ESTIMATING DRUG ABUSE EPISODES FROM A NONRANDOM SAMPLE OF HOSPITALS

Richard E. Gruberg, CSR Inc., Arthur L. Hughes, NIDA
Richard E. Gruberg, CSR Inc., 1400 Eye St. N.W., Washington D.C. 20005

Key Words: Hospital Emergency Rooms, Drug Abuse Data, Empirical Study

I. Introduction

Data on hospital emergency room episodes involving the abuse of licit and illicit drugs are collected by the Drug Abuse Warning Network (DAWN), which is a voluntary reporting system sponsored by the National Institute on Drug Abuse (NIDA). Some of the major objectives of DAWN are: 1) to identify drugs or substances that are currently being abused and 2) to provide data for national and local area drug abuse policy and planning. For the purpose of reporting to DAWN, drug abuse is defined as the nonmedical use of a drug or substance for psychic effect, for suicide attempt or gesture, or because of dependence (NIDA, 1991). DAWN data are collected and transmitted in machine-readable form to NIDA monthly. Quarterly and annual weighted estimates of the total number of emergency room visits involving drug abuse in general (referred to as episodes), and the total number of visits involving the abuse of a specific drug (referred to as a drug mention) are currently produced for the nation as a whole as well as for 21 metropolitan areas and an area called the "National Panel" which is associated with the balance of the coterminous U.S. These estimates are based on data from a new sample recently implemented by NIDA, and on auxilliary data obtained from the American Hospital Association (AHA).

When DAWN was originally implemented in the early 1970's, a random sample of hospitals was chosen (DEA, 1974). The sample design consisted of 100 percent sampling in 20 metropolitan areas and less than 100 percent sampling in three metropolitan areas (L.A., N.Y., and Chicago) and in a national panel. The national panel was stratified by hospital bed size (0-99, 100-299, and 300+). This sample, however, gradually deteriorated over the years, as a result of attrition and nonrandom replacement, making the development of weighted estimates problematic (NIDA, 1982). NIDA assumed responsibility for DAWN in 1980 and deemed the sample appropriate only for presentation of raw (unweighted) data. Trend analyses could be conducted only by using panels of consistently reporting hospital emergency rooms (ERs). Early in its sponsorship of DAWN, NIDA adopted the objective of implementing a new statistical sample of hospitals that could be used to produce representative estimates for the nation as well as for separate metropolitan areas. A panel of experts was assembled to develop a design for the new statistical sample. Once the design was determined and the units selected, the new sample required the recruitment of approximately 300 new hospitals. Included in the design was a sample maintenance plan to ensure the statistical integrity of the new sample over time.

The new sample design is stratified by metropolitan area and, within area, by location (central city, outside central city), and within location, by the presence of an organized outpatient department and/or an alcohol/chemical dependency inpatient unit (both, one only, or neither). Stratification by location is not included in the National Panel. Within each sampling stratum, hospitals are selected by simple random sampling. Once a hospital is selected, all of its ERs are included. Hospitals having more than 80,000 ER visits per year are selected with certainty.

Implementation of the new statistical sample, which started in 1986, has enabled NIDA to develop weighted, representative estimates of ER episodes and mentions. NIDA's 1990 Annual Report was the first DAWN publication to be based wholly on weighted estimates from the new sample, and these estimates are available starting with data from year 1988. However, representative estimates from DAWN for years prior to 1988 have not been available, owing to the original sample's loss of statistical integrity and incomplete implementation of the new sample until 1988.

Last year, the performance of several estimators was evaluated for the new statistical sample using a simulation study. The combined ratio estimator was found to perform well for this sample with minimal bias (Hughes, et al., 1991). The purpose of this paper is to present and evaluate weighting procedures for producing representative estimates from the old DAWN sample data (referred to as ex post facto estimates). Such weights will be used to generate both estimates and variance estimates from DAWN for years 1977 to 1987.

