Person-pair Sampling Weight Calibration using the Generalized Exponential Model for The National Household Survey on Drug Abuse

M.A. Penne, P. Chen, and A.C. Singh, Research Triangle Institute
M.A. Penne, Statistics Research Division, RTI, P.O. Box 12194,
Research Triangle Park, NC 27709 penne@rti.org

Key words: Weight Calibration; Extreme Weights; Bound Restrictions; Generalized Exponential Model; Multiplicity Factors

1. The 1999 NHSDA: a brief description

The National Household Survey on Drug Abuse (NHSDA) is designed to produce estimates of prevalence of use, the number of users and other measures related to illicit drugs, alcohol, cigarettes and other forms of tobacco. The respondent universe is the civilian, noninstitutionalized population aged 12 years or older residing within the United States and the District of Columbia. Consistent with the NHSDA designs since 1991, the 1999 NHSDA universe included residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories, and group homes), residents of Alaska and Hawaii, and civilians residing on military bases. Survey coverage before the 1991 NHSDA was limited to residents of the coterminous 48 states and it excluded residents of group quarters and all persons (including civilians) living on military bases. Persons excluded from the 1999 universe included those with no fixed household (HH) address (e.g., homeless transients not in shelters) and residents of institutional group quarters, such as jails and hospitals.

The 1999 survey marks the first year in which all interviewing was conducted using computer-assisted interviewing (CAI) methods as well as a transition to improved state estimates based on minimum sample sizes in all states. In order to obtain the required precision at the state level the sample size was increased to 70,000, approximately three times greater than previous NHSDAs. This large sample allowed the continuation of reporting demographic subgroups at the national level with adequate precision without the need to oversample specially targeted racial/ethnic demographics, as has been required in the past.

To help further facilitate the improved state estimates, a coordinated 5-year sample design was developed starting in 1999. This design implemented a 50% overlap in first-stage units between each two successive years from 1999 through 2003. This was intended to increase the precision of estimates in year-to-year trend analyses because of the expected positive correlation resulting from the overlapping sample between successive NHSDA years.

The 1999-2003 design provides for estimates by state in all 50 states plus the District of Columbia. States may therefore be viewed as the first level of stratification as well as a reporting variable. Eight states, referred to as the "big" states,[1] had a sample designed to yield 3,600 to 4,630 respondents per state for the 1999 survey. This sample size was considered adequate to support direct state estimates. The remaining 43 states,[2] had a sample designed to yield 900 respondents per state in the 1999 survey. In these 43 states, adequate data were available to support reliable state estimates based on small area estimation methodology.

Within each state, field interviewer (FI) regions were formed. Based on a composite size measure, states were geographically partitioned into roughly equal size regions. In other words, regions were formed such that each area yielded, in expectation, roughly the same number of interviews during each data collection period, thus distributing the workload equally among NHSDA interviewers. The smaller states were partitioned into 12 FI regions, whereas the eight "big" states were divided into 48 regions. Therefore, the partitioning of the United States resulted in the formation of a total of 900 FI regions.

For the first stage of sampling, each of the FI regions was partitioned into noncompact clusters[3] of dwelling units by aggregating adjacent Census blocks. Consistent with the terminology used in previous NHSDAs, these geographic clusters of blocks are referred to as *segments*. Sample segments were drawn with probabilities proportional to composite size measures using Chromy's algorithm (Chromy, 1981).

---

[1]For the 1999-2003 NHSDAs, the "big" states are California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas.

[2]For reporting and stratification purposes, the District of Columbia is treated the same as a state and no distinction is made in the discussion.

[3]Noncompact clusters (selection from a list) differ from compact clusters in that not all units within the cluster are included in the sample. While compact cluster designs are less costly and more stable, a noncompact cluster design was used because it provides for greater heterogeneity of dwellings within the sample. Also, social interaction (contagion) among neighboring dwellings is sometimes introduced with compact clusters (Kish 1965).

Two segments are drawn each quarter with a total of eight per year. To support the overlapping sample design, segments were formed to contain a minimum of 175 dwelling units[4] on average. In prior years, this average minimum segment dwelling unit size was only 90.

Once sample segments for the 1999 NHSDA were selected, specially trained field household listers visited the areas and obtained complete and accurate lists of all eligible dwelling units (DUs) within the sample segment boundaries. A *dwelling unit* in the NHSDA refers to either a housing unit or a group-quarters listing unit such as a dormitory room or a shelter bed. Using a random start point and interval-based (systematic) selection, the actual listing units were selected from the segment frame with an EPSEM (equal probability selection method) goal. After dwelling unit selections were made, an interviewer visited each selected dwelling unit to obtain a roster of all persons residing in the dwelling unit. Using the roster information obtained from an eligible member of the selected dwelling unit, 0, 1, or 2 persons were selected for the survey using an adaptation of Brewer's sampling scheme.

