

**ICPSR**  
**Inter-university Consortium for**  
**Political and Social Research**

---

Washington, DC, Metropolitan Area Drug Study  
(DC\*MADS), 1991:  
Household and Nonhousehold Populations

---

United States Department of Health and Human Services  
National Institute on Drug Abuse



WASHINGTON, DC, METROPOLITAN AREA DRUG STUDY (DC\*MADS), 1991:  
HOUSEHOLD AND NONHOUSEHOLD POPULATIONS

(ICPSR 2155)

Principal Investigator

United States Department of Health and Human Services  
National Institute on Drug Abuse

Second ICPSR Version  
November 1998

Inter-university Consortium for  
Political and Social Research  
P.O. Box 1248  
Ann Arbor, Michigan 48106



## BIBLIOGRAPHIC CITATION

Publications based on ICPSR data collections should acknowledge those sources by means of bibliographic citations. To ensure that such source attributions are captured for social science bibliographic utilities, citations must appear in footnotes or in the reference section of publications. The bibliographic citation for this data collection is:

U.S. Dept. of Health and Human Services.  
National Institute on Drug Abuse. WASHINGTON,  
DC, METROPOLITAN AREA DRUG STUDY (DC\*MADS),  
1991 [Computer file]. 2nd ICPSR version.  
Research Triangle Park, NC: Research Triangle  
Institute [producer], 1995. Ann Arbor, MI:  
Inter-university Consortium for Political and  
Social Research [distributor], 1998.

## REQUEST FOR INFORMATION ON USE OF ICPSR RESOURCES

To provide funding agencies with essential information about use of archival resources and to facilitate the exchange of information about ICPSR participants' research activities, users of ICPSR data are requested to send to ICPSR bibliographic citations for each completed manuscript or thesis abstract. Please indicate in a cover letter which data were used.

## DATA DISCLAIMER

The original collector of the data, ICPSR, and the relevant funding agency bear no responsibility for uses of this collection or for interpretations or inferences based upon such uses.



## DATA COLLECTION DESCRIPTION

United States Department of Health and Human Services. National Institute on Drug Abuse

WASHINGTON, DC, METROPOLITAN AREA DRUG STUDY (DC\*MADS), 1991: HOUSEHOLD AND NONHOUSEHOLD POPULATIONS (ICPSR 2155)

SUMMARY: This study examines the prevalence of illicit drug, alcohol, and tobacco use among members of household and nonhousehold populations and a combined aggregate population aged 12 and older in the District of Columbia Metropolitan Statistical Area (DC MSA). In addition, selected characteristics of three drug-abusing subgroups in the household and aggregate populations are examined: crack cocaine users, heroin users, and needle users. The study had three methodological objectives: (1) to investigate the effect that combining data from household and nonhousehold populations has on estimates of the prevalence of drug use and number of users, (2) to determine whether the addition of nonhousehold populations allows more detailed demographic analyses to be conducted for specific drug-using behaviors, and (3) to identify important methodological issues when combining and analyzing data from household and nonhousehold populations. Household population data were collected as part of the DC MSA oversample of the NATIONAL HOUSEHOLD SURVEY ON DRUG ABUSE, 1991 (ICPSR 6128). Nonhousehold population data were subsetted from the 1991 DC\*MADS Institutionalized Population Study and WASHINGTON, DC, METROPOLITAN AREA DRUG STUDY (DC\*MADS), 1991: HOMELESS AND TRANSIENT POPULATION (ICPSR 2346). Household survey topics included age at first use as well as lifetime, annual, and past-month usage for the following drug classes: marijuana and hash, cocaine (and crack), hallucinogens, heroin, inhalants, alcohol, tobacco, anabolic steroids, nonmedical use of prescription drugs including psychotherapeutics, and polysubstance use. Respondents were also asked about substance abuse treatment history, problems resulting from use of drugs, perceptions of the risks involved, personal and family income sources and amounts, need for treatment for drug or alcohol use, mental health and access to care, illegal activities and arrests, and needle-sharing. Demographic data for the household population include gender, race, age, ethnicity, marital status, motor vehicle use, educational level, job status, income level, veteran status, and past and current household composition. Topics from the Institutionalized Study and the Homeless and Transient Population Study included history of homelessness, alcohol and drug treatment or counseling, illegal activities and arrests, physical health, pregnancy, mental health, mental health treatment, employment, income and expenditures, living arrangements and population movement, and specific and general drug use. Drugs

covered include tobacco, alcohol, marijuana and hash, inhalants, cocaine and crack, hallucinogens, heroin, stimulants, and tranquilizers. Data also provide information on age at first use, route of administration, polysubstance use, perceived risks, and insurance coverage. Demographic data for the nonhousehold population include gender, age, marital status, race, education, military service, and number of children and other dependents.

UNIVERSE: For the Household Study: the civilian, noninstitutionalized population of the coterminous United States (Alaska and Hawaii excluded) aged 12 and older. For the Institutionalized Study: persons in institutional (e.g., correctional facilities, mental or psychiatric hospitals, noncorrectional facilities for juveniles) and noninstitutional group quarters (e.g., group homes, transitional homes for people leaving treatment for substance abuse). Nursing homes and hospitals or wards providing drug and alcohol treatment were excluded. For the Homeless and Transient Population Study: persons who were either literally homeless or at imminent risk of becoming homeless, including persons who spent the previous night in an emergency shelter, in a nondomicile (e.g., vacant building, city park, car, or on the street), or who were using soup kitchens or emergency food banks.

SAMPLING: The household and nonhousehold data were drawn from three sources: (1) the DC MSA oversample of NATIONAL HOUSEHOLD SURVEY OF DRUG ABUSE, 1991 (ICPSR 6128) (NHSDA), (2) the DC\*MADS 1991 Institutionalized Population Study, and (3) WASHINGTON, DC, METROPOLITAN AREA DRUG STUDY (DC\*MADS), 1991: HOMELESS AND TRANSIENT POPULATION (ICPSR 2346). The NHSDA oversample consisted of interviews with 2,547 residents from a sample of 5,399 households and selected group quarters (e.g., college dormitories, homeless shelters). The Institutionalized Study consisted of 1,203 interviews with residents of 42 institutions stratified into four groups: 888 interviews from 20 correctional institutions, 207 interviews from 6 psychiatric institutions, 55 interviews from 7 noncorrectional institutions for juveniles, and 73 interviews from 9 group homes. The Homeless and Transient Population Study consisted of 908 interviews from four overlapping sampling frames: 477 interviews with residents in 93 shelters, 224 interviews with patrons of 31 soup kitchens and food banks, 143 interviews with "literally homeless" people from 18 major cluster encampments, and 64 interviews with literally homeless people from an area probability sample of 432 census blocks in the MSA. Overlap between the samples may have occurred. The overlap was minimized by eliminating institutions from the DC\*MADS surveys that would increase overlap. For example, students living in dormitories were included in the NHSDA but not the DC\*MADS Institutionalized Study.

For the household, institutional, and homeless studies, the household/institutional response rates were 93.5 percent, 87.5 percent, and 82.6 percent, respectively. The individual interview response rates were 82.1 percent, 89.4 percent, and 86.1 percent. The overall response rates were 76.8 percent, 78.2 percent, and 71 percent.

NOTE: (1) Users should note that the National Institute on Drug Abuse does not make a public use file available for the DC\*MADS Institutionalized Population Study. (2) The codebook is provided as a Portable Document Format (PDF) file. The PDF file format was developed by Adobe Systems Incorporated and may be accessed using PDF reader software, such as the Adobe Acrobat Reader. Information on how to obtain a copy of the Acrobat Reader is provided through the ICPSR and SAMHDA Websites on the Internet.

EXTENT OF COLLECTION: 1 data file + machine-readable documentation (PDF) + SAS data definition statements + SPSS data definition statements

EXTENT OF PROCESSING: CONCHK.PR/ CONCHK.ICPSR/ DDEF.ICPSR/ FREQ.PR/ MDATA.PR/ SCAN/ REFORM.DATA/ UNDOCCHK.ICPSR/ REFORM.DOC

DATA FORMAT: Logical Record Length with SAS and SPSS data definition statements

File Structure: rectangular

Cases: 4,658

Variables: 48

Record Length: 117

Records Per Case: 1

RELATED PUBLICATIONS:

United States Department of Health and Human Services. National Institute on Drug Abuse. "Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the Washington, DC, Metropolitan Area Homeless and Transient Population, 1991" (Technical Report No. 2). Rockville, MD: United States Government Printing Office, 1993.

United States Department of Health and Human Services. National Institute on Drug Abuse. "Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the Washington, DC, Metropolitan Area Household and Nonhousehold Populations, 1991" (Technical Report No. 8). Rockville, MD: United States Government Printing Office, 1994.

United States Department of Health and Human Services. National Institute on Drug Abuse. "Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the Washington, DC,

Metropolitan Area Institutionalized Population, 1991" (Technical Report No. 4). Rockville, MD: United States Government Printing Office, 1994.

\*\*ICPSR Processor Note\*\*  
DC\*MADS 1991:  
HOUSEHOLD AND NONHOUSEHOLD POPULATIONS  
(ICPSR 2155)

- 1) Column locations detailed in the codebook document the original data file. Users should refer to the SAS or SPSS data definition statement files for column locations of variables in the ICPSR data file.



# **WARNING! DATA USE RESTRICTIONS**

## **Read Carefully Before Using**

The Public Health Service Act (Section 301) provides that the data collected by the National Institute on Drug Abuse (NIDA) may be used only for the purpose of statistical reporting and analysis.

Any effort to determine the identity of any reported case is prohibited by this law.

NIDA does all it can to assure that the identity of subjects cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are omitted from the data set. Any intentional identification or disclosure of a person or establishment violates the assurances of confidentiality given to the providers of the information. Therefore, users must agree to:

1. Use the data in this data set for statistical reporting and analysis only.
2. Make no use of the identity of any person or establishment discovered inadvertently and advise the Director of NIDA of any such discovery.
3. Not link this data set with individually identifiable data from other NIDA and non-NIDA data sets.

**By using these data you signify your agreement to comply with the above-stated statutorily based requirements.**



## TABLE OF CONTENTS

<b>Item</b>	<b>Page</b>
Use of DC*MADS Data on Household and Nonhousehold Populations.....	1
Data Disk Characteristics .....	2
SUMMARY OF DATA ITEMS AND LOCATION OF DOCUMENTATION.....	3
CODEBOOK SECTIONS:	
Demographic Characteristics.....	4
Drug Use Variables .....	4
Sample Design Variables .....	8
Appendix A: Summary of DC*MADS Study of Household and Nonhousehold Populations ..	A-1
Appendix B: Procedures for Combining Homeless, Institutionalized, and Household Data...	B-1
Appendix C: Other Methodological Notes.....	C-1
Appendix D: References .....	D-1
Appendix E: Alphabetical List of Variables .....	E-1



## **Use of DC\*MADS Data on Household and Nonhousehold Populations**

The National Institute on Drug Abuse (NIDA) sponsored the Study of Household and Nonhousehold Populations as part of the Washington, D.C. Metropolitan Area Drug Study (DC\*MADS), in 1991. Information on this study has been published in a technical report, a summary of which is reproduced as Appendix A in this document.

In preparing the public release version of the data file, all data containing geographic detail have been suppressed. Other items, such as race, were aggregated when very few individuals were found in more detailed categories. Information on attributes of the sampled institutions was suppressed to avoid potential disclosure. Some of the data contained in this file may not have been fully edited and will require user caution; these are self-report data which may be over- or under-reported.

The codebook is organized topically, beginning with demographic variables, followed by drug use variables and sample design variables, which include survey weights, stratum, and primary sampling unit variables. As described in Appendix C, the complex nature of this sample should be taken into account in the analysis of this data set. Drug use variables include a series of substance use variables commonly used in analyses of drug abuse data. These and other created variables were derived from compatible items in each of the component surveys. Each codebook item includes a suggested variable name along with the column location, codes and code descriptions, and the frequency distribution (unweighted) for that item. Appendix E provides an alphabetical listing of the suggested variable names with their respective page locations.

