TECHNICAL PAPERS ON
HEALTH AND BEHAVIOR MEASUREMENT

TECHNICAL PAPER 61

2000 NSBME: Final Weighting Strategy

Charles F. Turner, Maria A. Villarroel, Harper Gordek, and James R. Chromy

Reference Citation

Turner CF, Villarroel MA, Gordek H, Chromy JR. 2000 NSBME: Final Weighting Strategy. *Technical Papers on Health and Behavior Measurement,* No. 61, Washington DC: RTI Program in Health and Behavior Measurement, 2004.

2000 NSBME: Final Weighting Strategy

Charles F. Turner, Maria A. Villarroel, Harper Gordek, and James Chromy

1.INITIAL BASE WEIGHTS (W1).

Initial base weights were supplied by UMASS/CSR; these are the inverse of the sampling probabilities. These values are in the data file as W1INVER and are shown in the first panel of Table 1. Since only one eligible adult was selected in each sample, these initial base weights (W1INVER) were multiplied by the number of eligibles adults residing in the household (variable S1).[1] The resultant final base weight is designated W1 in the new data file.

2. IMPUTATIONS (for missing age & race)

Weight W1 can be calculated for every sample member since sample strata (gr), assigned experimental condition (oma), and eligibles 18-45 in household (s1) are known for every completed case. This was not true for the calculation of non-response adjustment weights. Calculation of the these adjustments require that we also know the gender, age, and race of each respondent. While gender was knowm for every case, age was missing (or invalid) for 8 cases (6 in National Stratum and 2 in Baltimore stratum) and race was missing for 17 cases (10 in National stratum and 7 in Baltimore stratum). We imputed values for missing ages (agegrp2a) and Black/AA race versus other racial groups (black2a) by estimating logistic regressions to predict the likelihood respondents would belong to a given age strata (18 to 30 vs. 31 to 45) or race (black vs. nonblack). The non-response adjustment are based on 32 strata defined by the assigned interview mode, race, age group, gender and sample (wtstrata).

Ages were imputed separately within each strata (Baltimore and national) using the following predictors:

> number of eligible household members, marital status, education, Hispanic origin, gender, proportion in exchange between a range of ages (0-17, 18-24, 25-34, 35-44, 45-54, 55-64, 65+), race, region of the country, medium income in exchange, racial preference of neighborhood, proportion white in exchange, proprtion black in exchange, proportion Hispanic in exchange, estimated proportion of homes that are owner occupied, and estimated proportion of homes that are non-owner occupied.

Race was predicted separately within each strata (Baltimore and national) using the following predictors:

> number of children in the home, Hispanic origin, age, marital status, education, gender, proportion black in exchange, proportion Hispanic in exchange, racial

[1]All 2,277 cases had a valid entry for S1 so there was no need to impute missing values for this variable. No household reported more than 7 eligible adults.

preference of neighborhood, region of the country, and proportion in exchange between a range of ages (0-17, 18-24, 25-34, 35-44, 45-54, 55-64, 65+).

3. Calculation of Census Estimates.

Table 2 provides the population estimates used in constructing the nonresponse adjustments. For the national sample, 1999 Census population estimates are available by single years of age for the four sex-race groups (black males, black females, non-black males, nonblack females). Thus we can use population estimates that exactly match the age range of our sampled population (18 to 45). For Baltimore City, however, the available 1999 Census estimates use 5-year age ranges that do not map precisely onto the eligible age range for this survey.

This problem required two estimations for Baltimore City. First, the numbers of 18 and 19 year olds in each race-gender group had to be estimated from 1999 Census estimates of the numbers of persons ages 15 to 19. For this estimation we have assumed that the proportions (within each race-sex group) of (1) 18-19 year olds to (2) 15-19 year olds was equivalent in 1999 and 1990. These proportions

$$\frac{N_{18\text{-}19}}{N_{15\text{-}19}}$$

for each of the race-sex groups area were calculated from the 1990 Census counts for Baltimore City (see Table 3). This proportions was then multiplied times the 1999 Census estimate of the number of Baltimore City residents who were 15 to 19 (in each race-sex group).

The above procedure provides an estimate of the Baltimore City population (in each race-sex group) ages 18 to 19, and the 1999 Census projections for ages 20 through 44 can be tabulated directly from the Census publication.

