North Central Texas Council of Governments
2009 National Household Travel Survey
Task 5: Reweighting the data

Final Weighting Report

Authors:
Lloyd Hicks
Amy Lin
Jing Kang

Table of Contents

Overview 1

This weighting report documents the new weights that were developed for the North Central Texas Council of Governments (NCTCOG) 2009 NHTS dataset. This report includes an overview of aspects of the new weights and includes documentation of the reweighting steps.

The goal of reweighting the 2009 NHTS NCTCOG data is to produce weights that are optimal for analysis of the North Central Texas region, consisting of Collin, Dallas, Denton, Ellis, Hood, Hunt, Johnson, Kaufman, Parker, Rockwall, Tarrant and Wise counties. Samples included in the reweighting are those originally selected in the twelve counties for the national NHTS sample or the NHTS Texas add-on sample. We replicated weighting steps done for 2008 NHTS with some modifications where necessary. For example, the original weights calculated for the year-long NHTS dataset took quarter in which the sample was selected into account. To simplify the process for the reweighting, we pooled all samples together for reweighting regardless the quarters they belonged to.

In the 2009 NHTS weighting, a household was defined as useable if at least 50% of the eligible adults followed through in completing the retrieval interview. For the reweighting of the 2009 NHTS NCTCOG sample, a "usable" household was redefined as all household members aged five and older completed the retrieval interview. As a result, the person level nonresponse adjustment is no longer needed.

The overall steps in the weighting process were as follows:

- Construction of base weights—the base weights are the reciprocals of the telephone frame sampling rates;
- Construction of jackknife replicate base weights—the replicate weights are designed to allow the user to easily produce valid jackknife variance estimators based on the sample design[1];
- Household-level nonresponse adjustments;
- Household-level raking and trimming;

[1] Each step below also processes the jackknife replicate weights

- Person-level raking and trimming;
- Computation of vehicle weights and trip weights.

The person level nonresponse adjustment was dropped in the reweighting process as there is no person nonresponse with the new definition of "usable" household. The NHTS weighting for the Texas add-on included raking to state control totals; the reweighting of these data includes raking to county household and population control totals.

The listing of chapters roughly follows the order of the weighting process. Chapter 2 describes the process for computing household-level base weights. Chapter 3 describes adjustments for nonresponse at the household level. Chapter 4 describes the raking procedure at the household level. Chapter 5 describes person-level base weights. Chapter 6 describes person-level raking adjustments. Chapter 7 describes special weights for vehicles and trips.

Household Level Weights—Base Weights 2

The primary component of the base weight is the inverse of the probability of selection of the telephone number in the frame. The sampling rate is simply the sample size divided by the total number of telephone numbers in the frame.

2.1 Base Weights at the Frame Level

The base weights are the inverse of the probability of selection for each telephone number from the frame and are calculated as the total number of working banks in a stratum multiplied by 100, divided by the number of sampled telephones in a stratum.

2.2 Replicate Base Weights at the Household Level

Replicate base weights were computed using the same methodology used for 2009 NHTS weights. The variance strata were generated based on the original frame order of the systematic sample, and were generated as 'blocks' of sampled telephone numbers based on this original frame order. The blocks had varying sizes across the sampling strata, based on the expected number of completed interviews per block. Each block was made large enough so that there would be at least one completed interview within each block.

Variance units were defined which split each variance stratum in half. This split into variance units was done randomly while keeping the selection order. A replicate was created for the variance stratum by deleting one variance unit randomly and retaining the other and doubling the weights for the retained one. For that variance stratum, all other replicate weights are equal to the base weight (only the designated replicate is 'perturbed' by setting half of the sample units' weights to zero (for that replicate) and doubling the other half of the sample units' weights for that replicate. This replicate variance methodology is called 'JK2' and is a variant of the jackknife method (see for example WesVar 4.3 User's Guide 2007, and also Rust 1986 for theoretical properties).

The last step was to combine the variance strata. The total number of replicate weights was 100. The variance strata were combined so that there were exactly 100 combined strata. In their original ordering within and across the sampling strata, variance strata 1 through 100 for example would be mapped to combined strata 1 through 100; variance strata 101 through 200 would be mapped to combined strata 1 through 100, etc. Replicate weight 1 would then correspond to combined stratum 1, corresponding to variance stratum 1, 101, 201, 301, etc. This guarantees that each replicate weight corresponds to 1/100 of the variance strata and 1/100 of the sample units, with that 1/100 spread as evenly as possible over the entire sample (all study groups and all sampling strata). This combining methodology has the purpose of reducing the number of replicate weights to an operationally feasible number. When combining is done, the original replicate strata should be combined in such a way that the covariances between the combined strata are as small in absolute value as possible, and combining across strata from differing sampling strata will accomplish this goal.