NIDA requests the cooperation of users of this data file in observing the following guidelines:

- Any published material derived from these data should acknowledge the National Institute on Drug Abuse as the original source. It should also include a disclaimer which credits any analyses, interpretations, or conclusions reached to the author and not to NIDA, which is responsible only for the initial data.
- Authors should provide NIDA with a reprint of published articles which utilize the DC\*MADS data from the Study of Household and Nonhousehold Populations. Please send reprints to:

DC\*MADS Data Management Staff  
Division of Epidemiology and Prevention Research  
National Institute on Drug Abuse  
Parklawn Building, Room 9A/53  
5600 Fishers Lane  
Rockville, MD 20857

## DATA DISK CHARACTERISTICS

### 1991 DC\*MADS Study of Household and Nonhousehold Populations

---

**Title:** 1991 DC\*MADS Study of Household and Nonhousehold Populations Data

**Data Set Name:** DCMHNNH.ASC (*Disk*)

**Record Length:** 133

**Number of Records:** 4658

**Number of Disks:** 1

**Recording Mode:** ASCII

**Notes:** An ASCII file named READ.ASC is provided with minimum specifications for a SAS read input program, and may be adapted to suit other software packages.

Also provided is an ASCII file named FORMATS.ASC to generate SAS formats consistent with codebook values.

**Contact Persons:** Elizabeth Lambert/Arthur Hughes  
Division of Epidemiology and Prevention Research  
National Institute on Drug Abuse  
Parklawn Building, Room 9A/53  
5600 Fishers Lane  
Rockville, MD 20857  
(301) 443-6543/(301) 443-6637

**SUMMARY OF DATA ITEMS AND LOCATION OF DOCUMENTATION**

<b>COLUMN LOCATION</b>	<b>VARIABLE NAME</b>	<b>VARIABLE LABEL</b>	<b>PAGE</b>
1-4	OBSNUM	Observation Number	4
5-6	CATAGE	Age Category	4
7-8	EDUCAT1C	Categorized and Collapsed Education	4
9-10	EMPSTAT2	Employment Status Recode	4
11-12	IRSEX	Sex - Imputation Revised	4
13-14	MARSTAT	Marital Status - Categorized	4
15-16	RACE	Race/Hispanic - Origin Recode	4
17-18	ALCMON	Alcohol - Past Month Use	4
19-20	ALCYR	Alcohol - Past Year Use	5
21-22	CIGMON	Cigarettes - Past Month Use	5
23-24	CIGYR	Cigarettes - Past Year Use	5
25-26	COCMON	Cocaine - Past Month Use (Includes Crack Use)	5
27-28	COCYR	Cocaine - Past Year Use (Includes Crack Use)	5
29-30	CRKMON	Crack - Past Month Use	5
31-32	CRKYR	Crack - Past Year Use	5
33-34	HALMON	Hallucinogens - Past Month Use	5
35-36	HALYR	Hallucinogens - Past Year Use	5
37-38	HERMON	Heroin - Past Month Use	5
39-40	HERYR	Heroin - Past Year Use	6
41-42	HVYDRK2	Drank Alcohol Heavily - Past 30 Days	6
43-44	IEMMON	Used Illicit Drugs Except Marijuana - Past Month	6
45-46	IEMYR	Used Illicit Drugs Except Marijuana - Past Year	6
47-48	INHMON	Inhalants - Past Month Use	6
49-50	INHYR	Inhalants - Past Year Use	6
51-52	MRJMON	Marijuana - Past Month Use (Includes Hashish)	6
53-54	MRJYR	Marijuana - Past Year Use (Includes Hashish)	6
55-56	NDSLHR	Needle Sharing Indicator	6
57-58	NEDFLAG3	Needle Use (Any Drug) - Lifetime Use	6
59-60	NEDMON	Needle Use (Any Drug) - Past Month Use	7
61-62	NEDYR3	Needle Use (Any Drug) - Past Year Use	7
63-64	PSYMON2	Any (Non-Medical Use) Psychotherapeutics - Past Month Use	7
65-66	PSYYR2	Any (Non-Medical Use) Psychotherapeutics - Past Year Use	7
67-68	STAMON	Sedatives/Tranquilizers/Analgesics - Past Month Use	7
69-70	STAYR	Sedatives/Tranquilizers/Analgesics - Past Year Use	7
71-72	STMMON	Stimulants - Past Month Use	7
73-74	STMYR	Stimulants - Past Year Use	7
75-76	SUMMON	Any Illicit Drug - Past Month Use	7
77-78	SUMYR	Any Illicit Drug - Past Year Use	7
79-80	DRUGOVLP	Overlap of Past Year Crack, Heroin, & Needle Use	8
81-82	OVERLAP	Overlap of Surveys	8
83-84	SURVEY	Survey Type	8
85-86	SURVEY1	DC Household	8
87-88	SURVEY2	Homeless	8
89-90	SURVEY3	Institutionalized	8
91-105	MULTWT	Multiplicity Adjusted Weight for Combined File	8
106-112	STRATUM	Stratum for SUDAAN	8
113-127	WEIGHT	Original Analysis Weights Combined	9
128-133	WPSU	PSU for SUDAAN	9

COLUMN LOCATION	VARIABLE NAME	COUNT	VARIABLE LABEL AND CODE VALUE
1-4	OBSNUM		Observation Number  <i>Demographic Characteristics</i>
5-6	CATAGE		Age Category  2 -9=Missing 755 1=12 to 17 1081 2=18 to 25 1392 3=26 to 34 1428 4=35 or over
7-8	EDUCAT1C		Categorized and Collapsed Education  11 -9=Missing 1319 1=Less than high school 1162 2=High school graduate 1411 3=Some college/college graduate 756 4=12 to 17 years old
9-10	EMPSTAT2		Employment Status Recode  27 -9=Missing 1614 1=Full time 399 2=Part time 655 3=Unemployed 1208 4=Other 755 5 =12 to 17 years old
11-12	IRSEX		Sex - Imputation Revised  2775 1 =Male 1883 2=Female
13-14	MARSTAT		Marital Status - Categorized  31 -9=Missing 2805 1 =Single 959 2=Married 863 3=Widowed, divorced, or separated
15-16	RACE		Race/Hispanic - Origin Recode  14 -9=Missing 1752 1=White 2370 2=Black 301 3=Hispanic 221 4=Other  <i>Drug Use Variables</i>
17-18	ALCMON		Alcohol- Past Month Use  10 -9=Missing 2684 0=Did not use within the past month 1964 1=Used within the past month

COLUMN LOCATION	VARIABLE NAME	COUNT	VARIABLE LABEL AND CODE VALUE
19-20	ALCYR		Alcohol - Past Year Use
		13	-9=Missing
		1482	0=Did not use within the past year
		3163	1=Used within the past year
21-22	CIGMON		Cigarettes - Past Month Use
		36	-9=Missing
		2649	0=Did not use within the past month
		1973	1=Used within the past month
23-24	CIGYR		Cigarettes - Past Year Use
		36	-9=Missing
		2349	0=Did not use within the past year
		2273	1=Used within the past year
25-26	COCMON		Cocaine - Past Month Use (includes crack use)
		4	-9=Missing
		4374	0=Did not use within the past month
		280	1=Used within the past month
27-28	COCYR		Cocaine - Past Year Use (includes crack use)
		4	-9=Missing
		3773	0=Did not use within the past year
		881	1=Used within the past year
29-30	CRKMON		Crack - Past Month Use
		7	-9=Missing
		4422	0=Did not use within the past month
		229	1=Used within the past month
31-32	CRKYR		Crack - Past Year Use
		6	-9=Missing
		3935	0=Did not use within the past year
		717	1=Used within the past year
33-34	HALMON		Hallucinogens - Past Month Use
		17	-9=Missing
		4613	0=Did not use within the past month
		28	1=Used within the past month
35-36	HALYR		Hallucinogens - Past Year Use
		17	-9=Missing
		4424	0=Did not use within the past year
		217	1=Used within the past year
37-38	HERMON		Heroin - Past Month Use
		12	-9=Missing
		4590	0=Did not use within the past month
		56	1=Used within the past month

COLUMN LOCATION	VARIABLE NAME	COUNT	VARIABLE LABEL AND CODE VALUE
39-40	HERYR		Heroin - Past Year Use
		12	-9=Missing
		4400	0=Did not use within the past year
		246	1=Used within the past year
41-42	HVYDRK2		Drank Alcohol Heavily - Past 30 Days (consumed 5 drinks in one sitting at least 5 times in the past 30 days)
		68	-9=Missing
		4247	0=Did not drink heavily within the past 30 days
		343	1=Drank heavily within the past 30 days
43-44	IEMMON		Used Illicit Drugs Except Marijuana - Past Month
		1	-9=Missing
		4282	0=Did not use within the past month
		375	1=Used within the past month
45-46	IEMYR		Used Illicit Drugs Except Marijuana - Past Year
		1	-9=Missing
		3543	0=Did not use within the past year
		1114	1=Used within the past year
47-48	INHMON		Inhalants - Past Month Use
		10	-9=Missing
		4612	0=Did not use within the past month
		36	1=Used within the past month
49-50	INHyr		Inhalants - Past Year Use
		10	-9=Missing
		4525	0=Did not use within the past year
		123	1=Used within the past year
51-52	MRJMON		Marijuana - Past Month Use (includes marijuana and hashish)
		17	-9=Missing
		4372	0=Did not use within the past month
		269	1=Used within the past month
53-54	MRJYR		Marijuana - Past Year Use (includes marijuana and hashish)
		15	-9=Missing
		3787	0=Did not use within the past year
		856	1=Used within the past year
55-56	NDLSHR		Needle Sharing Indicator
		109	-9=Missing
		4285	0= Has never shared a needle
		264	1=Has shared a needle in his/her lifetime
57-58	NEDFLAG3		Needle Use (Any Drug) - Lifetime Use
		102	-9=Missing
		4060	0=Has never used in his/her lifetime
		496	1=Has used in his/her lifetime

COLUMN LOCATION	VARIABLE NAME	COUNT	VARIABLE LABEL AND CODE VALUE
59-60	NEDMON		Needle Use (Any Drug) - Past Month Use
		57	-9=Missing
		4539	0=Did not use within the past month
		62	1=Used within the past month
61-62	NEDYR3		Needle Use (Any Drug) - Past Year Use
		62	-9=Missing
		4384	0=Did not use within the past year
		212	1=Used within the past year
63-64	PSYMON2		Any (non-medical use) Psychotherapeutics - Past Month Use
		4	-9=Missing
		4574	0=Did not use within the past month
		80	1=Used within the past month
65-66	PSYYR2		Any (non-medical use) Psychotherapeutics - Past Year Use
		4	-9=Missing
		4344	0=Did not use within the past year
		310	1=Used within the past year
67-68	STAMON		Sedatives/Tranquilizers/Analgesics - Past Month Use
		32	-9=Missing
		4559	0=Did not use within the past month
		67	1=Used within the past month
69-70	STAYR		Sedatives/Tranquilizers/Analgesics - Past Year Use
		32	-9=Missing
		4371	0=Did not use within the past year
		255	1=Used within the past year
71-72	STMMON		Stimulants - Past Month Use
		17	-9=Missing
		4620	0=Did not use within the past month
		21	1=Used within the past month
73-74	STMYR		Stimulants - Past Year Use
		18	-9=Missing
		4540	0=Did not use within the past year
		100	1=Used within the past year
75-76	SUMMON		Any Illicit Drug - Past Month Use
		1	-9=Missing
		4136	0=Did not use within the past month
		521	1=Used within the past month
77-78	SUMYR		Any Illicit Drug - Past Year Use
		1	-9=Missing
		3287	0=Did not use within the past year
		1370	1=Used within the past year

