rockville, MD: National Institute on Drug Abuse. Available at http://165.112.78.61/pdf/monographs/monograph167/download167.html

U. S. Census Bureau (2001) Census 2000. http://www.census.gov/main/www/cen2000.html (and for LA County, Census 2000 data at http://www.losangelesalmanac.com/topics/Population/po09m.htm)

APPENDIX 1: GLOSSARY

ADAM	The NIJ sponsored national Arrestee Drug Abuse Monitoring program, formerly known as the Drug Use Forecasting (DUF) program.
Age group	For the reported analyses, respondents were divided into three age groups: 18 – 24, 25 – 34, and 35 or older.
Anchor points	Survey sites comprising a calibrator sample for synthetic estimation (also see Calibrator and Synthetic estimation).
Arrestee	When referring to ADAM data, the term "arrestee" is used in this report for "booked arrestee" as defined in the sampling design for ADAM (e.g. Hunt & Rhodes, 2001b). For prevalence estimates, the term refers to any individual whose offense would be reported in the UCR.
CAL-DUF	California's Drug Use Forecasting survey conducted in 13 California counties during 1994 through 1996.
Calibrator	In the synthetic estimation techniques, a small survey sample acts as a calibrator providing estimated rate of drug use or treatment need to be used to project the prevalence in a large population. (also see synthetic estimation.)
CEWG	Community Epidemiology Work Group, conducted twice a year by NIDA.
DARC	The UCLA Drug Abuse Research Center, now called the Integrated Substance Abuse Programs (ISAP).
DAWN	Drug Abuse Warning Network, a national surveillance system that collects data on drug-related emergency department visits and deaths.
Dependent variable	In this study's logistic regression model, it is the urine-test positive rate as an indicator of drug use. (Also see Logistic regression model.)
DUF	The Drug Use Forecasting program (precursor project to ADAM).

Illicit drugs	Illicit drugs considered in this study included opiates, cocaine, PCP, marijuana, amphetamines (methamphetamine), non-prescription use of methadone, propoxyphene (Darvon), barbiturates, methaqualone, and benzodiazepine. A composite measure of "any illicit drug use" has referred to use of any of these ten types of drugs.
Independent variables	In this study's logistic regression models, independent variables are the socioeconomic indicators. (Also see Socioeconomic indicators.)
ISAP	UCLA Integrated Substance Abuse Programs.
Logistic regression model	A statistical method using a group of measures to predict the probability of occurrence of a specific event. (See *Technical Note 2*.)
MACR	Monthly Arrest and Citation Register, the arrest database maintained by the California Department of Justice, used in this study to calculate multiple capture rates.
MAPE	Mean absolute percent error, measuring discrepancy between observed and predicted values. (See *Technical Note 9*.)
Multiple capture rate	The proportion of arrests that represent duplicate counts of individuals within the specified time period, that is, rearrests (i.e. second, third, etc. arrests for individuals); used for translating counts of arrests to counts of arrestees. (Also see "rearrest" rates.)
NHSDA	The SAMHSA-funded National Household Survey on Drug Abuse.
NIDA	The National Institute on Drug Abuse.
NIJ	The National Institute of Justice.
ONDCP	The White House Office of National Drug Control Policy.
Prevalence estimate	A general term used by drug researchers to describe the projected numbers of drug users, or users who need or utilize treatment services. In this study, prevalence refers to numbers of drug-using arrestees (more specifically, arrestees with recent drug use who would test positive with urinalysis on at least one arrest occasion during the year 2000 if urine tests were applied to all arrests).

Rearrest rate	(See "multiple capture" rate.)
SACPA	The California Substance Abuse and Crime Prevention Act of 2000, through which certain drug-related offenders can be diverted to drug treatment instead of prison.
SAMHSA	Substance Abuse and Mental Health Services Administration.
Socioeconomic indicators	In this study, measures of socioeconomic status, hypothesized to be related to drug use and used in the estimation of drug use prevalence. This study used five indicators: overall population size, poverty, unemployment, education, and the proportion of youth in the population. (The term "social indicators" is also sometimes used.)
Stratification variables	Characteristics by which the sample is categorized into subgroups. Calibration and projection is done within these subgroups. In this study, the stratification variables were gender, age group, and offense category.
Synthetic Estimation	An estimation approach that determines rates of drug use or need for drug treatment in one sample or population and projects these rates to another similarly categorized populations.
UCR	The FBI Uniform Crime Reporting system.
Urinalysis	The use of a urine test to detect the presence of any of the ten types (in this study) of illicit drugs. (See illicit drug for the types of drugs tested.)

APPENDIX 2: TECHNICAL NOTES

1. The National Household Survey on Drug Abuse (NHSDA) data through the University of Michigan ICPSR website http://www.icpsr.umich.edu (on 1997 data available in Nov. 2001) were used to calculate the rate among the adult household population for the Pacific (census division) region of those reporting one or more arrests in the 12 months preceding the interview. This rate was then applied to the California adult population to project the number of arrestees in the state.

2. For this application, the model is formulated for grouped (by ADAM site) data. We assume that the rate (or probability) of drug use p is distributed as a binomial (p_h, n_h) where n_h is the sample size for ADAM sites $h=1,2,\ldots$ And the model for each gender i by age group j by offense category k subgroup can be written as

   $$\log_e \{p_h/(1-p_h)\} = X_h\beta + \varepsilon_h$$

   where X_h contains the census data for ADAM site h, β contains the intercept and regression coefficients, and ε is the error (c.f. Suciu et al., 2001; Demaris, 1992). The model assumes independence of observations; while the sampling may not provide strict independence, intraclass correlations (across sites within gender by age by offense subgroups) are very low (e.g. <.06) and not significant. However, exploration also included estimation using a model that could account for such possible correlations (see section *Evaluation of Estimation*).