Household Level Nonresponse Adjustments 3

Nonresponse unfortunately is a major and continuously growing problem with every RDD survey. For 2009 NHTS, Westat did extensive work analyzing nonresponse and the potential for bias, and built in adjustments based on this analysis. To reflect the response patterns from 2009 NHTS, the recruitment nonresponse adjustment factors developed in 2009 NHTS were applied to the reweighting of 2009 NHTS NCTCOG sample. The following paragraph describes the methodology behind the nonresponse adjustments for the 2009 NHTS.

The nonresponse adjustments for 2009 NHTS are based on a paradigm generally used in survey research (see for example Oh and Scheuren 1983). Under this paradigm, nonresponse is treated as a subsampling process within carefully selected nonresponse-adjustment cells. The nonresponse-adjustment cells are selected to be heterogeneous in response propensity (the probability of responding) across cells, and homogeneous in response propensity within cells. The nonresponse bias analysis informed this cell selection process by finding characteristics that were related to response propensity (propensity to be successfully contacted, propensity to cooperate at the recruitment level, propensity to cooperate at the retrieval level). The final nonresponse adjustments were equal to the inverse of the weighted response rates within the selected nonresponse adjustment cells. These nonresponse adjustment cells nested within the strata utilized in sample selection. The cells were not smaller than 15 sample units, as cells with limited numbers of sample units generate unreliable (highly variable) nonresponse adjustments. In addition, cells with very low weighted response rates were collapsed with other cells to avoid extreme weighting adjustments which can increase variability. The cutoff was 1/3: cells with weighted response rates lower than one-third of the overall response rate for the study area, stratum and sample group were collapsed with other neighboring cells[2].

For the 2009 NHTS NCTCOG sample, a household was viewed as *usable* in terms of its completed retrieval interview *if 100% of the household completed the retrieval (travel) interview,* and if the household responded on a non-holiday weekday. This definition is slightly different from 2009 NHTS, where a household was viewed as usable if *50% of the adults enumerated within the household completed the retrieval (travel) interview*, regardless of the day of response. This results in a larger pool of retrieval (travel)

[2] Cells are defined as 'neighboring' based on their location in a tree generated by a binary search algorithm: see the discussion below.

interview nonresponse for the NCTCOG sample. These additional nonrespondents were compensated for in a household level raking procedure.

All of the nonresponse adjustments are fully replicated: the replicate nonresponse adjustment for a particular cell is the sum of the particular replicate weights for the numerator set divided by the sum of the replicate weights for the denominator set. The nonresponse-adjusted replicate weight is equal to the product of the replicate weight preceding the adjustment multiplied by the replicate nonresponse adjustment for the particular cell containing the sample unit. Replicating the nonresponse adjustments within the nonresponse cells allows the variance estimator to represent variance components induced from the computation of nonresponse adjustments, conditional on the selection of nonresponse cells.

3.1 Specification of Nonresponse Adjustment Cells

In 2009 NHTS, Westat's software routine WESSEARCH was utilized to define nonresponse cells within each sampling stratum for both screener nonresponse and household usability separately. WESSEARCH is based on a search algorithm produced by and used with the permission of the University of Michigan (http://www.isr.umich.edu/src/smp/search/).

For recruitment nonresponse adjustment cells, the WESSEARCH algorithm searched within each study, sample group, and stratum separately. In some cases, the final cell was the study—sample group—stratum combination alone (no breakdown of this basic sampling cell). The algorithm avoided cells with a sample size smaller than 15 or a recruitment respondent sample size smaller than 3, and did not allow for adjustments larger than three times the mean adjustment for the sample group—stratum combination. In case of violation of these norms, the cells were collapsed, by collapsing 'up the tree'. Every terminal cell found to be deficient was linked to a parent cell. All child cells of this parent cell were collapsed making the parent cell the new terminal cell.