II. Characteristics of the Old and New Samples

To compare the characteristics of the old and new DAWN samples, the retention of old sample units was examined over time and the amount of overlap between old and new sample units was determined. Both samples were also evaluated in terms of coverage, expressed in terms of the fraction

of total ER visits in the population captured by the sample.

Rate of Retention for Original DAWN Sample Units

Phoenix has the highest retention rate (defined in terms of the percentage of original old DAWN sample units remaining as of 1989) with 94%, followed by Atlanta (76%), Miami (75%), and Philadelphia (69%). San Francisco, Minneapolis, Dallas, New York City, Los Angeles, and the National Panel have the worst retention rates—all less than fifty percent.

Considering the percentage of original DAWN reporters contained in the 1989 new DAWN sample, Denver has the highest percentage (79%), followed by Seattle (76%), Philadelphia (76%), Buffalo (73%), and Phoenix (72%). Because the differences between the original sample and new sample are smallest in these locations, one would expect estimators that perform well for the new sample to also perform well for the old sample here.

Coverage of the Old and New Samples

For many metropolitan areas the new sample captures a significantly smaller percentage of total visits in the non-central city areas. The new sample also captures a smaller percentage of the total visits for facilities belonging to the National Panel (2 percent versus 5 percent).

For the central city areas, the pattern is less well defined. Comparing the new sample coverage in 1989 to the old sample coverage in 1988, the coverage is considerably higher in the 1989 new sample for Atlanta (87 % vs. 64 %), Los Angeles (60 % vs. 40 %), and Miami (69 % vs. 27 %). In Dallas, Detroit, New York, Philadelphia, and Seattle, the coverage is lower in the new sample in 1989. For most of the metropolitan areas, the coverage in the central cities is comparable or a little higher in the new sample in 1989. In terms of the total visits data, the new sample allocates a larger fraction of its overall sample to the central city locations (this resulted from an optimal allocation scheme for the new sample based on the estimated variance of total drug episodes in each sampling stratum).

III. Methodology

Selection of Models using the Overlap Period

Hospitals belonging only to the old sample were not dropped from the DAWN reporting system immediately upon implementation of the new sample. Most such facilities remained in the system until 1989. This created an overlap period between 1988 and 1989 during which DAWN contained both "old" and "new" reporters. In particular, from the fourth quarter of 1988 to the second quarter of 1989, the new sample was virtually fully implemented and the old sample was still in place. This overlap period was used to evaluate various procedures for weighting the old sample data (from 1977 to 1987): for each particular method under consideration, weighted estimates were generated for the overlap period using the old sample data, and then compared to weighted estimates for the same period obtained from the new statistical DAWN sample. The performance of each ex post facto model estimator was then measured in terms of how well estimates derived from them agreed with the new estimates for this period. For the purposes of this study, the new published estimates for this period were taken to be "true" values, and the discrepancies between old and new estimates have been expressed in terms of relative bias. These new estimates have sampling errors (as well as nonsampling errors) associated with them, and, obviously, there is no way of ensuring that they are in fact more accurate than the old estimates, although we consider this to be a reasonable assumption. At the very least, however, this approach will enhance the trendability of the ex post facto estimates. Since the overlap period corresponds to the period of maximum sample deterioration for the old sample, we expect that in areas where successful weighting models are found for the overlap, such models should perform well for previous years as well.

Old DAWN sample units which also belonged to the new DAWN sample were assigned to a sampling stratum based upon their stratum in the new sample. For those old sample units not belonging to the new sample, stratum assignments were made using PROC NEIGHBOR, a multivariate classification procedure in SAS, based upon a nearest neighbor discriminant analysis. The variables used for this analysis were cocaine, heroin, and marijuana ER mentions.