2. Pairwise Selection and Weights

Starting in 1998, the NHSDA samples were selected so that every pair of survey eligible persons residing within the same dwelling unit would have some known non-zero probability of selection (Penne & Witt, 1998). The primary purpose for this was to both allocate the number of selected pairs suitable for different pair-age groups, as well as increase the utility of the NHSDA data for analysts by allowing them to use survey results to measure family and other household effects. Additionally, by achieving positive probabilities for all pairs within DUs this permits unbiased estimation of the within dwelling unit component of variance.

As described in Chromy & Singh (2001), issues of extreme weights and multiplicities arise when dealing with pair data analysis.

With the notion that all possible pairs within a DU have a non-zero probability of selection, these probabilities are directly related to the household size. As indicated in **Table 1**, the number of pairs increase dramatically with the increase of the total household size. This aspect alone demonstrates the potential for large variation among pair selection probabilities and hence the pair weight.

The problem of multiplicities arises because for pair

---

[4]Dwelling unit counts were obtained from the 1990 Decennial Census data supplemented with revised population counts from Claritas.

data, initial pair design weights represent the population of all pairs, which may include multiple contributions

Table 1. Possible Pair Combinations By Persons Per Dwelling Unit

Persons	Pairs	Persons	Pairs
1	0	5	10
2	1	6	15
3	3	7	21
4	6	8	28

from the same pair domain (e.g. parent-child). The reason for this is that several pairs in a household may correspond to the same domain. However, with respect to the target parameters, typically an average contribution from all pairs from a HH, corresponding to a domain, may be desirable. This implies that the HH design weight should be divided by the pair multiplicities for the domain of interest. Because the multiplicity factors are domain-specific, determination of final calibration weights becomes somewhat nonstandard (see Chromy & Singh (2001)).

3. Generalized Exponential Modeling (GEM): An Effective Solution

GEM is a unified modeling approach which capitalizes on a generalization of the logit-type model of Deville and Särndal (1992) by not requiring bounds on weights to be uniform. It allows an analyst to address the three major concerns often associated with proper weight calibration: (i) variance inflation of small domain estimates due to extreme values (*ev*), (ii) bias due to nonresponse (*nr*) and (iii) bias due to under/over coverage, often realized through some post-stratification (*ps*) adjustment to nonrandom controls for model covariates. Specific to the NHSDA, the *ev* adjustment entails identifying extreme weights, imposing unit-specific bounds on them, and using GEM calibration to reproduce the sample distribution for various demographic domains.

Several additional aspects of the GEM weight calibration process should also be noted which assist in making it a versatile and effective device. Among these are several features:

(i) *Extreme Weight Treatment*: Not only in the *ev* adjustment step, but in every other adjustment step, unit-specific bounds can be applied on initially defined extreme weights (high or low) in order to keep them under control. Extreme weight definitions are based on some standard criteria for various design-driven domains.

(ii) *Types of Control Totals and Covariates*: The GEM process allows for internal as well as external

control totals. In the instances of *ev* and *nr,* control totals are internal to the data, such that respondent's weights are calibrated to the weights of the entire selected sample. However, in *ps* control totals come from sources separate from the data (e.g., Census Population Estimates). Additionally, the GEM can effectively model any combination of discrete and continuous covariates given that a solution exists.

(iii) *Built-in Quality Control Measures*: GEM provides descriptive statistics concerning the weights before and after each modeling step. These statistics include Unequal Weighting Effect (UWE)(Kish, p. 427), max over mean for individual weight components as well as weight products, and the percentage of extreme weights. For extreme weights, this includes an unweighted percentage, a weighted percentage and a winsorized percentage which is the amount of weight that would be removed if these were trimmed to their respective critical values of extreme weight definition. After each weight adjustment step, the GEM allows for a slippage analysis which indicates the total amount and percentage that weighted covariate totals do not match external control totals. For all covariates directly controlled for within the model these values will be zero.

(iv) *Model Covariate Selection Diagnostics*: To assist in model fitting when model convergence is an issue or for creating parsimonious models, GEM provides some selection diagnostics. Among these are t-test values and their associated p-values, sample sizes of respondents and nonrespondents for determining which covariates may potentially be difficult to control for and for each covariate, the relative difference between the unadjusted weight sum and the control total (the amount the weights must be adjusted).