COLUMN LOCATION	VARIABLE NAME	COUNT	VARIABLE LABEL AND CODE VALUE
79-80	DRUGOVL		Overlap of Past Year Crack, Heroin, & Needle Use - Based on CRKYR (past year crack use), HERYR (past year heroin use), and NEDYR3 (past year needle use).
		70	-9=Missing
		511	1=Crack user only
		23	2=Heroin user only
		25	3=Needle user only
		63	4=Crack and heroin user only
		26	5=Crack and needle user only
		47	6=Heroin and needle user only
		113	7=Crack, heroin and needle user
		3780	8=Crack, heroin and needle nonuser
81-82	OVERLAP		Overlap of Surveys - Based on SURVEY (type of DC*MADS survey), INSSTYPE (type of institution), HOMLIT (literally homeless), ZSAMLOC (sample location), and GQTY (group quarters indicator).
		2542	1=NHSDA only
		1130	2=Institutional only
		349	3=Homeless only
		82	4=Not literally homeless
		73	5=Institutional group homes
		477	6=Homeless, Shelters
		5	7=NHSDA and Institutionalized, NHSDA and Homeless, or Institutionalized and Homeless
			<i>Sample Design Variables</i>
83-84	SURVEY		Survey Type
		2547	1=1991 DC Household
		908	2=Homeless
		1203	3=Institutionalized
85-86	SURVEY1		DC Household
		2111	0=Other
		2547	1=DC Household
87-88	SURVEY2		Homeless
		3750	0=Other
		908	1=Homeless
89-90	SURVEY3		Institutionalized
		3455	0=Other
		1203	1=Institutionalized
91-105	MULTWT		Multiplicity Adjusted Weight for Combined File (IMPLIED DECIMAL POINT, FORMAT 15.8) Used for analyses of aggregate data: see Appendix C.
106-112	STRATUM		Stratum for SUDAAN

<b>COLUMN LOCATION</b>	<b>VARIABLE NAME</b>	<b>COUNT</b>	<b>VARIABLE LABEL AND CODE VALUE</b>
113-127	WEIGHT		Original Analysis Weights Combined (IMPLIED DECIMAL POINT, FORMAT 15.8) Used only for analyses of survey-specific data from the component subpopulations: see Appendix C.
128-133	WPSU		PSU for SUDAAN



## **Appendix A**

### **SUMMARY OF DC\*MADS STUDY OF HOUSEHOLD AND NONHOUSEHOLD POPULATIONS\***

This study examines the prevalence of illicit drug, alcohol, and tobacco use among members of household and nonhousehold populations and a combined "aggregate" population aged 12 and older in the District of Columbia Metropolitan Statistical Area (DC MSA). In addition, it examines selected characteristics of three drug-abusing subgroups in the household and aggregate populations: crack-cocaine users, heroin users, and needle users. Findings are presented in the context of three methodological objectives:

- (a) to examine the effect that combining data from household and nonhousehold populations has on estimates of the *prevalence* of drug use and *numbers* of users;
- (b) to examine whether the addition of nonhousehold populations allows more detailed demographic analyses to be conducted for specific drug-using behaviors (i.e., crack-cocaine, heroin, and needle use); and
- (c) to identify important methodological issues when combining and analyzing data from household and nonhousehold populations.

Household population data were collected as part of the DC MSA oversample of the 1991 National Household Survey on Drug Abuse (NHSDA), and consisted of interviews with 2,547 residents from a sample of 5,399 households and selected group quarters (e.g., college dormitories, homeless shelters). Nonhousehold population data were drawn from the 1991 DC\*MADS Institutionalized Study and the 1991 DC\*MADS Homeless and Transient Population Study. The Institutionalized Study consisted of 1,203 interviews with residents of 42 institutions stratified into four groups: 868 interviews from 20 correctional institutions; 207 interviews from 6 psychiatric institutions; 55 interviews from 7 noncorrectional institutions for juveniles; and 73 interviews from 9 group homes. The Homeless and Transient Population Study consisted of 908 interviews from four overlapping sampling frames: 477 interviews with residents in 93 shelters; 224 interviews with patrons of 31 soup kitchens and food banks; 143 interviews with literally homeless people from 18 major clusters of encampments; and 64 interviews with literally homeless people from an area probability sample of 432 census blocks in the MSA.

For the household, institutionalized, and homeless studies, respectively, the household/institutional response rates were 93.5%, 87.5%, and 82.6%; the individual interview response rates were 82.1%, 89.4%, and 86.1%; and the overall rates were 76.8%, 78.2%, and 75.0%. Data were combined from the household, institutionalized, and homeless populations to produce an aggregate population for the DC MSA based on interviews from 4,658 individuals. Aggregate data were adjusted for potential sampling overlap across the surveys. Estimates of population characteristics are based on data from these three surveys.

---

\*Adapted from the Executive Summary and Chapter 5 of *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the DC Metropolitan Area Household and Nonhousehold Populations: 1991* (National Institute on Drug Abuse, 1994a).

Some of the key findings for the household, institutionalized, and homeless populations include the following:

- An estimated 47.7% of the DC MSA household population were male; 61.9% were white, 27.2% were black, and 5.2% were Hispanic. Only 12.9% of adults in the household population had not completed high school. In contrast, the institutionalized and homeless populations covered by DC\*MADS were predominantly male (90.7% and 75.9%, respectively) and black (69.3% and 75.8% respectively). In addition, an estimated 64.8% of institutionalized and 40.1% of homeless and transient adults had not completed high school.
- Rates of any illicit drug use in the household population were 39.9% in the lifetime, 11.7% in the past year, and 5.7% in the past month. Marijuana was the most commonly used drug for all time periods, having ever been used by 36.2% of the household population, by 8.1% in the past year, and by 4.1% in the past month.
- An estimated 2.2% of the household population had used crack-cocaine during the lifetime, 0.9% during the past year and 0.3% during the past month. Crack-cocaine use in the past year was significantly greater among the household population of the District of Columbia (DC), than among household residents in the Maryland portion of the DC MSA (2.9% vs. 0.3%). Similarly, crack-cocaine use in the past month was significantly greater among DC household residents (1.2%) than among residents of the DC MSA who lived in Maryland or Virginia (0.2% and 0.1%). The prevalence of past month crack-cocaine use was also significantly greater among household residents of low socioeconomic status (SES) areas of the DC MSA (1.1%) compared with that among household residents of other SES areas (0.1%).
- Rates of any alcohol use in the household population were 84.6% in the lifetime, 73.5% in the past year, and 55.9% in the past month. An estimated 4.2% were heavy alcohol users in the past month (i.e., five or more drinks per occasion on 5 or more days in the 30 days). Rates of cigarette smoking were 69.5% in the lifetime, 28.1% in the past year, and 23.1% in the past month.
- Among residents of institutions covered by the DC\*MADS Institutionalized Study, the prevalence of any illicit drug use was 84.4% in the lifetime, 49.9% in the past year, and 8.1% in the past month. The most commonly used drugs among this population in the past year were any form of cocaine (36.9%), marijuana (31.7%), and crack-cocaine (30.4%).
- Among the homeless and transient population, rates of any illicit drug use were 80.0% in the lifetime, 57.7% in the past year, and 34.3% in the past month. For both the past year and past month, any form of cocaine (past year: 48.4%; past month: 27.5%) and crack-cocaine (past year: 44.8%; past month: 25.7%) were the most commonly used illicit drugs. An estimated 27.5% were heavy alcohol users in the past month.

The prevalence of illicit drug use was substantially higher in the two DC\*MADS nonhousehold populations than in the household population. However, direct comparisons of

this type may be misleading because of demographic differences between the populations that have also been shown to be related to substance use. For example, the proportion of males is higher in the nonhousehold populations than in the household population, and males generally have higher rates of illicit drug and heavy alcohol use than do females. More detailed analyses that control for these differences are not provided here, but are needed to make accurate comparisons between household and nonhousehold populations.

Some of the key findings for the DC MSA aggregate population include the following:

- An estimated 99.2% of the 1991 DC MSA aggregate population lived in households compared with an estimated 0.6% who were residents of institutions covered by DC\*MADS and 0.2% who were homeless.
- Estimates of the *percentages* of the DC MSA aggregate population who engaged in illicit drug use in the past year differed little from the corresponding household estimates. For example, past year use of any illicit drug increased by only 0.3%, from 11.7% for the household population to 12.0% for the aggregate.
- Adding nonhousehold population data to household data substantially enlarged estimates of the *numbers* of past year crack-cocaine users and needle users. There were over 9,000 more past year crack-cocaine users in the aggregate population (38,433) than in the household population (29,027) and nearly 3,000 more past year needle users in the aggregate population (8,740) than in the household population (5,987). Stated another way, household estimates alone would have missed approximately one-fourth of the past year crack users and one-third of the past year needle users in the DC MSA aggregate population.
- In the year before the survey, over 50,000 people in the DC MSA aggregate population (53,241) engaged in one or more of the specific drug-using behaviors of crack-cocaine use, heroin use, or needle use.
- There were 15,549 past year heroin users in the DC MSA aggregate population and 8,740 needle users. This finding suggests that a substantial number of heroin users had not injected the drug in the past year. Other possible routes of administration for heroin include smoking or intranasal administration (i.e., "snorting").

Combining data from the household, institutionalized, and homeless and transient populations offers insights about the strengths and limitations of the coverage from the household population on prevalence rates of drug use, estimated numbers of drug users, and drug use behaviors.

The aggregate population consists of the combined data from the DC MSA oversample of the 1991 National Household Survey on Drug Abuse (NHSDA) and the DC\*MADS Institutionalized and Homeless and Transient Population Studies. Information is presented on the methodology for combining data from these three studies to produce the aggregate population data set, including adjustments for potential multiplicity in the sample frames. The aggregate population covered the large majority of the DC MSA population, but did not include

the entire population because some types of institutions were excluded from the Institutionalized Study.

Data provided in *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the DC Metropolitan Area Household and Nonhousehold Populations: 1991* (National Institute on Drug Abuse, 1994a) show what effect adding data from nonhousehold populations had on estimates of prevalence of drug use and numbers of users compared with those obtained from the household data alone. From a methodological standpoint, past year and past month prevalence estimates were of greater importance than lifetime prevalence estimates for addressing questions about the effects of combining data from household and nonhousehold populations. Accordingly, findings presented in the report focus on the past year and past month periods only.

## **A.1 DC MSA Aggregate Population Methodology and Analytical Approach**

### **A.1.1 Data Sources for the DC MSA Aggregate Population**

The estimates presented in *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the DC Metropolitan Area Household and Nonhousehold Populations: 1991* (National Institute on Drug Abuse, 1994a) were based on data combined from three separate sample surveys conducted in the DC MSA during 1991. These surveys are the

- 1991 NHSDA's DC MSA oversample,
- DC\*MADS Institutionalized Study, and
- DC\*MADS Homeless and Transient Population Study,

For each of these DC\*MADS studies, a technical report was developed. Findings specific to the 1991 oversample of households in the DC area are included in DC\*MADS Technical Report #8, *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the DC Metropolitan Area Household and Nonhousehold Populations: 1991* (National Institute on Drug Abuse, 1994a), and the drug answer sheets of the data collection instrument are documented by the Substance Abuse and Mental Health Services Administration [SAMHSA] (1993). Findings on the institutionalized population are reported in DC\*MADS Technical Report #4, *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the DC Metropolitan Area Institutionalized Population: 1991* (National Institute on Drug Abuse, 1994b). This report also includes the study questionnaire. Findings on the homeless and transient population are reported in DC\*MADS Technical Report #2, *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the Washington, DC Metropolitan Area Homeless and Transient Population: 1991* (National Institute on Drug Abuse, 1993), which also includes the study questionnaire.

The subpopulations surveyed in the separate studies were generally defined in terms of where respondents were residing at the time of data collection. The NHSDA surveyed the civilian, noninstitutional population, including civilians living on military bases and persons living in noninstitutional group quarters (e.g., rooming houses, dormitories, shelters for homeless

people and group homes). There were 2,547 respondents in the DC MSA oversample (Substance Abuse and Mental Health Services Administration, 1993).