There is, however, no 1999 estimate for persons aged exactly 45 years of age. Furthermore, we were unable to find a breakdown of the Baltimore City population in the 1990 Census that provided population counts by single years for persons in their forties (within race-sex groups). (This would have allowed us to use a procedure parallel to that used for 18-19 year olds.) Given this lack of data, we have assumed that 1999 population of 45 year olds in Baltimore City was one-fifth of the estimated population ages 45 to 49. While this estimate will be slightly inaccurate, it should not have a major impact on our overall estimate for the Baltimore population aged 18 to 45 years.

4. Base weight adjusted for number of telephone lines in household (W1_adj)

The NSBME survey is telephone-based. To accurately assess the chance a person might be contacted, an adjustment of the initial base weight must be made to take account of the number of telephones lines each individual has access to in their household is necessary. Unfortunately the number of telephones lines was not

collected during the survey. To account for this missing information, a model was made from a second data set. This data set (from the NSHS 1996 survey) was from a similar population and contained many similar variables as well as the number of phone lines.

Since every observation was contacted by phone, modeling had to be done in a way to ensure that the predicted value would not be less than one. Instead of modeling to the number of phone lines, the logarithm of the number of phone lines minus one was used. Theoretically, we were modeling the 'extra' number of phone lines. The use of the logarithm made sure that the value predicted would always be greater than zero. The resulting predicted value from the coefficients in this model would be added to one to get the predicted number of phone lines in the household (predphonlines; predphon in SPSS dataset).

Modeling was done using the LOGLINK procedure in SUDAAN. Using a model consisting of age group, education level, marital status / living arrangements, hispanicity, gender, number of children, race, region, and urbanicity, coefficients were calculated. These coefficients were then applied to similar variables in our NSBME data set to predict the number of phone lines. Then, the initial sampling weight was adjusted by multiplying it by the inverse of the predicted number of phone lines (W1 x 1/predphonelines) to create an adjusted weight (W1_adj) that was used as the starting weight for the rest of the adjustments.

5. Response Rate Adjustments (W2, W3, W4, W5)

A number of the interviews, though not fully completed, had finished sections of the questionnaire. We could still use such data for section specific calculations . Therefore, separate adjustments were calculated for observations with completed drug sections (W2), sex sections (W3), std sections (W4), and entire substantive questionnaires (W5).

The adjustments were made in such a way that at each of these sections, the sum of the weights for respondents summed to the census estimates for the thirty-two strata (defined as Type of Interview (CATI vs T-ACASI), by the two sample strata (Baltimore and National), by Age (18-29 vs. 30-45), by Ethnicity (Black vs. Non-Black), by Gender (Male vs. Female)). The current weight was multiplied by a zero/one indicator of response to create a temporary weight for the section. The census estimates and the temporary weight were summed by the thirty two strata. Then the summed census estimate was divided by the summed temporary weight. This value was the section-specific adjustment weight. It was multiplied by the same zero/one indicator as before so as to avoid the confusion of assigning an adjustment factor to a weight that had been zeroed out. The product of the adjustment weight and the temporary weight would become the new weight for the section.

For example: The weight from completion of the drug section (WT12) was multiplied by the indicator if a value completed the sex section or not (SEXSAMP). This creates zero values for anyone who did not complete the sex section. This temporary weight (CKCSSEX), as well as the census estimates, were summed by the thirty two strata. These values were taken and division took place as described above, such to create 32 separate nonresponse adjustments. Further, we multiplied these 32 values by the

indicator to create a new adjustment weight (W3) that existed only for observations in the sample. The temporary weight and adjustment weight were multiplied together to form the new weight for the sex section (WT123).

The adjustments made were not simple nonresponse adjustments. Since everyone in the analysis data set had at least the first section complete, the adjustment made at this level, the first level looked at, was a post-stratification to meet control totals. Each subsequent nonresponse adjustment was done to the same levels as the first adjustment, so they also matched the control totals. Thus, these steps could be considered nonresponse/post-stratification combinations, but, for ease of reading, are referred to as nonresponse in this document.

These new weights were then normalized to the unweighted sample size of respondents of that section. The section weights (drug - WT12, sex - WT123, std - WT1234, and complete - WT12345) and their normalized versions (DRGFINWT, SEXFINWT, STDFINWT, and CMPFINWT respectively) adjust for the initial probabilities of selection, nonresponse, and deviations from Census projections of the population by age, race, and sex within the two major sampling strata (National and Baltimore samples). These weights must be used in any analysis that wishes to generate estimates for the National or Baltimore City population. (Note also that the data set is intended for analysis *within* these two major sampling strata, i.e., national sample and Baltimore City sample.)