3. Because the purpose of this calibration is prediction rather than hypothesis testing, we have used liberal probability levels to indicate "acceptable" fit and utility of individual predictors. Other criteria, such as replicability of ADAM site observed rates, have also been used to assess the calibration models (see the section *Evaluation of Estimation*).

4. This description of the relationship is only in general terms. The model is actually relating the values of the predictors to the log of the logit (see Note 2 above).

5. Strict application of guidelines for significant digits in calculations would suggest that interpretation of estimates should be limited to 2 digits, thus suggesting rounding of the national estimate to 6,400,000 or 6.4 million. Such guidelines, however, are not strictly enforced in the statistical literature. In addition, we feel that quantities of zeroes can be easily misinterpreted, and translating quantities partially to words (millions, thousands) presents a consistency problem for the report since we have a considerable range of values. Therefore, we have presented estimates rounded only to the nearest integer, but recommend caution in interpreting small differences within large values.

6. While the NHSDA has found in the U.S. as a whole that rural and "less urban" non-metropolitan areas report slightly lower rates of illicit drug use, such counties include a very small proportion of the U. S. population. However, this finding may be geographically specific and may not be applicable to the arrestee population. For example, a treatment needs assessment in California found higher rates of drug use for

the most rural area of the state than for some metropolitan areas, for both household and arrestee populations (Brecht et al., 2002; Ebener et al., 2002; Shen et al., 2002).

7. Calibration values for the ADAM counties cover approximately the same range as do state values for percent poverty (ADAM county range 7.3-31.2% vs. state range 6.5-19.9%), unemployment (2.0-7.0% vs. 2.2-6.6%), young (7.2-12.9% vs 8.1-14.2%), and high school graduates (53.0-90.6% vs. 72.9-88.3%). Only seven states have populations outside the range of ADAM county population.

8. Differences in the multiple capture rate, used to translate numbers of arrests to numbers of arrestees, will affect the prevalence estimates. While such rates may differ from state to state, relevant data are not easily or consistently available for each state for the gender by age group by offense categories used in this analysis. Recidivism (or rearrest) rates are more readily available for parolees or drug court or other special program participants; however, these may not be representative of broader classes of arrestees and thus not optimal for deduplicating the arrest count from the UCR data.

   While ADAM data contain self-reported numbers of arrests that could be used to calculate multiple capture rates, the current study did not adopt these rates as adjustors for the following reasons. ADAM data represent only arrestees who have used drugs; other studies have suggested that rearrest rates may be higher for drug user arrestees than for other arrestees (e.g. Rhodes, 2002). Using an artificially high multiple capture rate would cause our prevalence estimates to be too low. In addition, there may be bias from the self-report nature of numbers of prior arrests.

   Thus, for this project, multiple capture rates calculated from California arrest records were applied to all states. As an illustration of the sensitivity of the estimation approach to differences in the multiple capture rates, the estimation approach was applied using rates modified to be 10%, 20%, and 30% of the rate below and above the California rate for the subgroup (i.e. rate + .10*rate, rate + .20*rate, etc.). The adjusted prevalence estimates for the U.S. for any illicit drug are the following:

Scenario	Prevalence Estimate
Estimate using CA rates	6.4 million
CA rates -.30*rate	7.0 million
CA rates -.20*rate	6.8 million
CA rates -.10*rate	6.6 million
CA rates +.10*rate	6.2 million
CA rates +.20*rate	6.0 million
CA rates +.30*rate	5.7 million

While it was beyond the scope of this study to do a comprehensive exploration of multiple capture rates, we recommend that in future applications of the methodology rates be specialized to geographic areas of interest wherever such specialized data are available.

9. The mean absolute percent error (MAPE) is the average (over ADAM sites) of the absolute value of the percentage discrepancy between observed and predicted prevalence:

$$(\sum \text{ across sites } (|\text{Observed} - \text{Predicted}|/\text{Observed}) * 100) / \text{number of sites}$$

APPENDIX 3: SOCIOECONOMIC INDICATORS FOR ADAM COUNTIES IN 2000[a,b]