(v) *Grouped Data*: Since all records with the same covariate values as those controlled for within a model will receive the same weight adjustment, the GEM allows for the collapsing/grouping of similar records. This entails summing the weight products of each similar record into a single record such that weights of all records are represented. This ultimately reduces the total number of records that are required to be processed. This may have the effect of dramatically reducing the amount of computational time.

4. 1999 NHSDA PAIR Weight Calibration

Exhibit 1 displays the different weight components associated with a final pair analysis weight. This paper will focus on the Phase II Pair specific components only. To expedite the completion of the pair weights and to reduce the burden of computer computation time, the entire sample was divided into two modeling groups based on Census Region definitions (North East + South and North Central +

West). For further explanation of the dwelling unit components and their respective calibration methods refer to the 1999 Sample Design Report (2001) and Sampling Weight Calibration for the 1999 NHSDA (2001).

The GEM methodology of weight calibration has a built-in control on extreme weights which may be sufficient for most calibration problems. However, a separate ev step after the final ps step may be performed, if necessary. This is basically a repeat post-stratification with the same ps controls but with tighter bounds on the extreme weights obtained after the ps step. Thus it redistributes the total weight such that sample distributions for various demographic domains are preserved. In our application an initial ev step before the ps (for selected persons) was necessary because of the presence of unusually extreme weights. This is weight #12. As it turned out that even with this step, there were still quite a few extreme weights. Note that use of Brewer's method for selection of pairs could give rise to such extreme weights. So we decided to use the traditional method of trimming as the first adjustment of the pair weights of the selected pair sample except that the trimming was restricted to top 1% of the weights in each extreme weight defining domain. Here, extreme weights are identified by whether they fall outside the interval, median $\pm$ 3* Interquartile range, for some pre-specified domains in a hierarchy of four domains: (a) States, (b) 3 level pair age grouping, (c)States by a 6 level pair age grouping, and (d) States by 6 level pair age grouping by household size (2 persons, 3 persons or 4+ persons). Pair age groupings are classified in the following manner: 12-17 year-old = youth, 18-25 = young adult and 26+ = Older Adult. The 6 level pair age grouping consists of all possible pairs of these 3 categories and the 3 level pair grouping combines young adults with older adults. The hierarchy of domains works by requiring a minimum of 60 observations in each cross-classification cell. If this minimum is not met at a lower level, then the critical values for extreme weights at the next higher level is utilized and so on until the minimum observation limit is met. Trimming involved identifying those weights that were both classified as extreme and within the top 1 % of the weight distribution for that particular domain of interest. Weights that were within this category were trimmed to their respective critical value.

The first component in Phase II to utilize the GEM is weight # 12, the Selected Pair Extreme Weight Adjustment as mentioned earlier.

Weight Component #13 takes full advantage of the inherent two-phase design of NHSDA. Thus the large first phase sample of screener households is used to provide post-stratification controls (for the

population of all pairs) for use at the second phase of selected pairs. Note that external controls for the pair population are not available unlike the census counts for the population of persons. It should be noted that the control totals of the full sample of possible pairs consist of counts of all pairs within a specific domain (or level of a factor in the model) using DUs with weights being the product of components 1-9.

Nonresponse adjustments (Component # 14) are calculated in the manner of adjusting the final responding pair weights to meet control totals for selected covariates.

For Respondent post-stratification, no external control totals are available for population counts. Instead, the same sample of all possible pairs of eligible rostered persons as was used in adjustment # 13 is utilized. This step starts with the same set of initial controls as in components 13 & 14. If all the controls were met in step #14, then this step would simply act as an *ev* step. Moreover, in all previous adjustment steps no actual questionnaire information is known for nonrespondents or non-selected pairs and hence all available data is based on solely dwelling unit screener information. This allows for the updated use of the data source for each adjustment. Although small in number, some discrepancies do occur between screener data and questionnaire data. This often is a result of recall bias on the part of the person providing the screener information. Hence, this final post-stratification is required since for the NHSDA, questionnaire data has priority over screener data and will be utilized in all final analysis.

The final adjustment before Hajek-type adjustment involving multiplicities (see Chromy and Singh, 2001) is another extreme weight adjustment. Sometimes, even with the built-in separate bounds for extreme weights, some weights will become extreme in comparison to other sample weights. This can be due to large differences in external controls versus observed design based weights or possibly due to large nonresponse. Therefore, after post-stratification, weight summary stats are reviewed and if the results would likely result in unreliable estimates, we have the option of performing another round of extreme weight adjustment. This would be performed for the same domains as indicated in weight component # 12 to maintain consistency in our procedures.