Potential cells were generated based on Genesys exchange level information. The nonresponse cells were dichotomous cells (above-median and below-median) using weighted medians of Genesys exchange characteristics within the study, sample group, and stratum. For example, one set of cells was above-median percentage of Hispanics and below-median percentage of Hispanics for a particular stratum and sample group. The above-median percentage cell included telephone numbers in exchanges that are above the cutoff (the weighted median) in percentage Hispanics, as estimated

by Genesys. Not every set of cells was chosen: only those that registered as significantly correlated to response propensity within the stratum and sample group were chosen. And, not every Genesys characteristic was tested: only those that registered as important at the study level in the ¾ nonresponse bias analysis were tested.

Raking Procedures—Household Level 4

It is well-known that RDD surveys do not cover the full population of households of interest, as not every household has a telephone. An estimated 2-4 percent of households in the United States do not have a telephone. In addition to non-telephone households, households with telephone numbers in zero banks (sets of 100 telephone numbers with the same prefix that have no listed residential numbers: these are not included on the sampling frame for efficiency reasons) and cellphone-only households are not covered in the main NHTS sample. There may be significant numbers of households in zero banks, and we know there is a growing number of cellphone-only households (estimated at approximately 20% at the time of the NHTS data collection). These various sources of undercoverage can result in people of a certain demographic characteristic being underrepresented in the final sample (e.g., Hispanics or males). It is possible to adjust for undercoverage through a poststratification weighting process called "raking," where the weights are iteratively adjusted to independent controls totals for various demographic categories. The process has the effect of differentially adjusting the weights of the sampled households within groups of demographically similar households, so that the total sum of weights for the sampled households equals the corresponding independent control totals for all households (including those not covered by the RDD sample).

Raking and trimming steps were performed iteratively at the household level. The trimming steps included a 'pre-trim' step preceding the first household raking step, and a 'post-trim' step following each household raking step.

The pre-trim step consisted of checking for weights that were more than 3.0 times the median weight. If less than 1% of the weights fell into this category, then all such weights were trimmed back to equal the cutoff. If more than 1% of the weights fell into this category[3], then the largest 1% set of the weights were trimmed back to equal the 99th percentile of the weights. If there were less than 100 observations with at least one beyond the cutoff, then exactly one observation was trimmed back.

[3] The number of weights affected by this rule was 1% of the number of sample units, rounded up to the smallest larger integer. For example, if the sample size was 120, then the number of trimmed weights was 2 (1.2 rounded up). In particular, there was always at least one weight trimmed if any weights exceeded 3.0 times the median weight. The actual percentage of trimmed weights then could be slightly larger than 1%.

The trimming steps following the raking steps (the 'post-trim steps') targeted for trimming any weights that were 4.5 times smaller or 4.5 times larger than the median weight. A maximum of 2.5% of the weights could be trimmed on the high side and a maximum of 2.5% of the weights could be trimmed on the low side for each post-trim step[4]. If more than 2.5% of the weights were greater than 4.5 times the median weight (less than 4.5 times the median weight), then the largest (smallest) 2.5% of the weights were trimmed back to the 97.5th percentile (the 2.5th percentile).

The iteration of raking and trimming steps were complete when all of the trimming factors for that final putative trimming step were between 0.99 and 1.01. We found in practice that trying to trim to a stricter tolerance led to a large number of spurious iterations of the algorithm, with little value added (the changes in weights in each iteration were less than 1%). Out of the 3,053 raked households, 90 households had their weights trimmed.

Each household raking step in the cycle was done to a tolerance of ±5,000 (i.e., the weighted household totals will be raked until they are within 5,000 of the household control totals).

All of the raking and trimming adjustments are fully replicated. The raked and trimmed replicate weight is equal to the product of the replicate weight preceding raking and trimming multiplied by the replicate raking and trimming factor for the particular cell containing the sample unit. Replicating the raking and trimming adjustments allows the variance estimator to represent variance components induced from the computation for raking and trimming.

Section 4.1 below describes the initial raking dimensions for household raking.

4.1 Raking Dimensions for Households

The 2011 American Community Survey (ACS) data was used to develop the control totals for the reweighting of the NCTCOG dataset. The control totals for all dimensions were derived from five-year 2011 ACS estimates.

The dimensions were as follows[5].

[4] This count of trimmed weights was 2.5% of the number of sample units, rounded up to the smallest larger integer. For example, if the sample size was 100, then the maximum high side or low side number of trimmed weights was 3 (2.5 rounded up). The actual percentage of trimmed weights on the high and low side could be slightly larger than 2.5%.

- (1) Number of household workers * Number of household vehicles * County of residence
- (2) Household Size * Number of household workers * County of residence
- (3) Household Income * County of residence

In the event that convergence failed, the collapsing of cells was conducted in a stepwise, automated manner using the rules provided by NCTCOG.