ID[0]	ADAM SITE	COUNTY/STATE	POP[1]	POV2[2]	UNEMP[3]	YOUNG[4]	HSGRAD[5]
1	New York City	Manhattan_Borough/NY	1537	20.0	4.9	10.2	78.7
3	Portland	Multnomah/OR	660	12.7	4.3	10.3	85.6
4	San Diego	San Diego/CA	2814	12.4	3.0	11.3	82.6
5	Indianapolis	Marion/IN	860	11.4	2.9	10.0	81.6
6	Houston	Harris/TX	3401	15.0	4.3	10.3	74.6
7	Ft. Lauderdale	Broward/FL	1623	11.5	3.7	7.2	82.0
8	Detroit	Wayne/MI	2061	16.4	3.9	8.7	77.0
9	New Orleans	Orleans Parish/LA	485	27.9	5.7	11.4	74.7
10	Phoenix	Maricopa/AZ	3072	11.7	2.6	10.2	82.5
11	Chicago	Cook/IL	5377	13.5	4.7	9.9	77.7
12	Los Angeles	Los Angeles/CA	9519	17.9	5.4	10.3	69.9
13	Dallas	Dallas/TX	2219	13.4	3.5	10.7	75.0
14	Birmingham	Jefferson/AL	662	14.8	3.5	9.6	80.9
15	Omaha	Douglas/NE	464	9.8	3.1	10.3	87.3
16	Philadelphia	Philadelphia/PA	1518	22.9	6.1	11.1	71.2
17	Miami	Miami-Dade/FL	2253	18.0	5.3	9.1	67.9
18	Cleveland	Cuyahoga/OH	1394	13.1	4.6	8.0	81.6
19	San Antonio	Bexar/TX	1393	15.9	3.5	10.7	76.9
22	San Jose	Santa Clara/CA	1683	7.5	2.0	9.3	83.4
23	Denver	Denver/CO	555	14.3	3.0	10.7	78.9
24M	Atlanta (males)	Fulton/GA	816	15.7	3.7	11.0	84.0
24F	Atlanta (females)	Fulton & Dekalb/GA	1482	13.5	3.7	11.0	84.5
25	Albuquerque	Bernalillo/NM	557	13.7	3.2	10.3	84.4
26	Minneapolis	Hennepin/MN	1116	8.3	2.6	9.7	90.6
27	Sacramento	Sacramento/CA	1223	14.1	4.2	9.5	83.3
28	Tucson	Pima/AZ	844	14.7	2.8	10.9	83.4
29	Anchorage	Anchorage Borough/AK	260	7.3	4.7	9.6	90.3
30	Des Moines	Polk/IA	375	7.9	2.0	9.4	88.3
31	Laredo	Webb/TX	193	31.2	7.0	11.4	53.0
32	Las Vegas	Clark/NV	1376	10.8	4.1	9.2	79.5
33	Oklahoma City	Oklahoma/OK	660	15.3	2.6	10.9	82.5
34	Salt Lake City	Salt Lake/UT	898	8.0	3.0	12.9	86.8
35	Seattle	King/WA	1737	8.4	3.6	9.3	90.3
36	Spokane	Spokane/WA	418	12.3	5.6	10.6	89.1
37	Honolulu	Honolulu/HI	876	9.9	3.8	10.1	84.8
90	Albany	Capital Areas[c]/NY	594	10.4	3.3	10.2	85.6
91	Charlotte-Metro	Charlotte MSA[d]/NC	1499	9.3	2.5	9.1	80.5

Variables
(Note that for purposes of the calibration analysis, above values for variable 1 were divided by 1000 and variables 2-5 were divided by 10.)
0 ID= ADAM identifier (note that ADAM ID=24 has two entries for census data because of slightly different regional coverage, one for males, one for females)
1 POP=population in thousands
2 POV=% below poverty level
3 UNEMP=% unemployed (100-% of population 16+ years in labor force)
4 YOUNG=% population 18-24 years
5 HSGRAD=% high school graduate (from Educational Attainment of the Population 25+ Years)

Notes:
[a] Source: Variables 1, 2, 4, and 5 are from Census 2000; Variable 3 is from U.S. Bureau of Labor Statistics
[b] Year: Variables 1, 3, 4, and 5 are for Year 2000; variable 2 is for 1999
[c] The capital areas are composed of Albany, Rensselaer, and Schenectady, New York.
[d] Charlotte MSA includes Cabarrus, Gaston, Lincoln, Mecklenberg, Rowan, Union, Young Counties North Carolina and York County South Carolina

APPENDIX 4: CODING OF OFFENSES INTO FOUR CATEGORIES

ADAM Coding

INITIAL CODE	DESCRIPTION	RECODED VALUE	DESCRIPTION
1.01	AGGRAVATED ASSAULT	1	Violent
1.02	BLACKMAIL/EXTORTION/THREAT	1	Violent
1.03	KIDNAPPING	1	Violent
1.04	MANSLAUGHTER - NEGLIGENT	1	Violent
1.05	MURDER/HOMICIDE	1	Violent
1.06	ROBBERY	1	Violent
1.07	SEXUAL ASSAULT/RAPE	1	Violent
1.08	WEAPONS	1	Violent
1.09	DOMESTIC VIOLENCE	1	Violent
1.10	CHILD ABUSE	1	Violent
1.11	SPOUSE/PARTNER ABUSE	1	Violent
1.12	OFFENSE AGAINST FAMILY/CHILD	1	Violent
1.13	VIOLATION PROTECT ORDER	1	Violent
1.14	OTHER ASSAULT	1	Violent
1.15	OTHER CRIME AGAINST PERSONS	1	Violent
2.01	DWI/DUI	4	Other
2.02	DRUG POSSESSION	2	Drug Related
2.03	DRUG SALE	2	Drug Related
2.04	LIQUOR	4	Other
2.05	POSSESSION OF ALCOHOL	4	Other
2.06	UNDER INFLUENCE OF SUBSTANCE	2	Drug Related
2.07	OTHER DRUG OFFENSE	2	Drug Related
3.01	ARSON	3	Property
3.02	BRIBERY	3	Property
3.03	BURGLARY	3	Property
3.04	BURGLARY TOOLS	3	Property
3.05	DAMAGE/DESTROY PROPERTY	3	Property
3.06	FORGERY	3	Property
3.07	FRAUD	3	Property
3.08	LARCENY/THEFT	3	Property
3.09	STOLEN PROPERTY	3	Property
3.10	STOLEN VEHICLE	3	Property
3.11	TRESPASSING	3	Property
5.01	PROSTITUTION	4	Other
5.02	EMBEZZLEMENT	4	Other
5.03	FARE BEATING	4	Other
5.04	FLIGHT/ESCAPE	4	Other
5.05	GAMBLING	4	Other
5.06	OBSCENITY	4	Other
5.07	OBSTRUCTION OF JUSTICE	4	Other