The final step is a multivariate calibration adjustment (Hajek-type) such that for a selected set of pair domains, estimates based on the pair data for the number of persons belonging to the domain are calibrated to the estimates obtained from the larger sample of households in which single or pair of persons are selected for the drug questionnaire. The calibration controls are obtained from the larger sample of households in which single or pair of persons are selected for the drug questionnaire.

5. Model Covariates

Typical predictors used in any pair weight adjustment components consist of State, quarter of the year, population density, group quarters indicator, household size, pair age combinations (12-17, 18-25, 26+), pair gender combinations, pair race combinations (NonHispanic Other, Hispanic, NonHispanic Black, and NonHispanic White), race of householder, SES indicator, percentage of Blacks residing in segment, percentage of Hispanics residing in segment and percentage of owner occupied DUs within a segment. This also includes any higher order interactions of interest.

6. Numerical Results

Both model groups of region pair samples of the 1999 NHSDA are used to illustrate results obtained by utilizing the GEM methodology to calibrate a final analytic weight.. **Table 2** and **3** display the summary results of before and after each pair adjustment step in the calibration process, for the Northeast and South region and North Central and West regions, respectively. Please note that "after" results of each preceding adjustment step is synonymous with "before" results of every subsequent step. This table presents sample sizes, UWE, unweighted, weighted and outwinsor percentages of extreme values and the quartile distribution of both the weight component itself and the weight product up through that step. Note that changes in summary statistics between *sel.ps* and *res.nr* are a result of calculations being based on the sample of all selected pairs versus all responding pairs. In our example, steps *res.nr* and *res.ps* use the same set of controls and so, the changes between *res.nr* and *res.ps* are caused by definitions of demographic domains changing from screener based to questionnaire based.

From the table we can see that the UWE prior to the *sel.ev* step is still quite high after weight trimming, and shows the need for the *ev* step. We can see the apparent benefit of GEM by the reduction in UWE from 58.8 to 8.3 in the Northeast and South and a decrease from 81.9 to 11.1 in the North Central and West. We can clearly see that the GEM provides very favorable results in reducing the variance among the weights. Additionally, the GEM performs quite well in the *res.ev* adjustment step by keeping the UWE in check and simultaneously reducing the percentage of extreme values to negligible levels.

Furthermore, due to time constraints, the final pair domain specific post-stratification weight adjustment involving multiplicities is still in processing and could not be presented here.

Exhibit 1: Final Analytic Pair Weight Components

Phase I Dwelling Unit Level

Design Weight Components
1 Inverse Probability of Selecting Segment
2 Quarter Segment Weight Adjustment
3 Subsegmentation Inflation Adjustment
4 Inverse Probability of Selecting Dwelling Unit
5 Inverse Probability of Added Dwelling Unit
6 Dwelling Unit Subsampling/Release Adjustment

7 Dwelling Unit Nonresponse Adjustment
8 Dwelling Unit Post-stratification Adjustment
9 Dwelling Unit Extreme Value Check/Adjustment

Phase II Pair Level

Design Weight Component
#10 Inverse Probability of Selecting a Pair Within a Dwelling Unit

11 Extreme Weight Trimming (*sel.trim*)
12 Selected Pair Extreme Value Adjustment (*sel.ev*)
13 Selected Pair Post-stratification to All Possible Pairs from Eligible Rostered Persons (*sel.ps*)
14 Respondent Pair Nonresponse Adjustment (*res.nr*)
15 Respondent Pair Post-stratifcation (*res.ps*)
16 Respondent Pair Extreme Value Adjustment (*res.ev*)
17 Pair-domain Specific Post-Stratification

References

Bowman, K., Penne, M.A., Chromy, J.R., and Odom, D. (2001). 1999 NHSDA sample design report (prepared for SAMHSA, Contract No. 283-98-9008). Research Triangle Park, NC: Research Triangle Institute.

Chen, P., Gordek, H., Heller, D.C., Penne, M.A., Singh, A.C. and Westlake, M. (2001). Sampling weight calibration for the 1999 NHSDA (prepared for SAMHSA, Contract No. 283-98-9008). Research Triangle Park, NC: Research Triangle Institute.

Chromy, J.R., and Singh, A.C. (2001) Estimation for person-pair drug related characteristics in the presence of pair multiplicities and extreme sampling weights for the NHSDA. *ASA Proc. Surv. Res. Meth. Sec*.

Chromy, J.R. (1981). Variance estimation for a sequential selection procedure. In Current Topics in Survey Sampling (eds. Krewski, D., Platek, R., and Rao, J.N.K), *Academic Press*, New York, pp 329-347.