The DC\*MADS Institutionalized Study surveyed persons in institutional and noninstitutional group quarters. Institutional group quarters included correctional facilities, mental or psychiatric hospitals, and other institutions, such as noncorrectional facilities for juveniles. Noninstitutional group quarters included group homes for people who were mentally retarded, homes for people with physical disabilities, and transitional homes for people leaving treatment for alcohol or other drug abuse. Nursing homes and hospitals or wards providing treatment for alcohol or other drug abuse were excluded. There were 1,203 respondents from 42 institutions in this study (National Institute on Drug Abuse, 1994b).

The DC\*MADS Homeless and Transient Population Study surveyed persons who were either literally homeless or at imminent risk of becoming homeless, including persons who spent the previous night in an emergency shelter, in a nondomicile (i.e., vacant building, public or commercial facility, city park, car, or on the street), or who were using soup kitchens or emergency food banks for the homeless population. There were 908 respondents in this study (National Institute on Drug Abuse, 1993).

#### **A.1.2 Combining Data Sets for Aggregate Population Estimates**

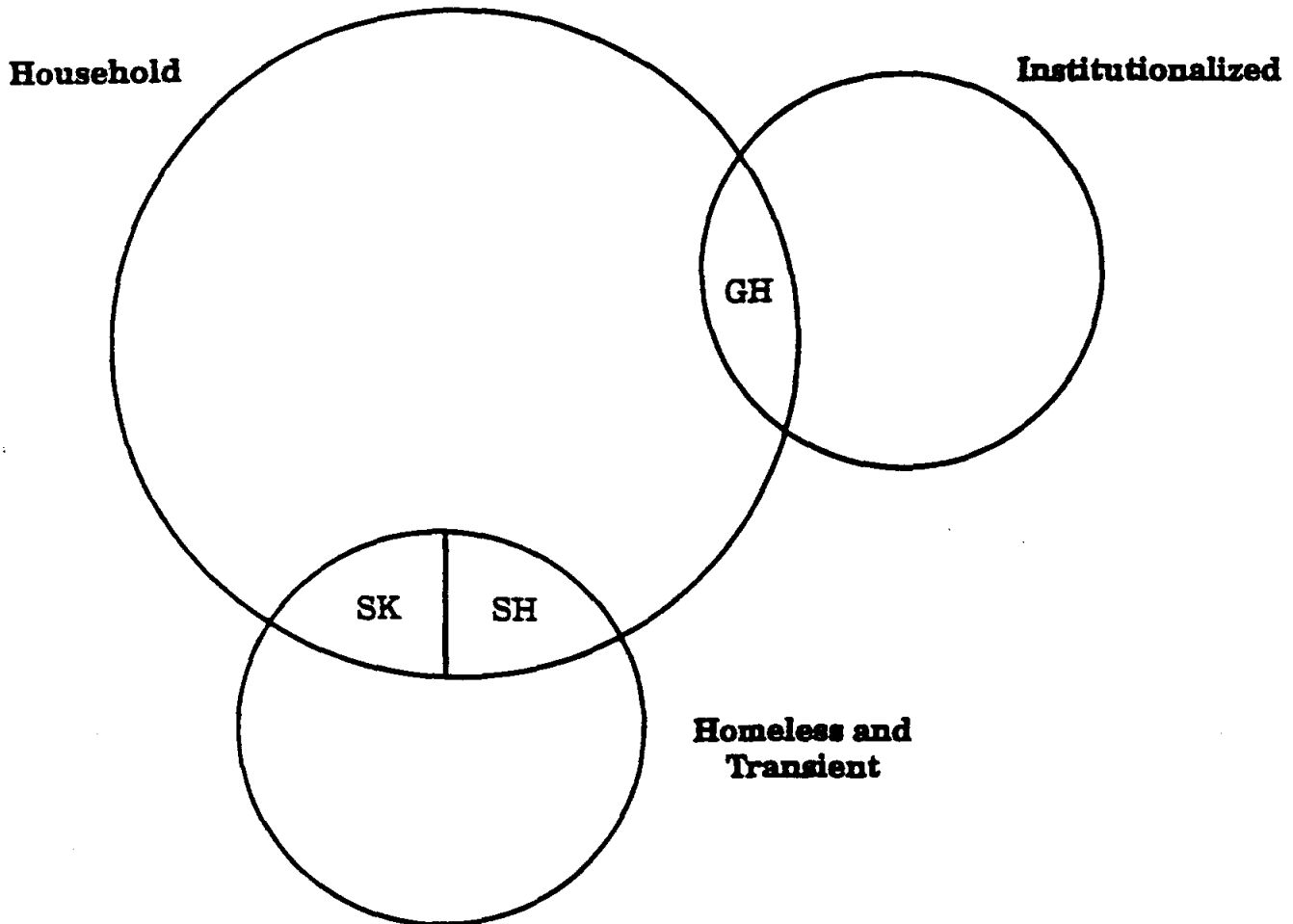
Steps were taken during the planning of these three studies to permit the integration of the data, including coordination of timing of data collection, definition of the subpopulations, structure and content of questionnaires, and estimation procedures.

Although the populations surveyed by the three studies were generally defined in terms of place of residence, there was some overlap in the target populations for the three studies. Both the NHSDA and the Institutionalized Study included some portions of the noninstitutionalized group quarters population. Both the NHSDA and the Homeless and Transient Population Study included persons living in homeless shelters and persons who, while not literally homeless, may have been at risk of homelessness as evidenced by their use of soup kitchens or food banks. The overlap was minimized, however, through careful planning. For example, students living in dormitories were surveyed in the NHSDA but were not included in the Institutionalized Study, and persons in emergency shelters for homeless people were excluded from the sample frame for the Institutionalized Study.

Figure A.1 shows graphically the potential overlap in the target populations for the three surveys. Of the 4,658 persons interviewed in the three studies, 637 could potentially have been interviewed in more than one of the studies (see Table B.3 in Appendix B). In terms of the total numbers of persons represented, however, the overlap was very small; less than 0.05% of the total combined population was potentially represented by more than one of the surveys. Appendix B contains further details of the overlaps of each of the surveys.

Nevertheless, because of these potential overlaps, it was necessary to make adjustments to avoid multiple counting of the subpopulations when producing aggregate estimates. To adjust for the potential overlap, respondents were first classified according to the number of surveys for which they could have been potentially selected. At most, the

**Figure A.1** Overlap of the DC MSA Household, Institutionalized, and Homeless Sampling Frames: 1991



GH = Noninstitutionalized Group Homes  
SK = Not Literally Homeless Who Used Soup Kitchens  
SH = Homeless Shelters

Note: Populations not drawn to scale.

Household data source: 1991 NIDA/SAMHSA National Household Survey on Drug Abuse: DC MSA.

Homeless data source: 1991 NIDA DC\*MADS Homeless and Transient Population Study.

Institutionalized data source: 1991 NIDA DC\*MADS Institutionalized Study.

overlap for subsets of individuals occurred only in two of the three surveys (i.e. household and homeless, household and institutionalized, or homeless and institutionalized). The analysis weights were then adjusted for persons who could have been selected for two surveys by dividing the weights by two. However, it was not known whether persons interviewed for the NHSDA may have been at risk of homelessness, as evidenced by their use of soup kitchens, because this information was not collected by the NHSDA. Thus, it was not possible to completely adjust for this potential multiplicity. It is assumed that only a small proportion of persons who were linked to the area frame used for the NHSDA were also linked to the soup kitchen frame. The procedures used for adjusting for multiplicity are discussed in detail in Appendix B.

### **A.1.3 Analytical Approach**

The aggregate data set provides unbiased estimates of the prevalence of illicit drug, alcohol, and cigarette use among the eligible population in the DC MSA. The SUDAAN software package (Research Triangle Institute, 1990) was used to compute prevalence estimates and associated standard errors. These computations took into account the stratified clustered designs of the surveys. The multiplicity-adjusted weights were used to produce the estimates for the aggregate population data set.

### **A.1.4 Strengths and Limitations of the Methodology**

A major strength of the methodology for producing combined estimates is that these three studies were carefully planned, coordinated, and designed to the extent possible to be combined for making aggregate estimates for the DC MSA. Steps were taken to minimize nonsampling types of differences that might cause differences in estimates in the three populations. The questionnaires were designed so that they have similar structure and content. Data were edited and analysis variables were created following similar algorithms for the three studies.

Nonetheless, in spite of attempts to minimize them, there are some differences in the three studies. First, the mode of administration of the questionnaires differed. For the NHSDA, most of the questionnaire was self-administered, whereas the two DC\*MADS instruments were interviewer-administered because many of the institutionalized and homeless respondents may have had limited reading ability. Although extra steps were taken to protect and reassure respondents of the confidentiality of their data, some respondents might have been less likely to report drug use in the interviewer-administered questionnaire. Although this difference in data collection methods could be viewed as a potential weakness in the methodology, use of self-administered questionnaires in DC\*MADS would likely have resulted in data of poor quality due to respondent difficulty in completing the questionnaires.

Another difference concerns timing of data collection. These three studies were initially designed so that data collection for all three would be conducted in the first half of 1991 (January through June 1991). Data collection for the Institutionalized Study actually extended from April to December 1991, however. In combining data from the three studies, the assumption was made that drug use would be fairly stable in these populations over the time period. However, if drug use showed seasonal variations over the year for the institutionalized population covered by DC\*MADS, then the estimates obtained from combining the data could

vary from the estimates that would be obtained had all three studies been conducted in the first half of 1991. Nevertheless, the actual effect of this variation in the data collection periods is likely to be small because the institutionalized population is only a small fraction of the total population. Furthermore, past month estimates would likely be most vulnerable to any seasonal variations (i.e., lifetime and past year periods would be expected to include periods when use of specific drugs is more prevalent), but past month use tended to be relatively less prevalent for the residents of institutions covered by DC\*MADS.

Finally, estimates generated from this file are for the combined household, institutional, and homeless populations, but they do *not* reflect estimates for the entire DC MSA population. Groups excluded from the studies include those living in nursing homes and those in the military. These groups represent either a small fraction of the total population, or are not likely to be drug users. Some of these groups are covered by other population surveys, such as the Worldwide Surveys of Substance Abuse and Health Behaviors Among Military Personnel (Bray et al., 1992).

## ***Appendix B***

### **PROCEDURES FOR COMBINING HOMELESS, INSTITUTIONALIZED, AND HOUSEHOLD DATA\***

This appendix describes the procedures used for integrating information from three surveys conducted in the District of Columbia Metropolitan Statistical Area (DC MSA) during 1991:

- National Household Survey on Drug Abuse (NHSDA) DC MSA oversample (2,547 respondents),
- DC\*MADS Institutionalized Study (1,203 respondents), and
- DC\*MADS Homeless and Transient Population Study (908 respondents).

The subpopulations surveyed in these three separate studies were generally defined in terms of where respondents were residing at the time of data collection (although there is some overlap in the target populations for the three studies). Combining the data from the three studies permits prevalence estimates to be made for the combined household, homeless, and institutionalized populations covered by DC\*MADS in the DC MSA. Certain steps were taken to permit the integration of information across the various data collection efforts, including coordination of the

- timing of data collection,
- definition of subpopulations,
- content of questionnaires and mode of administration, and
- estimation procedures.

This appendix begins by examining assumptions and issues underlying the estimation procedures. It next presents formulas for obtaining estimates of prevalence within the three surveys. It then offers a summary of the survey designs and their sampling frames, including the potential overlap among them. This information is used to develop a multiplicity estimator that takes this overlap into account. The appendix concludes with a discussion of the statistical test used to make comparisons between household and aggregate data.

#### **B.1 Assumptions and Issues Underlying Estimation Procedures**

Five basic assumptions underlie the estimation procedures described in this appendix:

- data were collected during a common period;

---

\* Adapted from Appendix E of *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the DC Metropolitan Area Household and Nonhousehold Populations: 1991* (National Institute on Drug Abuse, 1994a).

