RATIO ESTIMATION OF HARD-CORE DRUG USE

Douglas Wright, Joseph Gfroerer, Joan Epstein, Substance Abuse and Mental Health Services Administration
Douglas Wright, 5600 Fishers Lane, Room 16C-06, Rockville, Md. 20857

KEY WORDS: ratio estimation; benchmarking; hard core drug use.

1. INTRODUCTION

The need for accurate estimates of the size of the so called "hard core" drug using population is substantial. Regardless of how it is specifically defined, this population of heavy drug users is likely to need significant resources for treatment of their drug problem and associated medical and other problems. Hard core drug users have also been shown to be responsible for a disproportionate amount of crime.

This paper describes a method for estimating the prevalence of "hard core" drug use based on the National Household Survey on Drug Abuse (NHSDA) in conjunction with outside sources and the methodology of ratio estimation. In ratio estimation, one can often obtain a better estimate of a population total if there is a known population total for a related variable. Then the estimate of the total is $X' = (x/y)*Y$, where x is the variable of interest, y is the related variable, and Y is the known population total for the related variable.

Another way of describing this method is to say that it "inflates" (i.e., gives more weight to) the drug prevalence data from the NHSDA for populations with characteristics that are known to be related to hard core drug use but are also underestimated. In this case we know that the NHSDA undercounts arrestees and drug treatment populations, so we "ratio adjust" the NHSDA hard core drug use estimates upward to externally derived counts of arrestees and treatment clients that are believed to be accurate.

In survey sampling theory, ratio estimation is often associated the desire to improve the precision of an estimate. The ratio estimate will be better, in the sense that it will have a smaller variance, than the simple expansion estimator $X'' = \Sigma w_i x_i$ that is commonly used, when certain conditions are met. (See section 6. on precision of the estimates.)

However, in this application, we are less interested in variance reduction and more interested in bias reduction. Ratio estimates have been used for a number of years to adjust for nonresponse and to adjust to known population counts, often based on a census. This application represents an extension of those earlier uses to one in which we use known population counts to adjust NHSDA sample estimates for underreporting and undercoverage.

2. BACKGROUND

The NHSDA is the primary source of statistical information on the use of illegal drugs by the United States population. The respondent universe for the 1992 NHSDA is the civilian noninstitutionalized population aged 12 years old and older within the United States, including the residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories) and residents of civilian housing on military bases.

The 1992 survey employed a multistage area probability sample of 28,832 persons interviewed from January through December 1992. The screening and interview response rates were 95 percent and 83 percent, respectively, for an overall response rate of 79 percent. The household interview takes about an hour to complete and incorporates procedures designed to maximize honest reporting of illicit drug use, including the use of self-administered answer sheets. Further information on the methodology of the NHSDA is provided elsewhere (SAMHSA 1993a).

Comparison of NHSDA estimates with a variety of sociodemographic variables from independent sources (e.g., the Current Population Survey) typically has shown good agreement. However, estimating the number of hard core drug users has historically been a difficult problem. Household interview surveys such as the NHSDA were not designed for this type of estimation and are believed to be inadequate tools for measuring hard core drug use because of the low prevalence of the behavior and difficulties in accessing this population. Underreporting (survey participants who don't report their drug use) and undercoverage (inability to roster hard core drug users) also affects this estimation.

In comparing the results of NHSDA estimates to those from various administrative records systems (e.g., drug treatment program data, parole, probation, or arrest data from the FBI, etc.), the apparent underreporting of these types of characteristics by the sample respondents has been significant. Research has shown that underreporting of drug use increases as the reference period approaches the present and as the perceived social disapproval increases. This suggests that hard core drug use is underestimated more than casual drug use. The underestimates could also be the result of

undercoverage of the populations with these characteristics (Turner, Lessler, and Gfroerer 1992).

Various methods have been used to estimate hard core drug prevalence including capture-recapture techniques, truncated Poisson, and modeling methods generally (Rouse, Kozel, and Richards, (Eds.) 1985; Hser et al 1992). These methods have been based primarily on data from administrative records such as treatment admission data, essentially ignoring household survey data. Other methods have supplemented household survey data with other sources of data (e.g., arrest data) to construct composite estimates of hard core drug use (Wish 1990-91; Rhodes 1993).

