Weighting Approach for a Statewide Dual-Frame RDD Survey

Liza Nirelli McNamee[1], Monica Reyna[1], Snehalatha Gantla[1]
Priscilla Salant[2], J.D. Wulfhorst[1]

[1]University of Idaho, Social Science Research Unit
[2]University of Idaho, McClure Center for Public Policy Research

Abstract
This paper describes a statewide dual-frame RDD survey weighting procedure developed in SAS using the *Raking SAS Macro with weight trimming* (Izrael, 2009). The AAPOR Cell Phone Survey Task Force Report states, *"There is no consensus regarding how RDD cell phone samples should be weighted, especially when combining them with RDD landline samples."* During the winter of 2014, a dual-frame, stratified random sample, including cell and landline telephone users, was utilized to ensure coverage of a statewide population. A total of 12 sub-sampling frames were created by using six non-overlapping strata representing six geographic regions and two frames, cell and landline, for each region. The weighting process consisted of developing design weights for selection probability, eligibility, non-response, multiplicity due to multiple phones, respondent selection procedure, and frame overlap. Raking and trimming techniques were used to adjust the sample to reflect the statewide telephone service benchmarks and demographics of interest. This paper presents the weighting methodology along with estimates at each stage, the limitations of the procedure, and the methodological lessons learned.

Key Words: design weights, raking, trimming, Raking SAS Macro with weight trimming, sample balancing, iterative proportional fitting

1. Statement of Problem

In survey research, a study sample often does not mimic the population. Weighting helps yield accurate population estimates by assigning individual adjustments to survey respondents. Well-respected survey centers describe their weighting methodology; however, most methods of implementation are not available outside of their organizations. Implementation can be overwhelming for other researchers or smaller centers. In addition, *"There is no consensus regarding how RDD cell phone samples should be weighted, especially when combining them with RDD landline samples"* (AAPOR, 2010). The objective of this paper is to document a comprehensive, research-based approach for weighting a dual-frame phone survey using SAS software. This paper is intended to aid researchers who do not have an institutional method to develop survey weights. It is not meant to be inventive; rather it combines and documents what has already been published.

2. Overview of Project Background

A short project background is provided to allow the reader a better understanding of the weighting process.

2.1 Sampling Approach and Screening

The Social Science Research Unit (SSRU) at the University of Idaho conducted telephone surveys using a dual-frame, stratified random sample, including cell and landline users, to ensure overall coverage of the population. This combined approach reduces sampling error, while increasing the efficiency of the estimators of overall population parameters (Lavrakas, 2008).

A total of 12 sub-sampling frames were created by using six non-overlapping strata representing the various regions and two frames, cell and landline, for each region. These sub-sampling frames were purchased from Survey Sampling International for each of the six regions to create a statewide sample stratified by the way the state is organized and managed. As estimated by the National Center for Health Statistics, the statewide prevalence of landline only, mobile only and dual phone type users were 4.9%, 52.3% and 40.2% (Blumberg 2013). Approximately 44% of the numbers purchased were for landline. This partition was chosen based on a variety of factors including cost, past RDD sample performance, and desired sampling margin of error. The goal was to achieve an overall state-level sampling margin of error of +/-3% and at least +/-8% at the region levels, with 95% confidence.

The project used a modified Dillman method (Dillman, 2009) implementing multiple call-backs for each member of the sample. All telephone interviewers received training in proper telephone interviewing, phone etiquette, and the use of WinCATI, a computer assisted telephone interviewing system. In addition, interviewers received training specific to the survey. Each interviewer was required to complete an online National Institutes of Health training course in human subjects research, including confidentiality rules and regulations. Trained supervisors monitored and oversaw interviewers during each calling session.

2.2 Sampling results

Calls were made from February 3rd through April 17th, 2014. SSRU called each phone number in the sample at least eight and up to eleven times in an effort to complete a survey. Interviewers made calls Monday through Friday in the mornings, afternoons, and evenings, as well as on Saturdays between 10:00 a.m. and 2:00 p.m. PST.