For this overlap period, the performance of various weighting procedures was considered for nine drug categories: total episodes, total mentions, heroin-morphine mentions, cocaine mentions, marijuana-hashish mentions, acetaminophen mentions, aspirin mentions, diazepam mentions, and phenobarbital mentions. These drug categories were selected, in part, based upon the drugs mentioned in the objectives stated in the National Drug Control Strategy—a 1991 White House publication which specifies targets for reductions in the total number of drug related medical emergencies, as well as emergency room mentions of cocaine, heroin, and other "dangerous drugs." In addition, other drugs were selected to ensure that some over-the-counter and prescription drugs were included in the analysis.

For each ex post facto estimator considered, the relative bias in each of the nine drug categories was calculated in each of DAWN's 21 metropolitan areas, as well as in the National Panel. With the

exception of the two categories, total episodes and total mentions, these relative biases were assigned a weight between zero and one, by taking the published estimate for the drug category for the overlap period, and dividing by the sum of the published estimates for all seven drug categories. The relative biases for the total episodes and total mentions categories were assigned a weight equal to unity. The weighted absolute relative biases were then added together within each metropolitan area to give a total relative bias for the region. This ensured that large biases occurring in instances where the published number of drug mentions is very small did not unduly influence the choice of estimator. For example, the estimated number of ER visits involving phenobarbital abuse for Newark during the overlap period, based on the published estimates, is ten, while the estimated number of ER visits involving cocaine abuse is close to 3600. Clearly, other things being equal, a model which estimates the number of phenobarbital mentions to be 5 (a 50 percent "bias") and cocaine mentions to be 3601 (less than one percent "bias") is superior to a model which estimates the number of phenobarbital mentions to be 9 (a 10 percent "bias") and the number of cocaine mentions to be 2500 (a 30 percent "bias"), even though the sum of the biases in the first case is smaller.

Weighting Procedures Considered

This evaluation of the performance of various weighting procedures for the old DAWN sample data focused on the following weighting characteristics:

1. Benchmarking

In DAWN a benchmark adjustment can be made to all survey estimates by producing an estimate from the sample for total ER visits (i.e., all visits, not just those related to drug abuse), and comparing this information to ER visits data available from AHA. The ratio of an AHA ER visits total to a DAWN sample estimate for ER visits constitutes a benchmark adjustment. This benchmarking was examined at the metropolitan area level and at the stratum level within metropolitan area. When performed at the DAWN area level, the procedure produces a "combined ratio" estimate. At the stratum level, the procedure produces a "separate ratio" estimator. Benchmarking the sample estimate should improve the quality of the estimate whenever there is a strong relationship between the parameter being measured (i.e., total episodes, total mentions, or total mentions for a specific drug), and the benchmarking parameter—in this case, ER visits.

Generally speaking, the separate ratio estimator will outperform the combined ratio estimator both in terms of bias and variance, although variance estimates for the separate ratio estimator may be biased if sample sizes in some strata are exceedingly small (Cochran, 1977). It should be noted, however, that each ex post facto estimator is evaluated in terms of how close it comes to matching the estimate produced from new DAWN sample, which uses a combined ratio estimator. Hence, one might expect the combined ratio estimation procedure to outperform the separate ratio procedure.

2. Equal versus Unequal Selection Probabilities

Applying PPS weights to sample units not selected by a PPS procedure may improve the quality of estimates in instances where the characteristics of smaller hospitals are more representative of the stratum universe. For the ex post facto estimates, PPS weighting was evaluated by assigning to each sample unit a weight equivalent to what it would have been given had it been selected according to a PPS procedure. The size measure parameters considered were ER visits obtained from American Hospital Association (AHA) files and ER visits obtained from DAWN.

3. DAWN versus AHA for ER Visits Information

As discussed above, weights based on PPS selection probabilities can be generated for the old DAWN sample data, based on either DAWN or AHA ER visits information. The performance of both size measures was evaluated. Benchmarking may also be based on either DAWN or AHA visits data. Again, both cases were considered.