There also has been significant research on various data collection methods that encourage honest responses to sensitive questions. Such methods include randomized response and nominative techniques (Rouse, Kozel, and Richards, 1985).

In the following discussion we want to focus on the ratio estimate's ability to correct for bias (in particular, the undercounting of hard drug users in the NHSDA) given a true population value of a related variable. To make the discussion more concrete, we will apply the estimation procedure to one measure of "hard core" drug use for 1992, the number of past year users of heroin. Results for other measures are presented in a table after the References section.

3. METHODOLOGY

The information that we wish to make use of is the count of the number of persons in treatment centers for drug abuse during the past year (1992) from the National Drug and Alcoholism Treatment Unit Survey (NDATUS) (SAMHSA 1993b) and the known count of the number of arrests (for any crime other than minor traffic violations) during the past year (1991) from the FBI Uniform Crime Reporting (Maguire, Pastore, and Flanagan 1993).

3.1. USING A SINGLE POPULATION COUNT

Let:

Nt = the estimated count of the number who received treatment for a drug problem during the past year derived from the NDATUS. The count was computed by multiplying the number of treatment slots times the average number of persons treated per year per slot, and includes an adjustment for multiple episodes by the same individual.

Na = the estimated count of the number of persons arrested during the past year. Na is calculated by taking the 1991 (latest available) FBI Uniform Crime Report estimated number of arrests -- 14,211,900, and dividing this by the average number of arrests per person arrested calculated from the NHSDA -- 1.46, resulting in an estimate of 9,734,178. Based on recent trends, the 1992 estimate would be expected to be slightly higher than the 1991 estimate.

The typical use of a ratio estimate occurs when the outside source fully overlaps the population of interest. Our situation is slightly different in that neither our treatment nor arrest population counts fully overlap the population of hard drug users. However, we can construct counts from the Census Bureau estimates for 1992 that are so precise at the national level that we can consider them to be population counts. With these we can develop counts that cover the population.

Given the Census estimate of the number of noninstitutionalized persons 12 and older, N = 205,713,000 (for July 1, 1992, the count at the midpoint of data collection for the NHSDA target population), we can form two pairs of counts that cover the population:

Number of persons in treatment: Nt = 1,789,000
Number of persons not in treatment:
N - Nt = 203,924,000

Number of persons arrested: Na = 9,722,671
Number of persons not arrested: N - Na = 195,990,329.

From these counts, two estimates of the number of persons using heroin during the past year (1992) are possible:

$H(t) = r(t)*Nt + r(t)^{*}*(N-Nt)$
$= (139,003/834,702) * 1,789,000$
$\quad + (171,136/200,656,309) * 203,924,000$
$= 471,844$ and
$H(a) = r(a)*Na + r(a)^{*}*(N-Na)$
$= (184,277/4,743,706) * 9,722,671$
$\quad + (125,861/196,747,306) * 195,990,329$
$= 503,070$, where

$r(t) = h_t/t$ = the estimated rate of hard core drug use (heroin, in this example) in population Nt. For heroin, it is the estimated number (from the NHSDA sample) in treatment and using heroin in the past year divided by the estimated number (from the sample) in treatment centers for drug use in the past year

$r(t)^{*}$ = the estimated rate (from the sample) of hard core drug use in population N-Nt.

$r(a) = h_a/a$ = the estimated rate (from the sample) of hard core drug use in population Na.

$r(a)^{*}$ = the estimated rate (from the sample) of hard core drug use in population N-Na.

The above estimates can be compared to the published "simple expansion" estimate for the number of past year users of heroin in 1992 of 323,000. [Actually, the published estimate is really more than a "simple expansion estimator" since NHSDA makes adjustments for nonresponse and benchmarks to known population Census totals.]

3.2. USING TWO KNOWN POPULATION COUNTS

Having the two separate estimates based on treatment and arrests counts raises the question, "Is there an alternate method that would make simultaneous use of both the treatment and arrest counts?"

The ideal situation when one has two variables, such as the number receiving treatment in the past year and the number arrested and booked in the past year, is to use known counts for the interior cells. In other words, we can make consistent estimates that make use of ratio estimation for each of the cells if we have the following matrix:

	Treatment	No Treatment
Arrested	N(11)	N(12)
Not Arrested	N(21)	N(22)

where N(11) is the known count of the number in treatment and arrested and booked in the past year, etc. [It is interesting to note that if we use estimates of the number in each cell based on the NHSDA sample multiplied by the sample cell estimates of prevalence, we obtain the NHSDA simple expansion estimate. The impetus for using external counts is that the sample counts tend to underestimate counts of those in treatment and those arrested and booked.]