After each of the sections had received their adjustment, new weight, and normalized weight, the data was checked. Each new weight should add up to the census total for the thirty-two levels and each normalized weight should add up to the number of observations that completed the corresponding section.

For QC purposes, the unequal weighting effect (UWE) was calculated for each of four strata, defined as type of interview by sampling strata, or more explicitly (1) National CATI Sample (2) National T-ACASI sample (3) Baltimore CATI Sample and (4) Baltimore T-ACASI sample. The values for the UWE proved to be quite favorable with no result larger than 1.5. Also, a combination response/coverage rate was calculated for the each of the thirty-two levels. This rate indicates under-coverage in black males, specifically of ages 30-45, in relation to the other levels. Some of this apparent under coverage is due, of course, to the fact that our sample was drawn from the population of *households* with telephones. Not only were non-telephone owning households excluded, but all segments of the population living outside of households (e.g., in group quarters, dormitories, military bases, prisons, etc.) were not included in the population we sampled. This should be noted when looking at results for this specific age/race group.

Summary of weight variables used in calculation

Variable	Description
S1	number of eligible respondents
Partial	Questionnaire sections completed (1=drug, 2=sex, 3=std, 9=all)
not in dataset	1999 population **census estimates** - table: T:\ABS_2000\REWEIGHTING\finalized\abs_wghtrev4.xls
not in dataset	**Sampling probabilities** - from UMASS field report - table T:\ABS_2000\REWEIGHTING\finalized\abs_wghtrev4.xls
w1inver	inverse of sampling probabilities
wtstrata	32 categories defined by sample (national & Baltimore), interview mode (T-ACASI & T-IAQ), gender, age group (18-29, 30-45), race (black vs non-black) - (missing values for age or race group were inputed)
w1	Base weights: w1inver X S1
predphone	predicted number of phone lines from model derived using NSHS96
w1_adj	w1 X (1/predicted # of phone lines)
w2	drug section ratio adjustment (frequency of wtstrata weighted by w1_adj. to obtain current estimates. Census estimates / current estimates = drug section ratio adjustment for 32 strata) see also ratio adjustment in T:\ABS_2000\REWEIGHTING\finalized\abs_wghtrev4.xls
wt12	w2 X w1_adj (population weight for drug section)
drgfinwt	normalized weight for drug section (wt12/mean of wt12 over all observations that completed the drug section)
Not in dataset	**indicator of completion of sex section** or more partial=1 (drug), sex section indicator=0 partial=2, 3, 9, sex section indicator=1
W3	sex section ratio adjustment obtained by: 1) wt12 x sex section indicator = temp sex weight 2) frequency of wtstrata weighted by temp sex weight to obtain current estimates. 3) census estimates / current estimates = sex section ratio adjustment for 32 strata
wt123	w3 x wt12 (population weight for sex section)
sexfinwt	normalized weight for sex section (wt123/mean of wt123 over all observations that completed the sex section)

not in dataset	**indicator of completion of std section** or more partial=1 or 2 (drug & sex), std section indicator=0 partial=3 or 9, std section indicator=1
w4	std section ratio adjustment obtained by: 1) wt123 x std section indicator = temp weight for std section 2) frequency of wtstrata weighted by temp std weight to obtain current estimates. 3) census estimates / current estimates = std section ratio adjustment for 32 strata
wt1234	w3 x wt123 (population weight for std section)
stdfinwt	normalized std weight (wt1234/mean of wt1234 over all observations that completed the std section)
not in dataset	**indicator of completion of ALL sections** partial=1, 2 or 3 (drug, sex, std), completion indicator=0 partial=9, completion indicator=1
w5	Ratio adjustment for respondents completing all sections obtained by: 1) wt1234 x completion indicator = temp weight for completed interviews 2) frequency of wtstrata weighted by temp weight for completed interviews to obtain current estimates. 3) census estimates / current estimates = completed interview ratio adjustment for 32 strata
wt12345	w5 x wt1234 (population weight for completed interviews)
cmpfinwt	normalized weight for completed interviews (wt12345/mean of wt12345 over all completed interviews)