- survey estimates were based on common reference periods;
- drug use was stable over the data collection period;
- samples were allocated randomly; and
- surveys used a common mode of questionnaire administration.

These assumptions were necessary to permit the data to be combined because there were some differences in these dimensions across the surveys. The utility of the estimates relies on the acceptance of these assumptions. To the extent these assumptions are not justified, some bias will be introduced into the estimates.

### **B.1.1 Common Data Collection Period**

The first assumption for combining the data is that they were all conducted during a common data collection period. All three surveys were conducted during the period from January 1991 through December 1991 although the data collection period for the individual surveys varied within this total period. Data collection for the three surveys was as follows:

- 1991 NHSDA DC MSA oversample: January through June 1991;
- Homeless and Transient Population Study: February through June 1991; and
- Institutionalized Study: April through December 1991.

Thus, although the data collection periods for the three surveys were not identical, they all overlapped and occurred within a reasonably short window of time.

### **B.1.2 Survey Reference Periods**

The second assumption was that the reference periods for which estimates were made was common within and across surveys. The data collected in the three surveys cover a number of retrospective reference periods, including the past 30 days, the past year, and the respondent's lifetime. Each of these reference periods is calculated from the date of interview. The collection of data over several months yields rolling reference periods, and the extension of data collection over a number of months results in inexact boundaries of the period to which the subpopulation and overall estimates refer. The 30-day reference period, for example, is defined as the time interval that extends from 30 days prior to the date of the earliest interview to 30 days prior to the last interview for the study.

The subpopulations surveyed in the three separate studies were generally defined by where the person was residing at the time of data collection. Members of the DC MSA population were not static in regard to characteristics over the time period covered by the three surveys. Thus, to describe the characteristics of the DC MSA population, it was necessary to define a hypothetical population that reflects the average situation over the time period covered by the various surveys. This was done by centering the data collection periods on the midpoint of the reference period.

### **B.1.3 Stable Drug Use Over Data Collection Period**

Data collection extended from January through December 1991. The center is approximately June 30, 1991. Because data collection for the NHSDA and the Homeless and Transient Population Study occurred during the first half of the year and data collection for the Institutionalized Study extended through the end of the year, an additional assumption was made that drug use did not change very rapidly in these populations. Thus, for example, the combined estimate of 12-month prevalence is an average experience for the 12-month period that extends from January 1990 through December 1991. However, if drug use showed seasonal variations over the year for the institutionalized population covered by DC\*MADS, then the resulting aggregate estimates could vary from the estimates that would be obtained had all three studies been conducted concurrently (i.e., in the early part of 1991). Nevertheless, the institutionalized population comprises only a small portion of the aggregate population, so any effect on the estimates of prevalence for the aggregate population would be expected to be negligible.

### **B.1.4 Random Allocation of Samples**

It is assumed that all samples were randomly allocated to time within the data collection period so that the prevalence estimates reflect the averages over the data collection period. Of course, that was not strictly the case. To meet the practical demands of the fieldwork, it was not possible to randomly assign all of the respondents across time. The Homeless and Transient Population Study met this assumption the closest in that it included both winter and spring samples that were specifically designed to control for potential bias due to seasonal and time effects.

### **B.1.5 Common Mode of Questionnaire Administration**

To the extent possible, the structure and content of the questionnaires for the three studies were similar. However, the studies varied some in the mode of administration of the questionnaires. The questionnaires for the Homeless and Transient Population and Institutionalized Studies were interviewer-administered because individuals in these populations may have limited reading ability. The questionnaire for the NHSDA was self-administered. This difference in mode of administration may have resulted in differences in reporting in the populations. Although steps were taken to assure respondents that their responses were confidential, it is possible that respondents to the interviewer-administrated questionnaires were less likely to report drug use. Thus, it is possible that drug prevalence rates may have been slightly higher in the homeless and institutionalized populations had it been feasible to use a self-administered questionnaire.

## **B.2 Form of Estimators Within Surveys**

The approach used here to combine data collected over an extended period for a population of people whose characteristics change over that period was to use estimates that reflect the average number of drug users and the average number of people at risk over the period.

Let  $t$  index surveys,

$$Y^p(i, t) = \begin{cases} 1 & \text{if person } i \text{ in survey } t \text{ used drugs in the } p\text{th period prior to the interview} \\ 0 & \text{otherwise,} \end{cases}$$

where

$p$  = 30 days, 12 months, "lifetime,"

$w(i, t)$  = nonresponse-adjusted and multiplicity-adjusted sampling weight for person  $i$  in survey  $t$ , and

$w^*(i, t)$  = final analysis weight for person in survey  $t$  (this weight was adjusted for nonresponse and within-survey multiplicities; a poststratification adjustment may have also been used to adjust for noncoverage).

If  $w^*(i, t)$  was appropriate, then it was used for  $w(i, t)$  in the following equations. The quantity

$$\hat{N}(t) = \sum_{i \in (t)} w(i, t)$$

gives the estimated size of the target population for survey  $t$  on June 30, 1991. The quantity

$$\hat{Y}^p(t) = \sum_{i \in (t)} w(i, t) Y^p(i, t)$$

gives the estimated number of persons in the target population for survey  $t$  who used drugs in the  $p$ th period prior to data collection.

The 30-day prevalence was operationally defined as the estimated proportion of the population alive on June 30, 1991, who used drugs in the 30 days prior to their date of interview with

$$\hat{P}^{30}(t) = \frac{\hat{Y}^{30}(t)}{\hat{N}(t)}$$

This is the ratio of the average size of the drug-using population to the average size of the overall population, as the following illustrates.

Assume the sample for survey  $t$  is randomly allocated to  $d = 1, 2, \dots, D$  data collection periods, and let

$w(i, t, d)$  = nonresponse-adjusted weight for person  $i$  in survey  $t$  in data collection period  $d$

$$\hat{N}(t) = \frac{1}{D} \sum_{d=1}^D \sum_{i \in s(t,d)} w(i,t,d), \text{ and}$$

$$\hat{Y}^P(t) = \frac{1}{D} \sum_{d=1}^D \sum_{i \in s(t,d)} Y^P(i,t,d) w(i,t,d).$$

Now,  $w(i,t,d) = D w(i,t)$  because  $\pi(i,t,d) = \frac{1}{D} \pi(i,t)$ , where  $\pi(i,t) = \text{Prob} \{i \in s(t)\}$ . Then

$$\begin{aligned} \hat{N}(t) &= \frac{1}{D} \sum_{d=1}^D \sum_{i \in s(t,d)} D w(i,t) \\ &= \sum_{i \in s(t)} w(i,t). \end{aligned}$$

Similarly,

$$\begin{aligned} \hat{Y}^P(t) &= \frac{1}{D} \sum_{d=1}^D \sum_{i \in s(t,d)} D w(i,t) Y^P(i,t) \\ &= \sum_{i \in s(t)} w(i,t) Y^P(i,t), \end{aligned}$$

where  $i \in s(t,d)$  denotes sampled units  $i$  in survey  $t$ , time period  $d$ , and  $i \in s(t)$  denotes sampled units  $i$  in survey  $t$ .

Similarly, the 12-month prevalence was operationally defined as the estimated proportion of population ( $t$ ) who were alive on June 30, 1991, who reported drug use in the 12 months prior to the date of their interview.

### B.3 Summary of Surveys and Overlaps Among Sampling Frames

#### B.3.1 Survey 1: National Household Survey on Drug Abuse, DC MSA Oversample, 1991

The target population was the civilian, noninstitutionalized population, which included individuals in three types of dwellings:

- housing units, as defined by the U.S. Bureau of the Census (1982, pp. 20-21);
- noninstitutionalized group quarters, as defined for the 1980 census, which included persons living in rooming houses, group homes, religious group quarters, or college quarters off campus with 10 or more unrelated persons; persons living in college dormitories, military quarters, agricultural workers' dormitories, other workers' dormitories

(regardless of the number of persons living there); and emergency shelters for the homeless population; and

- civilian housing on military bases.

Details of the NHSDA sample design appear in *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the DC Metropolitan Area Household and Nonhousehold Populations* (National Institute on Drug Abuse, 1994a).

Response-adjusted weights were poststratified at the national level to Current Population Survey (CPS) estimates of the civilian, noninstitutionalized population as of mid-March 1991. The weights for the DC MSA oversample were not poststratified to a separate estimate for the DC MSA population. Prior to using these data for producing the aggregate estimates presented in this report, weight sums by location (DC, Virginia, Maryland), from the 1990 NHSDA's DC MSA oversample were compared with counts obtained from the 1990 census for persons aged 12 and older. Data from the 1990 NHSDA were used rather than data from the 1991 survey for a more direct comparison with the 1990 census data. Counts of persons in military quarters and in noninstitutional group quarters were subtracted from the census data before making the comparison because the NHSDA targets the civilian, noninstitutionalized population. Military personnel in civilian housing were not eligible for the NHSDA. However, counts of these persons in the DC MSA were not readily obtainable from the U.S. Bureau of the Census and thus could not be excluded prior to making this comparison.

Table B.1 compares the estimated number of persons aged 12 and older from the 1990 NHSDA and the 1990 census. The estimates obtained from the NHSDA are slightly lower than those given by the census; however the census totals are all covered by 95% confidence intervals around the NHSDA estimates. It was determined, based on this comparison, that no reweighting was necessary of the 1991 NHSDA data to poststratify the weights to an estimate of the DC MSA population.

### **B.3.2 Survey 2: DC\*MADS Institutionalized Study**

The target population was persons under institutional care or custody, regardless of their length of stay in that place and regardless of the number of people in that place. Institutional and noninstitutional group quarters defined for this study included

- mental or psychiatric hospitals;
- correctional institutions;
- hospitals for chronically ill people;
- homes, schools, and wards for people with mental disabilities;
- homes, schools, and wards for people with physical disabilities;
- homes for children who are dependent and/or neglected;

**Table B.1 Comparison of the Number of Persons Aged 12 and Older from the 1990 NHSDA and the 1990 Census, by Location Within the DC MSA**

<b>Survey</b>	<b>Location</b>			
	<b>DC MSA</b>	<b>DC</b>	<b>Maryland</b>	<b>Virginia</b>
<b>1990 NHSDA</b>				
Estimated persons aged 12 and older (Standard error)	2,909,323 (164,488)	466,565 (59,584)	1,336,098 (146,250)	1,106,660 (202,755)
<b>1990 Census</b>				
Total persons aged 12 and older	3,280,341	525,146	1,480,617	1,274,578
Institutionalized group quarters	40,699	14,070	12,407	14,222
Military group quarters	14,407	2,181	3,890	8,366
Difference <sup>1</sup>	3,225,235	508,895	1,464,320	1,252,020

<sup>1</sup>Computed by subtracting the count of persons in institutionalized group quarters and military group quarters from the total number of persons aged 12 and older. Military personnel residing in civilian housing were included in this difference because it was not possible to obtain these counts from the U.S. Bureau of the Census (1991) summary tape file 1a for the 1990 census.

Sources: National Institute on Drug Abuse, National Household on Drug Abuse, 1990: DC MSA; U.S. Bureau of the Census (1991).

- training schools for juvenile delinquents;
- detention homes; and
- dormitories for agriculture workers or other workers.

The Institutionalized Study also excluded certain segments of institutional and noninstitutional group quarters populations. Excluded from the study were

- nursing homes,
- religious group quarters,
- military quarters,
- rooming houses,
- college dormitories,
- hospitals or wards for alcohol or drug abuse, and
- emergency shelters for homeless people, runaways, or neglected or abused women.

The sampling frames consisted of lists of institutions located in the DC MSA and lists covering group homes, religious group quarters, and workers' dormitories. The survey design was a two-stage stratified sample of institutions and persons within institutions and noninstitutional group quarters. The four institutional and group quarters strata were

- correctional institutions,
- psychiatric institutions,
- other institutions, and
- group homes.