5.08	OTHER	4	Other
5.09	PUBLIC PEACE/DISTURBANCE	4	Other
5.10	PICKPOCKET	4	Other
5.11	SEX OFFENSE	4	Other
5.12	PROBATION/PAROLE VIOLATION	4	Other
5.13	TECH. VIOLATION	4	Other
5.14	TRAFFIC-RELATED	4	Other
5.15	CONT. DELINQUENCY OF A MINOR	4	Other
5.16	UNSPECIFIED WARRANT	4	Other
5.17	SALES NO LICENSE	4	Other
7.01	PC_AGG. ASSAULT	1	Violent
7.02	PC_BLACKMAIL/EXT.	1	Violent
7.03	PC_KIDNAPPING	1	Violent
7.04	PC_NEG.MANSLAUGTER	1	Violent
7.05	PC_MURDER/HOMICIDE	1	Violent
7.06	PC_ROBBERY	1	Violent
7.07	PC_SEXUAL ASSAULT	1	Violent
7.08	PC_WEAPONS	1	Violent
7.09	PC_DOMESTIC VIOLENCE	1	Violent
7.10	PC_CHILD ABUSE	1	Violent
7.11	PC_SPOUSE/PARTNER ABUSE	1	Violent
7.12	PC_OFFENSE AGAINST FAMILY/CHILDREN	1	Violent
7.13	PC_PROTECTION ORDER VIOLATION	1	Violent
7.14	PC_OTHER ASSAULT	1	Violent
7.15	PC_OTHER CRIMES AGAINST PERSONS	1	Violent
7.16	PC_DWI/DUI	4	Other
7.17	PC_DRUG POSSESSION	2	Drug Related
7.18	PC_DRUG SALE	2	Drug Related
7.19	PC_LIQUOR	4	Other
7.20	PC_POSSESS ALCOHOL	4	Other
7.21	PC_UNDER INFLUENCE OF CONT.SUB.	2	Drug Related
7.22	PC_OTHER DRUG OFF.	2	Drug Related
7.23	PC_PC_ARSON	3	Property
7.24	PC_BRIBERY	3	Property
7.25	PC_BURGLARY	3	Property
7.26	PC_BURGLARY TOOLS	3	Property
7.27	PC_DAMAGE PROPERTY	3	Property
7.28	PC_FORGERY	3	Property
7.29	PC_FRAUD	3	Property
7.30	PC_LARCENY/THEFT	3	Property
7.31	PC_STOLEN PROPERTY	3	Property
7.32	PC_STOLEN VEHICLE	3	Property
7.33	PC_TRESPASSING	3	Property
7.34	PC_PROSTITUTION	4	Other
7.35	PC_EMBEZZLEMENT	4	Other
7.36	PC_FARE BEATING	4	Other

7.37	PC_FLIGHT/ESCAPE	4	Other
7.38	PC_GAMBING	4	Other
7.39	PC_OBSCENITY	4	Other
7.40	PC_OBSTRUCTION OF JUSTICE	4	Other
7.41	PC_OTHER	4	Other
7.42	PC_PUBLIC PEACE/DESTURBANCE	4	Other
7.43	PC_PICKPOCKET	4	Other
7.44	PC_SEX OFFENSE	4	Other
7.45	PC_PROBATION/PAROLE VIOLATION	4	Other
7.46	PC_TECH. VIOLATION	4	Other
7.47	PC_TRAFFIC-RELATED	4	Other
7.48	PC_UNSPECIFIED WARRANT	4	Other
7.49	PC_IMMIGRATION	4	Other
7.50	PC_INCITING A RIOT	4	Other
8.01	FEDERAL VIOLATION	4	Other
8.02	ILLEGAL ENTRY INTOUS	4	Other

FBI UCR Coding

INITIAL CODE	DESCRIPTION	RECODED VALUE	DESCRIPTION
11	MURDER and NON-NEGLIGENT MANSLAUGHTER	1	Violent
12	MANSLAUGHTER BY NEGLIGENCE	1	Violent
20	FORCIBLE RAPE	1	Violent
30	ROBBERY	1	Violent
40	AGGRAVATED ASSAULT	1	Violent
50	BURGLARY - BREAKING OR ENTERING	3	Property
60	LARCENY - THEFT (EXCEPT MOTOR VEHICLE)	3	Property
70	MOTOR VEHICLE THEFT	3	Property
80	OTHER ASSAULTS	1	Violent
90	ARSON	3	Property
100	FORGERY and COUNTERFEITING	3	Property
110	FRAUD	3	Property
120	EMBEZZLEMENT	4	Other
130	STOLEN PROPERTY - BUYING, RECEIVING, POSS.	3	Property
140	VANDALISM	3	Property
150	WEAPONS - CARRING, POSSESSING, etc.	1	Violent
160	PROSTITUTION AND COMMERCIALIZED VICE	4	Other
170	SEX OFFENSES	4	Other
18	DRUG ABUSE VIOLATIONS (TOTAL)	2	Drug Related
180	SALE/MANUFACTURING (SUBTOTAL)	2	Drug Related
181	OPIUM, COCAINE AND DERIVATIVES	2	Drug Related
182	MARIJUANA	2	Drug Related
183	SYNTHETIC NARCARTICS--ADDICTIVE (e.g. DEMEROL, METHADONES)	2	Drug Related
184	OTHER DANGEROUS NON-NARC DRUGS (BARBITURATES, BENZEDRINE)	2	Drug Related
185	POSSESSION (SUBTOTAL)	2	Drug Related
186	OPIUM, COCAINE AND DERIVATIVES	2	Drug Related
187	MARIJUANA	2	Drug Related
188	SYNTHETIC NARCOTICS--ADDICTIVE (e.g. DEMEROL, METHADONES)	2	Drug Related
189	OTHER DANGEROUS NON-NARC DRUGS (BARBITURATES, BENZEDRINE)	2	Drug Related
190	GAMBLING (TOTAL)	4	Other
191	BOOKMAKING (HORSE and SPORT BOOK)	4	Other
192	NUMBER AND LOTTERY	4	Other
193	ALL OTHER GAMBLING	4	Other
200	OFFENSES AGAINST FAMILY AND CHILDREN	1	Violent
210	DRIVING UNDER the INFLUENCE	4	Other
220	LIQUOR LAWS	4	Other
230	DRUNKENNESS	4	Other
240	DISORDERLY CONDUCT	4	Other