Deville, J.-C., and Särndal, C.E. (1992). Calibration estimation in survey sampling, *JASA*, 87, 376-382.

Folsom, R.E., and Singh, A.C., (2000). A generalized exponential model for sampling weight calibration for extreme values, nonresponse and post-stratification. *ASA Proc. Surv. Res. Meth. Sec*. Pp 598-603.

Kish, L. (1965) Survey Sampling. Wiley-Interscience Publication: New York.

Penne, M.A. and Witt, M.B. (1998). 1998 National Household Survey on Drug Abuse: Determining optimal dwelling unit and person sample sizes (prepared for SAMHSA, Contract No. 283-96-0001). Research Triangle Park, NC: Research Triangle Institute.

Table 2. Summary Statistics for 1999 CAI NHSDA Pair Weight Adjustments: Census Northeast and South Regions

Summary Domains	sel.trim		sel.ev	sel.ps	res.nr		res.ps		res.ev
	Before	After	After	After	Before[5]	After	Before[6]	After	After
Sample Size	10,727	10,727	10,727	10,727	7,100	7,100	7,100	7,100	7,100
UWE	58.76	13.79	7.20	8.42	8.48	8.37		8.34	8.25
Extreme Values									
Unweighted	7.10%	6.07%	6.39%	6.26%	7.61%	5.82%	5.87%	5.75%	2.77%
Weighted	60.50%	39.80%	40.16%	39.14%	50.60%	41.20%	41.50%	34.20%	13.38%
Outwinsor	48.33%	25.17%	19.50%	13.84%	25.56%	16.05%	16.42%	6.73%	1.26%
Weight Distribution									
Adjustment Factor									
Minimum		0.00	0.09	0.09		0.24		0.03	0.33
25%		1.00	0.75	0.48		1.01		0.54	0.75
Median		1.00	0.95	0.92		1.08		0.75	0.90
75%		1.00	1.23	1.85		1.49		1.11	1.05
Max		1.00	3.97	4.50		10.00		3.99	1.99
Weight Product	(1-10)	(1-11)	(1-12)	(1-13)	(1-13)	(1-14)		(1-15)	(1-16)
Minimum	35	35	20	7	7	10		7	4
25%	788	787	693	574	540	633		425	335
Median	1,582	1,578	1,576	1,557	1,440	1,796		1,391	1,201
75%	5,015	4,930	4,885	5,336	4,726	6,757		6,352	5,302
Max	5,445,900	909,459	306,932	374,056	358,426	420,084		363,659	463,238
Max / Mean	5.0	143.2	48.3	41.8	47.2	31.1		26.9	25.6

Table 3. Summary Statistics for 1999 CAI NHSDA Pair Weight Adjustments: Census North Central and West

Summary Domains	sel.trim		sel.ev	sel.ps	res.nr		res.ps		res.ev
	Before	After	After	After	Before	After	Before	After	After
Sample Size	11,771	11,771	11,771	11,771	7,785	7,785	7,785	7,785	7,785
UWE	81.90	25.21	12.01	12.27	12.15	13.63		11.41	11.13
Extreme Values									
Unweighted	7.44%	6.25%	6.61%	7.31%	7.06%	5.45%	5.68%	4.01%	1.59%
Weighted	57.10%	37.10%	39.96%	40.28%	43.90%	36.29%	38.40%	23.80%	4.75%
Outwinsor	45.00%	23.12%	21.55%	17.30%	19.04%	12.16%	14.15%	3.70%	0.41%
Weight Distribution									
Adjustment Factor									
Minimum		0.01	0.54	0.06		0.18		0.09	0.33
25%		1.00	0.65	0.37		1.01		0.72	0.75
Median		1.00	0.88	0.79		1.20		0.87	0.90
75%		1.00	1.19	2.19		1.74		1.11	1.05
Max		1.00	4.50	6.00		9.00		5.42	1.99
Weight Product	(1-10)	(1-11)	(1-12)	(1-13)	(1-13)	(1-14)		(1-15)	(1-16)
Minimum	45	45	27	10	9	9		7	4
25%	572	571	469	365	362	495		413	335
Median	1,258	1,254	1,122	1,054	1,019	1,489		1,357	1,201
75%	3,793	3,717	3,514	3,896	3,553	5,493		5,340	5,302
Max	3,885,070	1,256,470	314,175	530,910	442,456	936,700		522,793	463,238
Max / Mean	573.1	259.7	64.9	74.6	67.0	87.0		48.6	43.0

[5] Summary stats here are representative of the responding pairs only.

[6] Extreme Value stats change here due to demographic definitions changing from screener based to questionnaire based.