In the earlier description of the calculation for heroin, all that was known at the national level were the marginals Nt, the number in treatment in the past year, and Na, the number arrested and booked in the past year. In the absence of known population counts, we proceed by developing independent sample estimates of one cell to estimate the remaining cells. To estimate one of the interior cells, we will use sample data from the 1990 Drug Services Research Survey (DSRS) in conjunction with data from the NHSDA and the estimated marginal count from NDATUS. DSRS is a national sample survey of treatment centers and records of discharged clients conducted in 1990. (Bigel 1992).

We wish to estimate the number of persons in treatment during 1992 who also were arrested and booked, N(11). Since we have the number of persons in treatment Nt, all we have to do is estimate the percent of those in treatment who also were arrested and booked in 1992.

From DSRS we estimate that the percent of those in treatment who were ever arrested was 77 percent. From NHSDA we estimate that the percent of those in treatment and ever arrested who were also arrested and booked in the past year was 52 percent. Therefore, we multiply .77 by .52 times the number in treatment 1,789,000 to obtain 716,315, the number in treatment and arrested and booked in the past year.

We can now obtain counts for the interior cells N(11), N(12), N(21), N(22) that are consistent with the marginal counts used earlier:

	Treatment	No Treatment	Total
Arrested	716,315	9,006,356	9,722,671
Not Arrested	1,072,685	194,917,644	195,990,329
Total	1,789,000	203,924,000	205,713,000

From the NHSDA, we obtain the corresponding ratios (i.e., prevalence rates) for heroin: r(11), r(12), r(21), and r(22).

	Treatment	No Treatment
Arrested	.308	.021
Not Arrested	.092	.00038

Then the estimate based on interior cells can be written as:
Estimate = H(t,a) = r(11)*N(11) + r(12)*N(12) + r(21)*N(21) + r(22)*N(22) = 587,966.

This estimate is larger than either of the estimates based on marginals. By analogy with stratification and poststratification, the interior-cells estimate is generally to be preferred to marginal estimates, especially if there are large differences in the usage ratios among the cells.

In two of the cells, the sample estimates of the counts significantly underestimate the known counts by a factor of 2 or more. For example, the estimated number of those arrested and booked in the past year but not in treatment (4,456,505) is less than half of the independent count of 9,006,356. The third cell, those in treatment but not arrested, is underestimated by a factor slightly less than 2. The last cell, those not in treatment and not arrested in the past year, is very close to the independent count, so that the impact of the ratio

estimate on this cell is minimal.

4. ASSUMPTIONS

One major assumption made in the above estimation methodology concerns the accuracy of our estimated ratios given that we expect underestimation of these hard drug populations from the household sample. A basic assumption being made in these ratios is that both numerator and denominator are being similarly underestimated. This would be the case, for example, if drug users underreport their drug use (or it is undercovered) at the same rate as the treatment population underreports their treatment and arrestees underreport being arrested. Taking the estimate of r(t), the expected value of h_t is assumed to equal $c_{ht}*H_t$, where c_{ht} is a constant and H_t is the true value. Similarly, the expected value of t is assumed to equal c_t*Nt, so that the expected value of r(t) equals approximately H_t/Nt when $c_{ht}=c_t$. (Another possible assumption is that $c_{ht}<c_t$ because some will assert that the NHSDA coverage of the heaviest users is poorest.)

For the complementary cell in the 2-cell estimate, the numerator of the ratio $r(t)^*$, the number of heroin users that are not in treatment, is probably underestimated. But the sample estimate of the denominator will generally not be an underestimate, so that the impact of ratio estimation on this cell is minimal.

Similarly, in the 4-cell estimate, the cell estimate of those who have not been in treatment or arrested and booked in the past year is very similar to the independent count for that cell. Here also, we expect that the assumption of equal underestimating of the numerator and denominator would not hold. Therefore, the ratio estimate still would underestimate this cell for any measure of hard drug use.