250	VAGRANCY	4	Other
260	ALL OTHER OFFENSES (EXCEPT TRAFFIC)	4	Other
280	CURFEW AND LOITERING LAW VIOLATIONS	4	Other
290	RUNAWAYS	4	Other

APPENDIX 5: STATISTICAL BACKGROUND FOR CONFIDENCE INTERVALS

[This appendix has been adapted to the current application from technical notes by J. de Leeuw and C. Peng, 10/19/1996 from an earlier prevalence estimation study by Hser et al., 1998].

Suppose $\hat{\beta}_{ijk}$ is the vector of regression coefficient estimates in the logistic regression in group *(i, j, k)*, where *(i, j, k)* indexes the 24 groups defined by gender *i*, age group *j*, and offense category *k*. The estimate variance-covariance matrices of the β_{ijk} are V_{ijk}.

We then estimate predicted arrest rates for geographic unit *h* in group (*i, j, k*) by

$$\hat{\pi}_{ijkh} = \frac{\exp x'_{ijkh}\hat{\beta}_{ijk}}{1+\exp x'_{ijkh}\hat{\beta}_{ijk}},$$

and transform them to arrest numbers by multiplying by the weights w_h. Then we take the total of these numbers over the geographic units (e.g. 50 states). Thus we can write for the estimated number of arrests in group *(i, j, k)*

$$\hat{n}_{ijk} = \sum_h w_h \hat{\pi}_{ijkh} .$$

If we linearize the $\hat{\pi}_{ijkh}$ around the true value π_{ijkh}, we find

$$\hat{\pi}_{ijkh} \approx \pi_{ijkh} + \sum_{s=1}^{S} \frac{\partial \pi_{ijkh}}{\partial \beta_{ijks}} (\hat{\beta}_{ijks} - \beta_{ijks}),$$

where *s* is the index for the regression coefficient, i.e. there are $s = 0,\dots, S$ predictors (here $S = 5$ to include intercept and 5 predictors). Now

$$\frac{\partial \pi_{ijkh}}{\partial \beta_{ijks}} = \pi_{ijkh}\ (1 - \pi_{ijkh})\ x_{ijkhs} .$$

Collect the quantities $\pi_{ijkh}(1-\pi_{ijkh})$ in the diagonal matrix D_{ijk}. It follows that the sampling variance of $\hat{n}_{ijk}$ is given by

$$\mathrm{V}\,(\hat{n}_{ijk}) = w'\, D_{ijk} X_{ijk} V_{ijk} X'_{ijk} D_{ijk}\, w,$$

and the standard error is the square root of this quantity. This we can estimate simply by substituting the sample quantities $\hat{D}_{ijk}$ and $\hat{V}_{ijk}$ for D_{ijk} and V_{ijk}.

It could be suggested that smaller standard errors (i.e. more precision) would have been found if we had only used significant logistic regression coefficients, or even do a stepwise search. The fewer coefficients we have to fit within the groups will be more precisely determined. But remember that this precision could very well be spurious. Actually, standard errors will be very difficult to determine in these cases, because the actual predictors that are selected for each groups are themselves a random variable, i.e. the predictor selection process itself will add instability.

Similar calculations can be applied to obtain standard errors for marginal prevalence estimates (subgroup totals). Here we compute $\hat{n}_{ih}$, the estimated number of users of gender i in geographic unit h, summed over age group and offense. The standard error of this quantity is

$$\text{s.e.}\,(\hat{n}_{ih}) = w_h \sqrt{\sum_{j=1}^{J}\sum_{k=1}^{K} \pi_{ijkh}^2 \left(1-\pi_{ijkh}\right)^2 x'_{ijkh} V_{ijk} x_{ijkh}},$$

where $j = 1,\ldots, J$ is age group and $k = 1,\ldots, K$ is offense category. Thus in this study $J = 3$ and $K = 4$.