5 Person-Level Base Weights

Retrieval interviews were completed for each eligible person within each of the recruited households. In terms of sampling, this means that every person in the household had a probability of selection equal to that of the household. In principle, each person's base weight was equal to the final household base weight. It is appropriate that the base person-level weights include the household nonresponse adjustments as the person sample at this point is only coming from the households that are judged to be useable, and include the household raking factors as the usable households represent all eligible households within the study area.

A household was defined as useable if 100% of the household members aged five and older completed the retrieval interview, and responded on a non-holiday weekday.

Person-Level Raking Adjustments 6

A person-level raking adjustment was conducted using the final questionnaire items from completed retrieval interviews. In the person-level raking adjustments, we utilized the 2011 ACS data for total persons by sex and age, by county of residence.

Person level trimming and raking followed an iterative process. An initial trimming was done, followed by an initial raking, followed by cycles of trimming and raking and trimming to convergence, similar to household level raking.

The pre-trim step consisted of checking for weights that were more than 3.0 times the median weight. If less than 1% of the weights fell into this category, then all such weights were trimmed back to equal the cutoff (3.0 times the median weight for the composite domain). If more than 1% of the weights fell into this category[6], then the largest 1% set of the weights were trimmed back to equal the 99th percentile of the weights. If there were less than 100 observations with at least one beyond the cutoff, then exactly one observation was trimmed back.

The trimming steps following the raking steps (the 'post-trim steps') targeted for trimming any weights that were 4.5 times smaller or 4.5 times larger than the median weight for the composite domain. A maximum of 2.5% of the weights could be trimmed on the high side and a maximum of 2.5% of the weights could be trimmed on the low side for each post-trim step[7]. If more than 2.5% of the weights was greater than 4.5 times the median weight (less than 4.5 times the median weight), then the largest (smallest) 2.5% of the weights was trimmed back to the 97.5th percentile (the 2.5th percentile).

The cycle of raking and trimming steps was complete when all of the trimming factors (the adjustments to the weight during the trimming step) for that potentially final trimming step were between 0.99 and 1.01. Out of the 7,202 raked persons, 324 had their weights trimmed.

---

[6] The number of weights affected by this rule was 1% of the number of sample units, rounded up to the smallest larger integer. For example, if the sample size was 120, then the number of trimmed weights was 2 (1.2 rounded up). In particular, there was always at least one weight trimmed if any weights exceeded 3.0 times the median weight. The actual percentage of trimmed weights then could be slightly larger than 1%.

[7] This count of trimmed weights was 2.5% of the number of sample units, rounded up to the smallest larger integer. For example, if the sample size was 100, then the number of trimmed weights was 3 (2.5 rounded up). The actual percentage of trimmed weights on the high and low side could be slightly larger than 2.5%.

The person-level raking process was done to a tolerance of ±10,000 (i.e., the weighted person totals were raked until the sum of weight was within 10,000 of the person control totals). In the event that convergence failed, the collapsing of cells was conducted in a stepwise, automated manner using the rules provided by NCTCOG.

Similar to the preceding steps, all of the raking and trimming adjustments are fully replicated.

Other Weights 7

7.1 Vehicle Weights

The raked vehicle weights can be used to analyze characteristics of vehicles reported by households and is equal to the raked household weights.

7.2 Trip Weights

The trip weights are equal to 239[8] times the person-level weights. These are the appropriate weights for counting trips for the year (e.g., for annual travel estimates). If daily trip estimates are required, the user should weight the data and divide the output by 239.

[8] The number of non-holiday weekdays in a given year.

References

Burke, J., Mohadjer, L., Green, J., and Waksberg, J. (1994). Composite estimation in national and state surveys, *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 873-878.

Kish, L. (1992). Weighting for unequal P_i. *Journal of Official Statistics 8*, 183-200.

Lohr, S. L., and Rao, J. N. K. (2007). Estimation in multiple-frame surveys, *Journal of the American Statistical Association*, 101, 1019-1030.

Oh, H. L. and Scheuren, F. J. (1983). Weighting adjustment for unit nonresponse, in W.G. Madow, I. Olkin, D. B. Rubin (eds.), *Incomplete Data in Sample Surveys, Vol. 2*, New York: Academic Press, pp. 143-184.

Rust (1986). Efficient replicated variance estimation. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 81-87.

WesVar 4.3 User's Guide (2007). Rockville MD: Westat, Inc.