Sample weights that reflect the inverse of the inclusion probability were calculated; these weights were adjusted for nonresponse using a weighting class adjustment. Further details about the sample design appear in Section 3.1.1 of *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the DC Metropolitan Area Household and Nonhousehold Populations* (National Institute on Drug Abuse, 1994a).

### **B.3.3 Survey 3: DC\*MADS Homeless and Transient Population Study**

The study targeted people who were either literally homeless or at imminent risk of moving into or out of homelessness. An eligible person met one of the following conditions on the sampled day:

- someone who stayed overnight in an emergency shelter for homeless people, runaways, or neglected or abused women;
- someone who stayed overnight in a house, apartment, or room paid for with municipal emergency housing funds;
- someone who stayed overnight in a nondomicile, such as a vacant building, public or commercial facility, city park, car, or on the street;
- someone whose regular place to stay was a nondomicile, regardless of where he/she stayed the prior night (e.g., people who traded sexual favors for shelter or spent one night in a hotel or hospital); or
- someone who was using a soup kitchen or emergency food bank for the homeless population.

The sampling frames consisted of lists of shelters and lists of blocks as defined by the census; persons in nonlocked private property and in cars; and lists of soup kitchens and food banks. Further details of the sample design appear in Section 4.1.1 of *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the DC Metropolitan Area Household and Nonhousehold Populations* (National Institute on Drug Abuse, 1994a). Sample weights were adjusted for frame multiplicities and nonresponse. The weight sums estimate the average number of homeless persons per day during the data collection period.

### **B.3.4 Overlap of the Surveys and Frames**

Table B.2 shows the various subpopulations included in the three studies and their potential overlap. Figure A.1 in this document diagrams the overlap. Table B.3 summarizes the sample sizes and estimated subpopulation sizes from the studies; the subpopulation sizes were obtained by summing the final analysis weights for each survey. Table B.2 also indicates that some persons were not in the target populations of any of the three surveys. These include noncivilians in housing units, persons in military group quarters, nursing homes, and hospitals or wards for alcohol or drug abuse. Data from the U.S. Bureau of the Census (1991) indicate that in 1990 in the DC MSA there were

- 14,407 persons in military group quarters (0.4% of total population),
- 20,480 persons in nursing homes (46% of total civilian institutionalized and group quarters population, or 0.6% of total population), and
- 2,014 persons in types of institutionalized group quarters other than correctional institutions, nursing homes, mental or psychiatric hospitals, and juvenile institutions

(these include hospitals or wards for alcohol abuse) (4.6% of total civilian institutionalized and group quarters population, or 0.1% of total population).

A count of military personnel in civilian housing in the DC MSA is not readily available from the census data. In addition, the census counts of institutionalized persons in this discussion and in Table B.1 include 283 persons under age 12 in the DC MSA for whom the type of institution cannot be determined; however, these counts give an indication of the numbers of individuals not included in the target populations for the three surveys. This total, 36,901, is 1.1% of the total persons age 12 and older in the DC MSA.

Because persons who were not literally homeless were included in the soup kitchen sample in the Homeless and Transient Population Study, these persons have links to both the soup kitchen frame used for that study and the area housing unit frame used for the NHSDA. Conversely, persons in the area household frame who used soup kitchens would also have been linked to both of these frames. Information on use of soup kitchens was not collected in the NHSDA, so adjustments for this potential multiplicity in the NHSDA sample are not possible. However, this is a minor problem because it is assumed that only a very small proportion of persons who were linked to the area household frame were also linked to the soup kitchen frame. The possibility of multiple linkages of persons on the soup kitchen frame to the area household frame is much greater, however. Of the 224 interviews obtained in soup kitchens, 82 were with individuals who were not literally homeless. Because both the Institutionalized Study and the NHSDA included noninstitutionalized group homes, persons in group homes also had multiple chances of being included in the surveys. In the Institutionalized Study, 73 interviews were conducted with residents of noninstitutionalized group homes.

Table B.3 indicates that only five interviews in the 1991 NHSDA's DC MSA oversample were conducted with residents of noninstitutionalized group quarters. These may have been either group homes (as included in the Institutionalized Study) or homeless shelters (as included in the Homeless and Transient Population Study). These sums of the weights in the overlap are only a small proportion of the total NHSDA weight sum (less than 0.5%).

#### **B.4 Combined Estimates of Prevalence**

The multiplicity of an individual is the number of links that an individual may have to different sampling frames. This section presents a technical definition and justification of the multiplicity estimator.

Because all estimates refer to the DC MSA populations as of June 30, 1991, these three surveys can conceptually be considered to be a single survey of a population divided into three (partially overlapping) super-strata denoted by  $i$  where

- $i$  = Super-stratum
- 1 = Household/group quarter population (NHSDA)
- 2 = Institutionalized population
- 3 = Homeless population

**Table B.2 Potential Overlap Among the Target Populations for the DC MSA Household, Institutionalized, and Homeless Surveys**

Survey Frames	Subpopulations Included in the Survey		
	Household	Institutionalized	Homeless
Civilians in housing units	✓		
Noncivilians in housing units <sup>1</sup>			
Civilians in military housing	✓		
<b>Noninstitutional group quarters</b>			
Rooming houses	✓		
Group homes <sup>2</sup>	✓	✓	
College dormitories or college quarters off campus	✓		
Religious group quarters	✓		
Military quarters <sup>1</sup>			
Agriculture workers' or other workers' dormitories <sup>2</sup>	✓	✓	
Emergency shelters for homeless people <sup>2</sup>	✓		✓
Emergency shelters for women who are dependent, neglected, or abused <sup>2</sup>	✓		✓
<b>Institutionalized group quarters</b>			
Correctional institutions		✓	
Nursing homes <sup>1</sup>			
Mental or psychiatric hospitals		✓	
Hospitals for people who are chronically ill		✓	
Schools, hospitals, or wards for people with physical disabilities		✓	
Hospitals or wards for people who abuse alcohol or drugs <sup>1</sup>			
Institutions for juveniles		✓	
<b>Other</b>			
Persons who use soup kitchens <sup>2</sup>	✓		✓
Persons in nondomiciles			✓

✓ Indicates subpopulations who had a chance to be included in the survey. The institutional and noninstitutional group quarters classifications are those used by the 1990 census.

<sup>1</sup>Not in the target populations of any of the three surveys.

<sup>2</sup>Overlapping frames across surveys.

**Table B.3 Sample Sizes and Estimated Subpopulation Sizes for the Three Studies**

<b>Study</b>	<b>Sample Size</b>	<b>Weight Sum</b>
<b>1. 1991 NHSDA DC oversample (total)</b>	<b>2,547</b>	<b>3,174,498</b>
1a. Group quarters	5	5,167
1b. Used soup kitchens	DK	DK
<b>2. Institutionalized Study (total)</b>	<b>1,203</b>	<b>19,757</b>
2a. Group homes	73	724
<b>3. Homeless and Transient Population Study (total)</b>	<b>908</b>	<b>10,387</b>
3a. Not literally homeless, but used soup kitchens	82	2,031
3b. Shelter users	477	5,844
<b>Total overlap (Sum of 1a, 2a, 3a, 3b)</b>	<b>637</b>	<b>13,766</b>

Overlaps: Persons in 1a. may also be in either 2a. or 3b. Persons in 1b. are also in 3a.  
 DK = don't know.

**B.4.1 Multiplicity Estimators**

Let  $\alpha$  index persons within the DC MSA,  $\alpha = 1, 2, \dots, N$ , and let  $N$  denote the total number of persons in DC MSA. Overall prevalence estimates are designed to estimate the parameter  $\frac{Y}{N}$  where

$$Y = \sum_{\alpha=1}^N Y(\alpha).$$

Here,  $Y(\alpha)$  is the (0-1) outcome variable for person  $\alpha$ . Let  $t$  index the target populations of interest in the three surveys, and let

$$\theta(\alpha, t) = \begin{cases} 1 & \text{if person } \alpha \text{ is a member of the target population of survey } t \\ 0 & \text{otherwise} \end{cases}$$

Then

$$N(t) = \sum_{\alpha=1}^N \theta(\alpha, t)$$

= total number of people in target population of survey  $t$ .

The goal of the individual surveys is to estimate  $\frac{Y(t)}{N(t)}$ . Analysis weights that have been adjusted for nonresponse and within-survey multiplicities are available for the three surveys.

For combining the three studies, a "superstage" multiplicity estimator described by Levy (1977) and Lessler and Kalsbeek (1992) was used. By counting the survey multiplicity, a multiplicity estimator for combining the three studies is developed.

Let

$$\gamma_{\alpha} = \sum_{t=1}^3 \theta(\alpha, t) = \text{number of linkages that person } \alpha \text{ has to the three surveys}$$

= survey multiplicity for person  $\alpha$ .

Note that  $\gamma_{\alpha}$  = either one or two for these three surveys because a person can be in the target population of either one or two of the surveys. Let

$$I(i, t) = \begin{cases} 1 & \text{if person } i \text{ is in the sample for survey } t \\ 0 & \text{otherwise} \end{cases}$$

$w^*(i, t)$  = final response-adjusted analysis weight for respondent  $i$  in survey  $t$ , adjusted for within survey multiplicities.

An estimate of the total,  $\hat{Y}$ , is formed by

$$\hat{Y} = \sum_{t=1}^3 \sum_{i=1}^{N(t)} \sum_{\alpha=1}^N \frac{\theta(\alpha, t) I(i, t) w^*(i, t) Y(\alpha)}{\gamma_{\alpha}},$$

where  $N(t)$  denotes the total persons in subpopulation  $t$ . The denominator is similarly computed as

$$\hat{N} = \sum_{t=1}^3 \sum_{i=1}^{N(t)} \sum_{\alpha=1}^N \frac{\theta(\alpha, t) I(i, t) w^*(i, t)}{\gamma_{\alpha}}.$$

That is, for combining the studies, the multiplicity estimator forms a new weight variable,  $w^{**}(i,t)$ , which is the final analysis weight for each respondent divided by the respondent's survey multiplicity:  $w^{**}(i,t) = w^*(i,t)/\gamma_a$ . This new weight is adjusted for multiplicities across the surveys.

#### **B.4.2 Optimizing the Multiplicity Adjustment**

Even though the use of multiplicity-adjusted weights in the overlapping portions reduces the bias (ideally to zero), it is plausible that this reduction is more than compensated for by an increase in sampling variance. To assess the trade-offs involved in the use of multiplicity estimates, the variances of key estimates were examined under alternative weighting schemes for four different options. These options differed in how the data in the overlapping portions of the target populations were treated:

- Option 1. Disregard the NHSDA portion of the overlap with the other two surveys (and assume that the number of interviews with users of soup kitchens in the NHSDA is negligible).
- Option 2. Use multiplicity-adjusted overlap (with  $\gamma_a$ ) for all overlapping portions.
- Option 3. Disregard the interviews with soup kitchen users who were not literally homeless from the homeless survey portion of the overlap, and use multiplicity-adjusted weights for shelter and noninstitutionalized group home interviews.
- Option 4. Disregard the interviews with soup kitchen users who were not literally homeless from the homeless survey portion of the overlap, and disregard the NHSDA portion of the overlap with the other two surveys.

For each of these four options, Table B-4 presents the estimated total number of persons in the union of the three populations and prevalence estimates related to past year and past month use of any illicit drug, crack-cocaine, heroin, and alcohol. Also shown are the estimated number of users, standard errors (SEs), and relative standard errors (RSEs) of all estimates. In general, estimates of prevalence are similar for the four options. Option 1, which would have disregarded NHSDA interviews in the overlap for producing the estimates, generally yields smaller RSEs than the other options, although the actual decrease was very small. It was undesirable to discard part of the observations in the estimation of the overlaps. Based on the examination of the RSEs, there was no overwhelming reason to do so. Based on this investigation, the multiplicity-adjusted weights (Option 2) were selected for use in producing estimates for the combined household, homeless, and institutionalized populations.