APPENDIX 6: LOGISTIC REGRESSION RESULTS FOR ANY ILLICIT DRUG USE, STEP 1 OF ESTIMATION (CALIBRATION)

Subgroup			Coefficients						Model Fit	
Gender[1]	Age[2]	Offense[3]	Intercept	POP[4]	POV[5]	UNEMP[6]	YOUNG[7]	HSGED[8]	ChiSquare[9]	Prob.
1	1	1	-5.564**	0.058	1.126**	0.026	-0.810	0.649**	24.240	<0.001
1	1	2	-2.39	-0.062	1.021**	-1.220	-1.523*	0.610**	16.720	0.005
1	1	3	0.005	-0.036	0.322	0.050	-0.826	0.149	4.692	0.455
1	1	4	-1.654	-0.021	0.596**	-0.994*	-1.458**	0.404**	21.382	<0.001
1	2	1	-3.417*	-0.009	0.866**	-1.483*	-0.041	0.367*	15.359	0.009
1	2	2	-7.744**	0.320*	1.054**	-0.335	-1.246	1.124**	46.299	<0.001
1	2	3	-4.252*	0.058	0.720*	0.429	-0.624	0.553**	10.435	0.064
1	2	4	-1.13	-0.054	0.822**	-1.871**	-1.576**	0.342**	29.140	<0.001
1	3	1	-0.782	-0.090	0.657**	-0.726	-1.396*	0.219	13.502	0.019
1	3	2	-3.679	0.279**	1.434**	-2.260*	-2.305**	0.808**	28.334	<0.001
1	3	3	-4.698**	0.078	1.042**	0.452	-0.562	0.553**	24.265	<0.001
1	3	4	-4.004**	0.055	0.837**	-0.893*	-0.598	0.484**	32.294	<0.001
2	1	1	3.154	0.023	-0.153	-2.046	0.134	-0.301	3.026	0.696
2	1	2	-3.058	0.002	1.239	-4.369	-4.465*	1.183*	27.015	<0.001
2	1	3	-4.767	0.061	0.687	-2.673*	-0.911	0.710*	9.856	0.079
2	1	4	-4.116	0.173*	0.856*	-2.565*	-1.231	0.642*	9.614	0.087
2	2	1	0.172	0.044	1.123*	-3.158*	-0.555	-0.012	10.265	0.068
2	2	2	-2.048	0.047	-0.634	2.791	2.381	0.194	2.838	0.725
2	2	3	-9.861**	-0.094	-0.331	5.929**	1.114	0.968**	20.477	0.001
2	2	4	-3.525	0.148*	-0.309	0.715	0.995	0.401	13.908	0.016
2	3	1	-5.899	-0.006	1.160*	-2.968	-1.829	0.925*	8.711	0.121
2	3	2	-8.287	0.341**	0.690	0.244	-2.241	1.358*	24.436	<0.001
2	3	3	-13.243**	0.140*	1.563**	1.522	-0.169	1.375**	17.519	0.004
2	3	4	6.675*	0.061	-0.009	-2.390*	0.505	-0.690*	18.336	0.003

[1] Gender: 1=male, 2=female

[2] Age: 1=18-24, 2=25-34, 3=35 & older

[3] Offense: 1=Violent, 2=Drug-related, 3=Property, 4=Other

[4] POP=(population in thousands)/1000 (i.e. population in millions)

[5] POV=% below poverty level (POV=POV/10 for estimation)

[6] UNEMP=% unemployed (100 - % of population 16+ years in labor force) (UNEMP=UNEMP/10 for est.)

[7] YOUNG=% population 18-24 years (YOUNG=YOUNG/10 for est.)

[8] HSGRAD=% high school graduate (from Educational Attainment of the Population 25+ Years) (HSGRAD=HSGRAD/10 for est.)

[9] from Score test, df=5

** p<.01, * p<.10

APPENDIX 7: ADDITIONAL EXPLORATION CALIBRATION MODEL

Weighted ADAM Data

One exploration of alternate calibration approaches used weighted data from ADAM, which produced urine positive rates for males (Table 7-A) that differed only slightly from rates based on unweighted data.

Table 7-A: Percentages of Positive Urinalysis in 2000 ADAM Calibration Data for Any Illicit Drug, by Gender, Age, and Offense Category, Using Weighted ADAM Data

Age/Offense	Rate
Age 18-24	**69.3**
Violent	62.4
Drug-related	86.4
Property	69.5
Other	64.1
Age 25-34	**64.7**
Violent	53.9
Drug-related	87.7
Property	68.8
Other	59.6
Age 35+	**64.2**
Violent	52.6
Drug-related	89.5
Property	72.7
Other	56.7
Total	**65.9**

Two approaches were used in estimation with weighted data. The first applied standard logistic regression, producing a total estimate for illicit drug use of 6.3 million. A second approach used logistic regression with adjustment for sampling weights to give more appropriate standard errors; STATA software was used for this analysis. Results of the second approach are given in Table 7-B and show a similar total prevalence estimate of 6.3 million arrestees with recent use of any illicit drug, but larger standard errors than with unweighted data.

Table 7-B: Prevalence of Recent Users of Any Illicit Drug Among United States Adult Arrestees in 2000, Stratified by Gender, Age, & Offense Category, Using Weighted ADAM Data and Adjustment to Standard Errors[1,2]