4. Scaling the Data Based on Ratio of Episodes to Visits

This procedure involves scaling the old DAWN sample data to account for the fact that the ratio of total episodes to visits in a given stratum may differ considerably from the old sample to the new sample. Multiplying the response values of the old sample data in each stratum by the ratio of episodes to visits in the new sample, and then dividing by the ratio of episodes to visits in the old sample, will scale the old sample data so that they are more consistent with the characteristics of the new sample.

In a sense, this procedure is similar to a benchmarking procedure. The benchmarking procedures discussed in (1), above, however, cannot make use of episodes information as it is not available from the AHA. With this scaling procedure, however, the relationship between total stratum episodes and visits may be determined from the new DAWN data only for the overlap period. Hence, to the extent that the relationship between episodes to visits changes from year to year for hospitals belonging to the new DAWN sample, this scaling procedure may be of limited use.

All together, 20 weighting schemes were chosen to be evaluated, based upon the considerations discussed above. Models based on regression and other similar techniques were not considered owing to the quantity of drug categories for which estimates are needed. These weighting models were derived from the following five standard survey estimators for population totals: the combined ratio estimator, the separate ratio ratio estimator, a simple stratified sample estimator with PPS sampling weights (referred to as the PPS estimator), and the combined ratio and separate ratio estimators with PPS sampling weights. The formulae for these estimators are given below.

1. Combined Ratio Estimator

$$\hat{Y}_{CR} = \sum_{j=1}^{L}\sum_{i=1}^{n_j} X \bullet \left(\sum_{j}^{L}\sum_{i}^{n_j} \frac{N_j}{n_j} x_i \right)^{-1} \bullet \frac{N_j}{n_j} \bullet y_i$$

2. Separate Ratio Estimator

$$\hat{Y}_{SR} = \sum_{j=1}^{L}\sum_{i=1}^{n_j} \left(\frac{X_j^{AHA}}{\sum_{i=1}^{n_j} x_i} \right) \bullet y_i$$

3. PPS

$$\hat{Y}_{PPS} = \sum_{j=1}^{L}\sum_{i=1}^{n_j} \left(\frac{X_j^{AHA}}{n_j x_i} \right) \bullet y_i$$

4. Separate Ratio PPS

$$\hat{Y}_{SR,PPS} = \sum_{j=1}^{L}\sum_{i=1}^{n_j} \left(\frac{X_j^{AHA}}{\sum_{i=1}^{n_j} \frac{N_j}{n_j} x_i} \right) \left(\frac{X_j^{AHA}}{n_j x_i} \right) \bullet y_i$$

5. Combined Ratio PPS

$$\hat{Y}_{CR,PPS} = \sum_{j=1}^{L}\sum_{i=1}^{n_j} X \bullet \left(\sum_{j}^{L}\sum_{i}^{n_j} \frac{N_j}{n_j} x_i \right)^{-1} \bullet \left(\frac{X_j^{AHA}}{n_j x_i} \right) \bullet y_i$$

where

j = the sampling stratum,
i = the hospital sample unit,
L= the total number of strata in the region,
X=number of ER visits in the population,
x= number of ER visits in the sample,
N=number of hospital units in the population,
n= number of hospital units in the sample,
y = response value ,
$\hat{Y}$ = estimate of population total for a particular response characteristic in a given metropolitan area.

These five estimators were evaluated using both AHA and DAWN data for the size measures in the PPS weighting and for the denominator of the benchmarking factor. The estimators were also evaluated with and without first scaling the raw data to account for differences between the old and new sample in the ratio of total emergency room drug related episodes to total emergency visits in each sampling stratum. Hence 20 models (5x2x2) models were evaluated.

IV. Findings

National Estimates

Representative national estimates of total drug episodes and total drug mentions for the overlap period for the coterminous U.S. were generated from the ex post facto modeling procedures by, first, finding the best weighting model for each metropolitan area (based upon the sum of the weighted absolute relative biases across all drug categories), and then adding these estimates together. Using this approach, the ex post facto modeling procedure yielded estimates from the old sample for the overlap period with relative biases as listed in the first column of Table 1.