What is the impact of this kind of estimation on other non-hard-core drugs? Generally speaking, it is not as dramatic as with the above hard-core drugs. We calculated the impact on use of marijuana in the past year (variable MRJYR). The simple expansion estimator was 17,400,273, while the ratio estimate based on four cells was 19,461,280. The latter estimate is only 12 percent larger. The reason for this is that marijuana is used more widely in the population, and most of the users fall in the no treatment/not arrested cell. Therefore, this cell (13,644,235 users) dominates the estimate. The relative differences in prevalence rates among cells also are not as dramatic as with the hard-core drugs.

5. PRECISION OF ESTIMATES

Estimates of variance were calculated for each of the above estimates using SUDAAN - a package that calculates the variance of complex sample surveys using Taylor Series (RTI 1992). We used the ratio estimation procedure with poststratification weights. It was assumed that the independent counts were estimated without error.

The estimated standard error for the ratio estimate of heroin was very similar to the estimated standard error for the simple expansion estimator, approximately 106,000. Even though the cell count of those who have neither been in treatment nor arrested and booked in the past year is the largest of the four cells, its contribution both to the estimate and to the estimate of variance was relatively small because of the estimated low prevalence in that cell. The 95 percent (2 σ) confidence interval for the estimate of past year users of heroin was 587,966 plus or minus 212,000.

6. COMPARISON WITH OTHER METHODS

Previous national estimates of hard core drug use have used widely varying methods. Estimates of heroin prevalence published by the National Institute on Drug Abuse in the 1970s relied on a small number of locally derived prevalence estimates that were projected to the entire nation using available heroin problem indicators available in other locations (Person, Retka, and Woodward 1977). These estimates of the number of heroin addicts ranged from 584,000 in 1974 to 420,000 in 1979. However, these are not comparable to estimates of any past year heroin use, because they do not include casual heroin users.

A nominative method of estimating heroin prevalence from the NHSDA produced an estimate of 1.9 million past year heroin users in 1982 (Rouse, Kozel, and Richards (Eds.) 1985). A recent estimate of 658,000 weekly heroin users in 1990 was derived from a "synthetic estimation" procedure that involved combining multiple data sources under various assumptions (Rhodes 1993). This same methodology was used to derive an estimate of 2.1 million weekly cocaine users in 1991. These recent synthetic estimates represent the most rigorous attempts to utilize multiple sources of data in estimating hard core drug use prevalence.

While there are many differences between the synthetic estimation model and the NHSDA ratio estimation, the large discrepancies in estimates from the two methods are largely explained by the assumptions made regarding the arrestee population. The synthetic model relied heavily on drug prevalence data from the Drug Use Forecasting System (not a representative sample of arrestees), resulting in an estimated 1.8 million weekly cocaine users (more than 80 percent of the total estimate) and 500,000 weekly heroin users (more than 80 percent of the total estimate) among arrestees. By contrast, the ratio estimation method relies

more heavily on NHSDA drug prevalence data for arrestees, and resulted in an estimated 329,626 weekly cocaine users and 414,265 past year heroin users among arrestees.

A complete evaluation and comparison of the ratio estimation procedure with other methods of estimating hard core drug use is beyond the scope of this paper. However, we can make some overall statements about ratio estimation.

- Ratio estimation does not fully account for underreporting and undercoverage in the NHSDA. In particular, for the population not arrested and not in treatment, the method does not adjust for underreporting at all. Thus, we consider these estimates of hard core drug use to be improvements on the generally published NHSDA estimates (using the simple expansion estimator) but still conservative estimates.

- The ratio estimation model, as applied in this case, relies primarily on regularly updated and consistently collected data from the NHSDA, NDATUS, and UCR, and a relatively small number of easily understood assumptions. Thus, it is likely to be able to provide more reliable trend information (given constant levels of underreporting) than the previously used methods which rely more heavily on assumptions that could change over time.

- Because it relies primarily on the NHSDA sample design and weighting, it is possible to develop estimates of the variances of ratio-adjusted estimates. This is generally not possible in the methods previously used.

7. POSSIBLE FUTURE RESEARCH/ APPLICATIONS

There are three primary areas for further investigation. One is in the population counts. Another is the assumptions made about the ratios used. The third involves a search for "unbiased" methods to estimate the ratio.