Age Group/ Offense Category	95% CI[3] Lower Limit	**Male Prevalence Estimate**	95% CI Upper Limit	95% CI Lower Limit	**Female Prevalence Estimate**	95% CI Upper Limit	95% CI Lower Limit	**Total Prevalence Estimate**	95% CI Upper Limit
Age 18-24	1,056,481	**1,564,051**	2,071,621	402,176	**466,038**	529,901	1,518,518	**2,030,089**	2,541,661
Violent	86,441	**196,991**	307,541	14,838	**44,454**	74,070	126,996	**241,445**	355,893
Drug	177,336	**441,796**	706,256	69,435	**104,570**	139,706	279,582	**546,366**	813,150
Property	59,268	**193,769**	328,271	55,241	**89,792**	124,343	144,693	**283,562**	422,430
Other	334,789	**731,495**	1,128,201	199,418	**227,222**	255,026	561,037	**958,717**	1,356,396
Age 25-34	1,076,195	**1,431,871**	1,787,547	422,336	**471,288**	520,241	1,544,130	**1,903,159**	2,262,188
Violent	110,059	**210,546**	311,032	17,476	**49,487**	81,498	154,571	**260,033**	365,495
Drug	510,241	**517,924**	525,607	112,386	**129,915**	147,444	628,700	**647,839**	666,977
Property	141,518	**188,194**	234,870	33,372	**60,146**	86,920	194,530	**248,340**	302,150
Other	177,316	**515,207**	853,098	213,097	**231,740**	250,384	408,542	**746,947**	1,085,352
Age 35+	1,317,108	**1,762,225**	2,207,341	537,501	**607,353**	677,205	1,919,013	**2,369,578**	2,820,142
Violent	48,268	**172,809**	297,350	8,717	**47,878**	87,038	90,134	**220,687**	351,240
Drug	582,508	**590,047**	597,587	181,521	**183,259**	184,997	765,569	**773,307**	781,044
Property	95,709	**187,457**	279,205	92,163	**108,882**	125,601	203,080	**296,39**	389,598
Other	394,606	**811,911**	1,229,217	211,987	**267,334**	322,681	658,285	**1,079,245**	1,500,205
Total	3,995,085	**4,758,146**	5,521,207	1,438,124	**1,544,680**	1,651,235	5,532,361	**6,302,826**	7,073,290

[1] Calibration analysis (logistic regression) was done using STATA in order to compute standard errors adjusted for sample weighting
[2] Note that subcategory cells may not add exactly to totals because of rounding
[3] CI=confidence interval

Note that weighted results are based on a smaller ADAM sample size, since weights were missing for approximately 4% of those with urine test results.

Comparative Models

While calibration model development was not within the scope of this project, selected comparisons of models were made. First, two alternative forms of the model were estimated: a simple linear model of prevalence rates across sites and a generalized estimating equations approach (GEE), using a logit link function and individual arrestee data. (See Table 7-C for summary of selected results.) The linear model was less satisfactory than the logistic regression approach: only 9 of the 24 subgroup calibration models fit with $p<.15$, the resulting estimate was particularly high, and the MAPE was higher than with logistic regression (5.7%). The GEE model appears a satisfactory alternative, with adequate fit in all 24 subgroup models, resulting in an estimate of 6.4 million and MAPE of 5.4%. The GEE approach is conceptually attractive since it allows within-site correlated observations and would allow the inclusion of individual-level characteristics in calibration, but it adds complexity to the estimation process.

Second, specification in the logistic regression calibration models was modified by dropping selected variables. The same variables were dropped across all 24 subgroup models to maintain consistency. Omitting population, which was infrequently significant, did slightly reduce the overall estimate of prevalence of arrestees with recent use of any illicit drug, and it also slightly decreased replicability of observed site data with MAPE of 5.8% (see Table 7-C). Omitting percent unemployment and percent young, also infrequently significant predictors, had a similar effect. Dropping percent poverty, which was a strong predictor in many of the subgroup models, had a substantial effect, decreasing the estimate to 5.7 million and increasing the MAPE to 6.9%.

Table 7-C: Selected Results of Alternative Calibration Models: Estimated Prevalence of Any Illicit Drug Use for U.S. Arrestees, Mean Absolute Percent Error (MAPE)

Model	Estimate	MAPE
Linear model of prevalence rates	7.6 million	5.7%
GEE, with logit link function, exchangeable correlation matrix	6.4 million	5.4%
Logistic with predictor omitted:		
Population	6.2 million	5.8%
Unemployment and % young	6.4 million	5.6%
Poverty	5.7 million	6.9%

APPENDIX 8: BIASED PREVALENCE ESTIMATES FOR OPIATES AND METHAMPHETAMINE

As explained in the report, the estimation method as applied in this study may be appropriate only when there is non-negligible opportunity for use of a specific drug in all areas, that is, where penetration rates are non-negligible. Because penetration rates for methamphetamine and opiates may not meet this criterion in the year 2000 ADAM calibration data, our prevalence estimates for these drugs may be biased and have not been included in the report text. Standard errors are particularly large for many of the subgroups, and prevalence is interpreted by the authors as overestimated. However, for completeness, we have included the results here.

Estimates

Estimated opiate prevalence among U.S. arrestees is 3,222,992 with 2,606,129 for males and 616,862 for females (see Table 8-A). Again, we see the greatest opiate prevalence estimates in the oldest age group (1,549,639) and substantially decreasing with age (918,212 for age 25-34 and 755,142 for age 18-24). By offense category, the largest estimated opiate prevalence is for the drug-related offense category (1,407,340), followed by "other" offenses (1,169,921), property (406,042), and violent crime (239,688), in that order.

Table 8-A: Prevalence of Recent Users of Opiates Among United States Adult Arrestees in 2000, Stratified by Gender, Age, & Offense Category[1]