Another approach to generating national estimates would be simply to ignore the stratification by metropolitan area and determine which estimator produces the best results at the aggregated level. Using this later approach, the best estimator was found to be the stratified estimator with PPS weighting, with the size measure based on DAWN reported visits information. The relative biases for this approach are given in second column in the table above. The relative biases across drug categories are less well behaved here when compared to the "best model" approach (first column). Another drawback to this one model approach is that it is less compatible with the objective of producing metropolitan area level estimates, since the model may not perform well for certain drugs in particular metropolitan areas.

Table 1. Relative Bias of Estimates

Drug Category	Best Regional Models	Best Aggregated Model
Total Episodes	3 %	1 %
Total Mentions	2 %	2 %
Cocaine	9 %	12 %
Heroin/Morphine	5 %	-1 %
Marijuana	9 %	22 %
Aspirin	9 %	-10 %
Acetaminophen	8 %	-5 %
Diazepam	5 %	6 %
Phenobarbital	2 %	5 %

Using the best models found for each metropolitan area, weighted and unweighted percent distributions of total episodes by age, gender, race/ethnicity, and hospital location were compared for the old and new sample, as shown in Table 2. The most notable difference between the old and new samples occurred in the weighted distribution by race/ethnicity. This may be due to the fact that the new sample is more concentrated in the central cities where more blacks reside. It is interesting to note that the old sample weighting does balance the rather different unweighted distributions between the old and new samples by hospital location.

Table 2. Demographics of Old and New Sample

	Percent Distribution			
	Old Sample		New Sample	
Age				
6-17	13	(9)	14	(8)
18-25	28	(27)	27	(26)
26-34	32	(32)	31	(36)
35+	26	(28)	27	(29)
Gender				
Male	49	(57)	48	(58)
Female	49	(42)	51	(41)
Race/ethnicity[1]				
White	54	(39)	58	(37)
Black	28	(40)	23	(43)
Hispanic	7	(11)	9	(12)
Location				
Central City	31	(62)	30	(70)
Suburbs	19	(27)	19	(24)
National Panel	50	(11)	51	(6)

Numbers in parentheses are unweighted.
[1] "Other" race/ethnicity is excluded.

Metropolitan Area Estimates

At the metropolitan area level, ten areas had the least relative bias when the separate ratio estimator was used (using either AHA or DAWN information in the denominator for the benchmark adjustment), four areas had best results with the combined ratio estimator, four areas with the PPS estimator, two regions with the combined ratio PPS estimator, and two areas with the separate ratio PPS estimator. Table 3 lists the best performing model in each metropolitan area, and gives the total weighted absolute relative bias.

In five metropolitan areas (Minneapolis, New Orleans, New York, Philadelphia, and Washington D.C.) useful ex post facto estimates were obtained only after scaling the raw data to account for stratum differences in the ratio of episodes to visits from the old sample to the new. Incorporating such an adjustment in the estimator limits its utility since there is no way of knowing whether the adjustment factors are valid for time periods other than the overlap period. In addition, estimation from the old sample was particularly problematic for the St. Louis and San Francisco metropolitan areas. With St. Louis, the problem may stem from the fact that one very large hospital having a considerable number of drug abuse ER episodes which was included in the old sample is not included in the new sample. In the case of San Francisco, the boundary definitions for the metropolitan area changed prior to the implementation of the new sample.

Three metropolitan areas had ex post facto estimates with relative biases of less than five percent in each of the nine drug categories: Denver (using the separate ratio estimator), Phoenix (using the combined ratio estimator), and San Diego (using the separate ratio estimator). Considering only two estimation categories, total drug episodes and total drug mentions, 13 metropolitan areas had estimates with relative biases of five percent or less in both categories without scaling the raw data, and 11 areas had estimates with relative biases of three percent in both categories. Including regions where the data were scaled, 18 metropolitan areas had estimates with relative biases of 5 percent or less in both categories.