**Table B.4 Estimation Options for Combining the Overlaps in the Three Surveys**

	Option 1			Option 2			Option 3			Option 4		
	EST	SE	RSE <sup>1</sup>	EST	SE	RSE <sup>1</sup>	EST	SE	RSE <sup>1</sup>	EST	SE	RSE <sup>1</sup>
<b>N (thousands)</b>	3,199	131	4.079	3,199	129	4.043	3,198	129	4.044	3,197	131	4.082
<b>Prevalence (percentage)</b>												
<b>Past year</b>												
Any illicit drug	12.075	1.147	9.501	12.028	1.146	9.527	12.014	1.146	9.538	12.048	1.147	9.524
Any cocaine	3.965	0.435	10.973	3.924	0.434	11.059	3.909	0.434	11.098	3.934	0.435	11.049
Crack	1.240	0.261	21.054	1.202	0.261	21.723	1.187	0.261	22.002	1.210	0.261	21.581
Heroin	0.493	0.263	53.320	0.486	0.263	54.069	0.484	0.263	54.323	0.489	0.263	53.813
Alcohol	73.430	1.225	1.668	73.443	1.220	1.662	73.440	1.221	1.662	73.424	1.225	1.669
<b>Past month</b>												
Any illicit drug	5.799	0.704	12.148	5.778	0.705	12.195	5.768	0.705	12.218	5.780	0.705	12.183
Any cocaine	1.607	0.494	30.727	1.590	0.494	31.047	1.682	0.494	31.203	1.691	0.494	31.031
Crack	0.441	0.072	16.236	0.424	0.071	16.721	0.417	0.071	16.955	0.426	0.071	16.656
Heroin	0.296	0.256	86.384	0.294	0.256	87.164	0.292	0.256	87.672	0.293	0.256	87.385
Alcohol	55.589	1.323	2.380	55.620	1.322	2.377	55.616	1.323	2.378	55.580	1.324	2.382
<b>Number of users (thousands)</b>												
<b>Past year</b>												
Any illicit drug	386	33	8.506	385	33	8.541	384	33	8.553	385	33	8.531
Any cocaine	127	13	10.083	126	13	10.187	125	13	10.228	126	13	10.163
Crack	40	9	21.667	38	9	22.348	38	9	22.632	39	9	22.204
Heroin	16	8	52.500	16	8	53.252	15	8	53.506	16	8	52.995
Alcohol	2,349	112	4.776	2,349	111	4.725	2,348	111	4.727	2,348	112	4.780
<b>Past month</b>												
Any illicit drug	186	23	12.267	185	23	12.315	184	23	12.339	185	23	12.315
Any cocaine	51	15	29.770	51	15	30.100	51	15	30.256	51	15	30.076
Crack	14	2	15.792	14	2	16.320	13	2	16.560	14	2	16.224
Heroin	9	8	85.353	9	8	86.136	9	8	86.644	9	8	86.354
Alcohol	1,779	88	4.929	1,779	87	4.871	1,778	87	4.873	1,777	88	4.933

<sup>1</sup>Relative standard errors (RSEs) are calculated as the standard error (SE) divided by the estimate (EST). RSEs are expressed in the table as percentages.



## **Appendix C**

### **OTHER METHODOLOGICAL NOTES**

#### **C.1 Use of Weight Variables**

Two alternative weight variables are included in this dataset, WEIGHT and MULTWT. For within-survey estimates, the weight variable WEIGHT should be used, which is equivalent to individual weight variables from each component survey. The WEIGHT variable *does not* include adjustments for potential multiplicity in the sampling frames and would be used only for analysis of data from a component subpopulation in the aggregate file. For example, if data for the household sample only (if SURVEY=1) are processed as a subset, then the appropriate weight to use is WEIGHT. Similarly for analyses of subgroups corresponding only to the homeless and transient segment or the institutionalized segment, the weight variable WEIGHT should be used. For across-survey or aggregate estimates, MULTWT should be used, since it is multiplicity-adjusted. Derivation of both weight variables are explained in detail in *Appendix B*.

#### **C.2 Calculating Standard Errors from a Complex Sample Design\***

As noted by Cohen, Xanthopoulos, and Jones (1986), national surveys conducted by government organizations, industry, political organizations, and market research firms need to provide the greatest precision in estimates from sample data for fixed cost and time constraints. Consequently, many national surveys are characterized by design components that include stratification, clustering, and disproportionate sampling.

Such design features complicate the data analysis while reducing the cost of data collection. Data from complex survey designs of this type deviate from the assumption of simple random sampling and require special consideration with regard to variance estimation and analysis.

Statistical software packages are currently available that accommodate these complex survey designs and allow for the generation of variance estimates of statistics expressed in terms of means, totals, ratios, and multivariate regression coefficients. The procedures vary, however, in program capabilities, computational efficiency, and user facilities. See Wolter (1985) for an overview of various computer software packages available and a discussion of criteria for selecting appropriate software for various situations. Three widely used and available software packages are the SUDAAN (Survey Data Analysis) procedure developed by Research Triangle Institute (1990), the WESVAR (WESStat VARIance Estimation) procedure developed by Westat (Flyer & Mohadjer, 1988), and the earlier procedures developed by the Statistical Laboratory at Iowa State University (Fuller, 1986).

To estimate proportions, means, and standard errors in SUDAAN in accordance with the sample design, a stratified, two-stage design was specified that incorporates with

---

\*Adapted from Appendix E of *The Washington, DC, Metropolitan Area Homeless and Transient Population Study, 1991, Public Use Data Tape Documentation* (National Institute on Drug Abuse, 1994c).

replacement sampling at the first stage. For the SUDAAN procedures, strata are represented by the variable STRATUM, and primary sampling units are represented by the variable WPSU. In this dataset, the variables STRATUM and WPSU have survey-specific definitions to incorporate the various sampling schemes and obtain estimates both within and across component surveys.

### **C.3 Identifying Low Precision Estimates\***

As with findings from any sample survey, prevalence estimates are subject to two kinds of error: nonsampling error and sampling error. Nonsampling error results from such factors as nonresponse, misreporting of data by the respondent, and miscoding of responses. Although the extent of nonsampling error cannot be precisely measured, attempts can be made to reduce it through quality control procedures and other means. Quality control procedures for the 1991 National Household Survey on Drug Abuse (NHSDA) are described in *Appendix B* of the 1991 *Main Findings* report (Substance Abuse and Mental Health Services Administration, 1993); quality control procedures that were used in the Homeless and Transient and Institutionalized Population Studies are described in detail in *Appendix A* of their respective reports (National Institute on Drug Abuse, 1993, 1994b).

Sampling error results from collecting data from a subset rather than from everyone in the population. Also known as sampling variability, sampling error is the variation among a set of estimates that would be observed if repeated samples of the same type were drawn from the same population. The magnitude of sampling error depends on (a) the inherent variability of the measured attribute in the population; (b) the sample size; (c) the extent of homogeneity of the sample on the variable in question (based on such factors as similarity of respondents within sample clusters and dissimilarity between clusters); and (d) the type of sampling and estimation procedures used. Sampling theory provides a basis for calculating the degree of sampling error; two commonly reported measures are the standard error (SE) and the relative standard error (RSE), defined as the ratio of the SE to the actual estimate and expressed as a percentage of the estimate. SEs are used to compute confidence intervals for estimates and also enter into the calculations required to test the statistical significance of the difference between two estimates.

Estimates subject to a high degree of sampling error are considered to have low precision. Low precision has been defined in many ways with no common definition used across various sample surveys. Thus, as in specifying an alpha level for conducting tests of significance or constructing confidence intervals, there will always be some subjectivity in defining low precision.

The precision criterion applied to the estimates in the *Technical Report* (National Institute on Drug Abuse, 1994a) was originally developed for the District of Columbia Metropolitan Statistical Area (DC MSA) oversample of the 1990 NHSDA and is based on the RSE of the natural logarithm of the estimate. The criterion, described further in the next section, is somewhat conservative and tends to require relatively large sample sizes to obtain an acceptable level of precision. When this criterion is used, low precision may occur if prevalence rates are close to 0% or 100%, or when the number of respondents in a particular

---

\*Adapted from Section 1.6 of *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the DC Metropolitan Area Household and Nonhousehold Populations: 1991* (National Institute on Drug Abuse, 1994a).

subgroup is small. Low precision estimates typically are omitted from all tables and figures and noted by an asterisk (\*), consistent with the procedure in the NHSDA (Substance Abuse and Mental Health Services Administration 1993). Consequently, readers should exercise caution in using these low precision estimates and are encouraged to use them in conjunction with their accompanying SEs.

In addition to flagging low precision estimates, very small estimates that were less than 0.05% but still met the precision criterion typically are shown in the tables as double asterisks (\*\*). Readers should keep in mind that these estimates are not actually zero, but that they round to zero when data are reported to only one decimal place.

#### C.4 Low Precision Rule for Prevalence Estimates\*

This section describes the rule used in analyses of this data set to identify and suppress unreliable prevalence estimates (i.e., rates that cannot be reported with confidence because they are based on small sample sizes or have large sampling errors). In defining a suppression rule, the goals are to capture unreliable estimates easily, to have broad applicability across both national and metropolitan area samples, and to have a rule that could be easily incorporated into table-producing software.

Prior to 1990, the rule used in the NHSDA was to suppress estimates with a relative standard error (RSE) greater than or equal to 50% of the prevalence estimate. The RSE is computed by dividing the standard error of the estimate by the estimate. Although the 50% RSE rule was easy to implement and understand, it was observed to have some undesirable properties. Specifically, the rule imposes a very stringent suppression requirement on small prevalence rates, but a very lax requirement on large rates. That is, small prevalence rates must have relatively large sample sizes to avoid being suppressed, whereas large rates require much smaller sample sizes.

To better address this limitation of the 50% RSE rule, a new suppression rule was adopted based on the RSE of the natural log of the estimate. Specifically, estimates are suppressed and shown as an asterisk (\*) in a table or figure when

$$RSE[-\ln(p)] > 0.175 \text{ when } p \leq 0.5$$

or

$$RSE[-\ln(p)] > 0.175 \text{ when } p > 0.5$$

Note that  $RSE[-\ln(P)] = RSE(P) / -\ln(P)$ . This is based on a first order Taylor series expansion of  $-\ln(P)$ .

For computational purposes, the above is equivalent to

---

\*Adapted from Appendix C of *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the DC Metropolitan Area Household and Nonhousehold Populations: 1991* (National Institute on Drug Abuse, 1994a).

$$\frac{SE(p) / p}{-\ln(p)} > 0.175 \text{ when } p \leq 0.5$$

or

$$\frac{SE(p) / (1-p)}{-\ln(1-p)} > 0.175 \text{ when } p > 0.5.$$

where  $SE(p)$  equals the standard error estimate of  $p$ . The log transformation of  $p$  is used to provide a more balanced treatment of measuring the quality of small, large, and intermediate  $p$  values. The switch to  $(1-p)$  for  $p$  greater than 0.5 provides a symmetric suppression rule across the range of possible  $p$  values.

This new suppression rule has been used in analyses of data from the DC MSA oversample of the 1990 NHSDA (National Institute on Drug Abuse, 1991) and data from the 1991 NHSDA (Substance Abuse and Mental Health Services Administration, 1993). In addition, this rule was used to identify low precision estimates in the DC\*MADS Institutionalized and Homeless and Transient Population Studies.

Compared with the 50% RSE rule, the new rule is more liberal in allowing small prevalence rates to be published, but more stringent in preventing large prevalence rates from being published. For example, under the new rule, it is impossible for prevalence rates of about 1% to be published unless they are based on a sample of 150 or more respondents. The 50% RSE rule would have required a sample size of 400 or more respondents. A 20% prevalence rate requires a minimum sample size of 50 under the new rule, whereas the old rule required only 16.