1. It would be useful to explore the development of more accurate estimates for the four cell counts or of alternative counts based on different variables. Estimating the counts used in this paper necessitated using multiple sources to make the counts comparable to what is collected by NHSDA. Generally, this is best accomplished by coordinating the questions on the NHSDA and other surveys with those systems used to develop administrative counts so that the definitions are as consistent as possible. Coordination of item wording among surveys will at a minimum make it possible to compare estimates across surveys. For 1994, the NHSDA question on being in treatment has been changed to agree exactly with the definition used in NDATUS.

Since it is known that age and race are major correlates of the rate of drug usage, yet another improvement would be to seek to find a source or a method of estimation that could provide further age/race breakouts to the treatment/arrest cell counts.

2. In the area of assumptions, where possible one can compare the distributions of persons for a variable used in the cross-classification based on the NHSDA to those of the population frames to see if they are similar. For example, we can compare the distribution of those in treatment from the NHSDA to the distribution of the population values from NDATUS that are available by age, race, and sex. Another possibility is the introduction of additional weights reflecting the proportion of the year that a person is in treatment or living in a household. This would serve to increase the size of the populations that are not year-round household residents.

3. With respect to the instrument, one could perhaps try to introduce methodology that would result in less undercounting of the variables that form the ratios: heroin, treatment, arrested and booked, etc., possibly using multiplicity methods or nominative techniques or using some new method, like hair tests if the methodology proves to be feasible, to confirm drug use or nonuse.

8. REFERENCES

BIGEL INSTITUTE FOR HEALTH POLICY, Brandeis University (1992). *Drug Services Research Survey - Final Report: Phase II.* Contract No. 271-90-8319/1. Submitted to the National Institute on Drug Abuse.

HSER, Y.I., ANGLIN, M.D., WICKENS, T.D., BRECHT, M.L., and HOMER, J.(1992). *Techniques for the Estimation of Illicit Drug-Use Prevalence: An Overview of Relevant Issues.* National Institute of Justice Publication No.(NCJ)133786.

MAGUIRE, K., PASTORE, A.L., and FLANAGAN, T.J. eds., *Sourcebook of Criminal Justice Statistics.* U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC: USGPO, 1993.

PERSON, P.H., RETKA, R.L., and WOODWARD, A.J. (1977). *A Method for Estimating Heroin Use Prevalence.* DHEW Publication No.(ADM)77-439. Washington, DC: Supt. of Docs., U.S. Government Printing Office, 1977.

RESEARCH TRIANGLE INSTITUTE (1992). *SUDAAN Users Manual.* Research Triangle Park, N.C.: Research Triangle Institute.

RHODES, W. (1993). Synthetic Estimation Applied to the Prevalence of Drug Use. *Journal of Drug Issues, 1993.*

ROUSE, B., KOZEL, N., AND RICHARDS, L. (Eds.) (1985)., *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity*, pp. 94-103. NIDA Research Monograph 57, Department of Health and Human Services Publication No. (ADM) 85-1402. Washington, DC:U.S. Government Printing Office.

SUBSTANCE ABUSE AND MENTAL HEALTH SERVICES ADMINISTRATION (1993a). *National Household Survey on Drug Abuse: Population Estimates, 1992.* DHHS Publication No.(SMA)93-2053. Washington, DC: Supt. of Docs., U.S. Government Printing Office.

SUBSTANCE ABUSE AND MENTAL HEALTH SERVICES ADMINISTRATION (1993b). *National Drug and Alcohol Treatment Unit Survey (NDATUS) 1991 Main Findings Report.* DHHS Publication No.(SMA)93-2007. Rockville, MD:U.S. Department of Health and Human Services.

TURNER, C.F., LESSLER, J.T., and GFROERER, J.C. (1992). *Survey Measurement of Drug Use: Methodological Studies.* National Institute on Drug Abuse. DHHS Publication No.(ADM)92-1929.

WISH, E.D. (1990-1991). Drug Policy in the 1990s: Insights from new data on arrestees. *Int J Addict.* 1990-1991;25:377-409.

Comparison of Ratio Estimate to Standard Estimate for Various Drug Use Measures

Drug Usage	Treatment/arrest ratio estimate	Standard estimate
past year heroin	587,966	323,000
weekly use of cocaine	829,017	642,221
past year dependence on any drug	2,869,242	2,104,508
past year needle use	1,019,165	659,292
past year use of marijuana	19,461,280	17,400,273