It is interesting to note that for the National Panel, relative biases for cocaine, heroin/morphine, and marijuana, were considerably higher than in most other areas (22%, 33 %, and 43 % respectively), while the relative biases for acetaminophen, diazepam, and phenobarbital were generally lower (all 6% respectively). This is consistent with the fact that the National Panel contains a large percentage of nonurban hospitals where illicit drug abuse is less common than in urban areas. The impact of the nonrandom sample here is greater since this creates a high degree of population variability for these drugs.

Table 3. Best Models Found by Metropolitan Area

Area	Model	Tot. Wgt. Abs. Rel. Bias
Atlanta	S.R.	.31
Baltimore	S.R.	.06
Boston	C.R.	.16
Buffalo	S.R.	.14
Chicago	S.R.	.11
Dallas	S.R.,PPS	.04
Denver	S.R.	.02
Detroit	PPS	.16
Los Angeles	S.R.,PPS	.16
Miami	PPS	.06
Minneapolis	S.R.(scaled)	.10
New Orleans	C.R.(scaled)	.09
New York	S.R. (scaled)	.11
Newark	C.R.	.08
Philadelphia	S.R. (scaled)	.14
Phoenix	C.R.	.07
St. Louis	C.R.,PPS	.66
San Diego	S.R.	.03
San Francisco	C.R.	.54
Seattle	PPS	.08
Wash., D.C.	S.R. (scaled)	.10
National Panel	PPS	.22

C.R. stands for combined ratio, S.R. stands for separate ratio.

V. Summary of Findings

Weighting models producing a five percent relative bias or less in all nine estimation categories (total episodes; total mentions; and mentions of cocaine, heroin-morphine, marijuana-hashish, aspirin, acetaminophen, diazepam, and phenobarbital) using original DAWN sample data were found for three of the metropolitan areas in DAWN for the overlap period (fourth quarter 1988 through second quarter 1989). When considering only total episodes and total mentions, 18 areas had at least one model producing estimates with relative biases of 5 percent or less in both categories. For national estimates based on the approach using the best model found in each region, the relative biases of the ex post facto estimates were less than 10 percent in all drug categories. For national estimates based on application of the same model for all metropolitan areas (i.e., ignoring stratification by area), a stratified PPS weighting model was found to produce the best results for the nine drug categories. This approach, however, was found to produce inconsistent results when applied across all metropolitan areas and drug categories investigated.

VI. Recommendations for Future Work

- For this study, the new sample estimates were assumed to be "true." The impact of the sampling error of these estimates upon the analysis will be evaluated.
- Better weighting models will be pursued in problem areas, such as St. Louis and San Francisco. If no improvement can be made, such areas may be incorporated into the National Panel, therefore not allowing for separate publication.
- Short-term trends based on estimates from the old and new samples will be compared for the 3-quarter overlap period.

REFERENCES

[1] Cochran, W., (1977), Sampling Techniques, 3rd Edition, John Wiley and Sons, New York: pp. 164-169.

[2] Drug Enforcement Administration, (1974), DAWN 2 Analysis, Phase II Report, July 1973-March 1974: pp. i-xi.

[3] National Institute on Drug Abuse (1982), Annual Data 1981, Series I, Number 1, DHHS no. (ADM) 82-1227: pp. 1-3.

[4] National Institute on Drug Abuse (1991), Annual Emergency Room Data 1990, Series I, Number 10A, DHHS No. (ADM) 91-1839: pp. 1-13, 115-116.

[5] Hughes, A.L., Elliott, S.D., Colliver, J.D., and Gruberg, R.E. (1991). "Estimating Drug Abuse Episodes from a Sample of Hospital Emergency Rooms," American Statistical Association, Proceedings of the Section of Survey Research Methods, 1991, pp. 326-331.