### C.5 Calculation of Confidence Intervals\*

Sampling error occurs due to the random process of sampling the total population of inferential interest (i.e., the civilian, noninstitutionalized population age 12 and older of the United States). Following procedures used in the NHSDA, 95% confidence intervals are calculated to quantify sampling error. Because estimates in the NHSDA are frequently small percentages, the confidence intervals were based on logit transformations. Logit transformations yield asymmetric interval boundaries that provide a more suitable measure of sampling error for small percentages.

To illustrate, let the proportion  $P_d$  represent the true prevalence rate for a particular analysis domain  $d$ . Then the logit transformation of  $P_d$ , commonly referred to as the "log odds," is defined as

$$L = \ln[p_d / (1 - p_d)]$$

where "ln" denotes the natural logarithm.

---

\*Adapted from Appendix C of *National Household Survey on Drug Abuse: Main Findings 1992* (Substance Abuse and Mental Health Services Administration, 1995).

Letting  $P_d$  be the estimate of the proportion, the log odds estimate becomes  $\hat{L} = \ln[p_d / (1 - p_d)]$ . Then, the lower and upper limits of  $L$  are calculated as

$$A = \hat{L} - K \left[ \frac{\sqrt{\text{var}(p_d)}}{P_d(1 - p_d)} \right]$$

$$B = \hat{L} + K \left[ \frac{\sqrt{\text{var}(p_d)}}{P_d(1 - p_d)} \right]$$

where  $\text{var}(p_d)$  is the variance estimate of  $p_d$ , and  $K$  is the constant chosen to yield the proper level of confidence (e.g.,  $K = 1.96$  for 95% confidence limits).

Applying the inverse logit transformation to  $A$  and  $B$  above yields a confidence interval for  $p_d$  as follows:

$$P_{d,lower} = \frac{1}{1 + \exp(-A)}$$

$$P_{d,upper} = \frac{1}{1 + \exp(-B)}$$

where "exp" denotes the inverse log transformation. The upper and lower limits of the percentage estimate are obtained by simply multiplying the upper and lower limits of  $p$  by 100.

Corresponding to the percentage estimates, the number of drug users,  $Y_d$ , can be estimated as

$$\hat{Y}_d = \hat{N}_d * P_d$$

where

$\hat{N}_d$  = estimated population total for domain  $d$

$P_d$  = estimated proportion for domain  $d$ .

The confidence interval for  $\hat{Y}_d$  is obtained by multiplying the lower and upper limits of the proportion confidence interval by  $\hat{N}_d$ .

In addition, the variance of  $\hat{Y}_d$  can be estimated as

$$\text{var}(\hat{Y}_d) = \hat{N}_d^2 * \text{var}(p_d).$$

For the NHSDA, the design-based variance was estimated using a Taylor series linearization. For a given variance estimate, the associated design effect is the ratio of the design based variance estimate over the variance that would have been obtained from a simple random sample of the same size. Because the combined design features of stratification, clustering, and unequal weighting are expected to increase the variance estimates, the design effect should virtually always be greater than one. However, for prevalence rates near zero, the variance inflating effects of unequal weighting and clustering were sometimes underestimated, resulting in design effects less than one. Because the corresponding variance estimates were considered anomalously small, two other variance estimates were computed as quality control measures. The first was based only on the stratification and unequal weighting effects and the second was based on no effects or simple random sampling. The reported variance estimate was then the maximum of these three estimates.

### C.6 Example Program for Complex Variance Estimation (SUDAAN)

This section provides an example of SUDAAN statements used to generate estimates and standard errors. Prior to this SUDAAN segment, data in the flat file need to be read into a SAS file with variables properly set up, missing values assigned, and the file sorted by the NEST variables.

```

1  PROC CROSSTAB DATA=" [STUDY1A]EXAMPLE"
           FILETYPE=SAS
           DESIGN=WR MEANS;
2  NEST    STRATUM WPSU;
3  WEIGHT  MULTWT;
4  SUBGROUP RACE SUMYR SUMMON;
5  LEVELS  4 2 2 ;
6  TABLES RACE*(SUMYR SUMMON);
7  SETENV  LINESIZE=80
           PAGESIZE=60
           DECWIDTH=5
           COLWIDTH=10
           LABWIDTH=25;
8  TITLE  "Any Illicit Drug Use by Race";

```

This example and description are not intended as a guide to using SUDAAN, but rather as an example of its use with the DC\*MADS Study of Household and Nonhousehold Populations. Briefly, the SUDAAN program reads data from the SAS file that has been sorted by the sampling levels used in the sampling design, STRATUM and WPSU, listed on the *NEST* statement. The default sampling design, with replacement (WR) sampling at the first stage, is used to generate better variance estimates for multistage designs. The data have been weighted using the survey multiplicity-adjusted analysis weight, MULTWT, listed on the *WEIGHT* statement. This example requests the computation of weighted joint distributions and standard errors. The *TABLES* statement specifies the cross-tabulations for which estimates are to be calculated. The variables on the *TABLES* statement must be listed in the *SUBGROUP* statement and their number of levels provided on the *LEVELS* statement. The estimates are printed by specified instructions using the SETENV and PRINT options.

**For survey-specific estimates, the data should be weighted by WEIGHT, the survey-specific analysis weight. When analysis is done on a subset of the variable SURVEY, or when a variable of interest is crossed with SURVEY, use of the weight variable WEIGHT will produce the equivalent of separate estimates for the different surveys without overlap adjustment.**



## **Appendix D**

### **REFERENCES**

- Bray, R.M., Kroutil, L.A., Luckey, J.W., Wheelless, S.C., Iannacchione, V.G., Anderson, D.W., Marsden, M.E., & Dunteman, G.H. (1992). *1992 Worldwide Survey of Substance Abuse and Health Behaviors Among Military Personnel*. Research Triangle Park, NC: Research Triangle Institute.
- Cohen, S.B., Xanthopoulos, J.H., & Jones, G.K. (1986). An evaluation of available software for the analysis of complex survey data. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*. Washington, DC: American Statistical Association.
- Flyer, P., & Mohadjer, L. (1988). *The WESVAR Procedure*. Rockville, MD: Westat, Inc.
- Fuller, W.A. (1986). *PC CARP*. Ames, IA: Statistical Laboratory, Iowa State University.
- Lessler, J.T., & Kalsbeek, W.D. (1992). *Nonsampling Error in Surveys*. New York: John Wiley & Sons.
- Levy, P. S. (1977). Optimum allocation in stratified random network sampling for estimating the prevalence of attributes in rare populations. *Journal of the American Statistical Association*. 72, 768-763.
- National Institute on Drug Abuse. (1991). *National Household Survey on Drug Abuse: Main Findings 1990* (DHHS Publication No. ADM 91-1788). Rockville, MD: Author.
- National Institute on Drug Abuse. (1993). *The Washington, DC, Metropolitan Area Drug Study. Prevalence of Drug Use in the Washington, DC, Metropolitan Area Homeless and Transient Population: 1991* (Technical Report No. 2). Rockville, MD: Author.
- National Institute on Drug Abuse. (1994a). *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the DC Metropolitan Area Household and Nonhousehold Populations: 1991* (Technical Report No. 8). Rockville, MD: Author.
- National Institute on Drug Abuse. (1994b). *The Washington, DC, Metropolitan Area Drug Study: Prevalence of Drug Use in the Washington, DC, Metropolitan Area Institutionalized Population: 1991* (Technical Report No. 4). Rockville, MD: Author.

- National Institute on Drug Abuse. (1994c). *The Washington, DC, Metropolitan Area Homeless and Transient Population Study: 1991: Public Use Data Tape Documentation*. Rockville, MD: Author.
- Research Triangle Institute. (1990). *Software for SURvey DATA ANalysis (SUDAAN) Version 5.30*. Research Triangle Park, NC: Author.
- Substance Abuse and Mental Health Services Administration. (1993). *National Household Survey on Drug Abuse: Main Findings 1991* (DHHS Publication No. SMA 93-1980). Rockville, MD: Author.
- Substance Abuse and Mental Health Services Administration. (1995). *National Household Survey on Drug Abuse: Main Findings 1992* (DHHS Publication No. SMA-94-3012). Rockville, MD: Author.
- U.S. Bureau of the Census. (1982). *1980 Census of Population and Dwelling: Users' Guide* (PHC80-R1). Washington, DC: U.S. Government Printing Office.
- U.S. Bureau of the Census. (1991). *1990 Census of the Population and Housing: Summary Tape File 1 [machine-readable data files]*. Washington, DC: U.S. Bureau of the Census (Producer and Distributor).
- Wolter, K.M. (1985). *Introduction to Variance Estimation*. New York, NY: Springer-Verlag.

**Appendix E**

**ALPHABETICAL LIST OF VARIABLES**

<b>VARIABLE NAME</b>	<b>COLUMN LOCATION</b>	<b>VARIABLE LABEL</b>	<b>PAGE</b>
ALCMON	17-18	Alcohol - Past Month Use	4
ALCYR	19-20	Alcohol - Past Year Use	5
CATAGE	5-6	Age Category	4
CIGMON	21-22	Cigarettes - Past Month Use	5
CIGYR	23-24	Cigarettes - Past Year Use	5
COCMON	25-26	Cocaine - Past Month Use (Includes Crack Use)	5
COCYR	27-28	Cocaine - Past Year Use (Includes Crack Use)	5
CRKMON	29-30	Crack - Past Month Use	5
CRKYR	31-32	Crack - Past Year Use	5
DRUGOVL	79-80	Overlap of Past Year Crack, Heroin, & Needle Use	8
EDUCAT1C	7-8	Categorized and Collapsed Education	4
EMPSTAT2	9-10	Employment Status Recode	4
HALMON	33-34	Hallucinogens - Past Month Use	5
HALYR	35-36	Hallucinogens - Past Year Use	5
HERMON	37-38	Heroin - Past Month Use	5
HERYR	39-40	Heroin - Past Year Use	6
HVYDRK2	41-42	Drank Alcohol Heavily - Past 30 Days	6
IEMMON	43-44	Used Illicit Drugs Except Marijuana - Past Month	6
IEMYR	45-46	Used Illicit Drugs Except Marijuana - Past Year	6
INHMON	47-48	Inhalants - Past Month Use	6
INHYR	49-50	Inhalants - Past Year Use	6
IRSEX	11-12	Sex - Imputation Revised	4
MARSTAT	13-14	Marital Status - Categorized	4
MRJMON	51-52	Marijuana - Past Month Use (Includes Hashish)	6
MRJYR	53-54	Marijuana - Past Year Use (Includes Hashish)	6
MULTWT	91-105	Multiplicity Adjusted Weight for Combined File	7
NDSLHR	55-56	Needle Sharing Indicator	6
NEDFLAG3	57-58	Needle Use (Any Drug) - Lifetime Use	6
NEDMON	59-60	Needle Use (Any Drug) - Past Month Use	7
NEDYR3	61-62	Needle Use (Any Drug) - Past Year Use	7
OBSNUM	1-4	Observation Number	4
OVERLAP	81-82	Overlap of Surveys	8
PSYMON2	63-64	Any (Non-Medical Use) Psychotherapeutics - Past Month Use	7
PSYYR2	65-66	Any (Non-Medical Use) Psychotherapeutics - Past Year Use	7
RACE	15-16	Race/Hispanic - Origin Recode	4
STAMON	67-68	Sedatives/Tranquilizers/Analgesics - Past Month Use	7
STAYR	69-70	Sedatives/Tranquilizers/Analgesics - Past Year Use	7
STMMON	71-72	Stimulants - Past Month Use	7
STMYR	73-74	Stimulants - Past Year Use	7
STRATUM	106-112	Stratum for SUDAAN	8
SUMMON	75-76	Any Illicit Drug - Past Month Use	7
SUMYR	77-78	Any Illicit Drug - Past Year Use	7
SURVEY	83-84	Survey Type	8
SURVEY1	85-86	DC Household	8
SURVEY2	87-88	Homeless	8
SURVEY3	89-90	Institutionalized	8
WEIGHT	113-127	Original Analysis Weights Combined	9
WPSU	128-133	PSU for SUDAAN	9