Age Group/ Offense Category	95% CI[2] Lower Limit	**Male Prevalence Estimate**	95% CI Upper Limit	95% CI Lower Limit	**Female Prevalence Estimate**	95% CI Upper Limit	95% CI Lower Limit	**Total Prevalence Estimate**	95% CI Upper Limit
Age 18-24	194,611	**673,024**	1,151,436	37,054	**82,118**	127,182	274,611	**755,142**	1,235,672
Violent	0	**30,213**	165,874	0	**991**	2,106	0	**31,204**	166,870
Drug	47,530	**226,261**	404,991	38,432	**62,846**	87,259	108,716	**289,106**	469,497
Property	0	**13,023**	42,791	0	**10,086**	45,355	0	**23,108**	69,261
Other	0	**403,528**	825,006	0	**8,195**	21,964	0	**411,723**	833,426
Age 25-34	344,066	**657,045**	970,025	174,451	**261,167**	347,883	593,442	**918,212**	1,242,982
Violent	0	**66,693**	212,645	0	**7,869**	48,838	0	**74,562**	226,155
Drug	362,021	**391,045**	420,068	86,169	**98,202**	110,235	457,828	**489,247**	520,666
Property	64,731	**116,151**	167,571	12,856	**46,107**	79,357	101,024	**162,258**	223,492
Other	0	**83,157**	353,652	41,233	**108,989**	176,744	0	**192,146**	470,998
Age 35+	814,041	**1,276,060**	1,738,079	200,751	**273,578**	346,404	1,081,914	**1,549,638**	2,017,361
Violent	0	**92,586**	281,443	0	**41,337**	83,245	0	**133,922**	327,373
Drug	456,760	**481,579**	506,397	137,459	**147,408**	157,357	602,249	**628,987**	655,725
Property	136,717	**166,975**	197,234	22,485	**53,701**	84,916	177,202	**220,676**	264,150
Other	115,084	**534,920**	954,757	0	**31,132**	80,871	143,279	**566,052**	988,825
Total	1,871,081	**2,606,129**	3,341,178	494,985	**616,862**	738,740	2,477,908	**3,222,992**	3,968,076

[1] Note that subcategory cells may not add exactly to totals because of rounding
[2] CI=confidence interval

Estimated methamphetamine prevalence among U.S. arrestees is 3,374,158, with 2,926,424 for males and 447,735 for females (see Table 8-B). We see the greatest methamphetamine prevalence estimates in the oldest age group (1,158,152) with only slightly lower numbers in the younger age groups (1,142,925 for age 25-34 and 1,073,081 for age 18-24). By offense category, the largest estimated methamphetamine prevalence is for the "other" offense category (2,148,674), followed by property (533,404), violent crime (365,817), and drug-related (326,262), in that order.

Table 8-B: Prevalence of Recent Users of Methamphetamine Among United States Adult Arrestees in 2000, Stratified by Gender, Age, & Offense Category[1]

Age Group/ Offense Category	95% CI[2] Lower Limit	Male Prevalence Estimate	95% CI Upper Limit	95% CI Lower Limit	Female Prevalence Estimate	95% CI Upper Limit	95% CI Lower Limit	Total Prevalence Estimate	95% CI Upper Limit
Age 18-24	604,368	**959,468**	1,314,567	61,993	**183,457**	304,921	767,626	**1,142,925**	1,518,224
Violent	0	**42,575**	169,992	0	**12,100**	66,887	0	**54,675**	193,371
Drug	0	**106,026**	305,944	0	**17,589**	40,384	0	**123,615**	324,829
Property	51,249	**152,681**	254,112	18,142	**61,824**	105,506	104,067	**214,505**	324,943
Other	414,044	**658,185**	902,327	0	**91,944**	188,506	487,585	**750,130**	1,012,674
Age 25-34	697,706	**939,211**	1,180,715	8,511	**133,870**	259,230	800,979	**1,073,081**	1,345,183
Violent	19,280	**123,121**	226,962	0	**16,076**	58,735	26,935	**139,197**	251,459
Drug	0	**75,143**	179,681	2,287	**14,695**	27,103	0	**89,838**	195,110
Property	125,271	**154,084**	182,896	0	**25,110**	60,949	133,209	**179,194**	225,178
Other	397,698	**586,863**	776,028	0	**77,989**	189,599	445,216	**664,852**	884,488
Age 35+	640,084	**1,027,746**	1,415,407	43,618	**130,407**	217,195	760,895	**1,158,152**	1,555,410
Violent	0	**128,583**	271,371	15,607	**43,363**	71,118	26,485	**171,945**	317,406
Drug	0	**98,920**	280,239	0	**13,890**	29,458	0	**112,809**	294,797
Property	742	**96,342**	191,943	0	**43,363**	112,646	21,639	**139,705**	257,772
Other	407,461	**703,901**	1,000,340	0	**29,792**	71,257	434,367	**733,692**	1,033,018
Total	2,347,890	**2,926,424**	3,504,958	252,797	**447,735**	642,672	2763,665	**3,374,158**	3,984,652

[1] Note that subcategory cells may not add exactly to totals because of rounding
[2] CI=confidence interval

Possible Adjustments

Some possible adjustments for negligible penetration have been mentioned in the text. We applied one possible simple adjustment as an illustration, scaling the prevalence estimates by the proportion of ADAM calibration sites in which use of opiates (or methamphetamine) was non-negligible. These adjustment factors were .56 for opiates (positive urinalysis rates for opiates were 5% or above in 20 of 36 sites) and .44 for methamphetamine (non-negligible use rates in 16 of 36 sites). For opiates, the adjusted prevalence is 1,804,875 (18% of UCR arrestees) and for methamphetamine, 1,484,630 (15% of UCR arrestees). These adjusted figures for opiates and methamphetamine are still higher than overall ADAM rates but now fall into the plausible range. One possible adjustment to explore in future work is expansion of the calibration model to include regional or site-specific penetration rates or other indicator of supply as a predictor. This would require additional supply-side data for both the calibration sample and for the geographic units (e.g. states) used in the projection phase. An alternative post hoc adjustment might be to use penetration rates for each geographic subcomponent (e.g. states for the national estimate) in the projection phase of the estimation, without altering the calibration model. We recommend further work in this area.