A total of 1,062 interviews were completed. The response rate, cooperation rate, refusal rate, and contact rate are defined and reported below (AAPOR, 2015):

- The response rate is the proportion of completed interviews out of all estimated eligible cases in the sample (RR3). The response rate for the entire study was 54.2%.
- The cooperation rate is the proportion of eligible cases who were interviewed out of those that were contacted (COOP1). The cooperation rate for the study was 83.5%.
- The refusal rate is the proportion of all eligible cases who refused to be interviewed or broke off the interview (REF2). The refusal rate for this study was 10.2%.
- The contact rate is the proportion of all incidences when a phone number in the sample reached a person (CON2). The contact rate for this study was 64.9%.

2. Why Weight?

When conducting surveys, researchers wish to make generalizations beyond those who were sampled. This is only possible when the surveyed individuals are representative of the population of interest. However, a sample will seldom mimic the population it came from and be representative with respect to all variables measured in the survey. To account for survey error, the original sample is reconfigured by applying weighting adjustments to each respondent. Persons in under-represented groups get a weight greater than one, while those in over-represented groups get a weight smaller than one. Subsequently, weight adjusted variables are used in the computation of means, totals, and percentages.

3. Weighting Process

Weighting a sample reflects sampling design decisions. Additionally, weighting incorporates the use of auxiliary data to improve the efficiency of estimators and ensure that the sample more accurately reflects the characteristics of the population of interest. The weighting process has three parts: identifying the design weights, raking, and trimming.

3.1 Design Weights
Design weights reflect the sample design and respondent selection procedure. The components of the design weights are identified in the planning stage of the study. Some common components are selection probability (*PS*), undetermined eligibility (*Elig*), nonresponse (*NR*), multiplicity due to multiple phones (*Num*), number of adults in the household (*NumAdult*), and frame overlap (*cell/landline*). Equations 1 and 2 below represent the design weights for the landline and cellphone respondents used in this study *(Kennedy, AAPOR Webinar 2012)*.

$$W_{LL\,i} = W_{PS\,LLi} * W_{Elig\,LLi} * W_{NR\,LLi} * 1/Num_j * NumAdult_j * .5^{IDual} \qquad (1)$$
$$W_{CL\,i} = W_{PS\,CLi} * W_{Elig\,CLi} * W_{NR\,CLi} * 1/Num_j * NumAdult_j * .5^{IDual} \qquad (2)$$

- The subscript *i* on $W_{LL\,i}$ and $W_{CL\,i}$ indicates that because there are 6 strata, there will be a base weight for each mode and stratum.
- The subscript *LL* refers to a landline sample while *CL* refers to a cell phone sample.
- The subscript *j* corresponds to the individual respondent.
- *Num* stands for number of telephone lines in the household.
- *NumAdult* stands for number of adults in the household.
  - *NumAdult* and *Num* were set to 1 for cell phone users because a cell phone is typically used by only one person and most people only have one cell phone. *NumAdult* and *Num* were capped at 3 (Kennedy, 2007) for landline users in order to reduce the amount of variance in the weights.
- For overlapping frames, a ½ compositing estimator was used, where ***IDual*** is an indicator variable which takes on the value of 1 if the respondent is a dual user, and 0 otherwise. As a result, dual users were down-weighted by 0.5 This adjustment accounted for the fact that cell- or landline-only users have a lower probability of being contacted than users who have both lines.

The adjustment for the overlapping frame is a key part of equations 1 and 2. There are alternative choices that can be used to represent the dual frame, but the ½ compositing estimator was chosen because it is the simplest to implement and in this project the design weight variance was reasonably small. To determine what other compositing estimator would best serve the sample design, the design weights for each of the dual frame estimators would need to be computed and the resulting variances compared. In general, a smaller variance is preferred (*Kennedy, AAPOR Webinar 2012*).

3.2 Raking and Trimming

Raking (i.e., sample balancing or iterative proportional fitting) is one of the most common methods used to adjust for auxiliary data. It adjusts the design weights so that the weighted sample aligns with the external population distribution for multiple categorical variables simultaneously (Izrael, 2000). Lastly, trimming is used to reduce extreme weights to cutoffs, thereby improving variance properties (Potter, 1990).

The basic idea behind raking is to ensure that the sample margin means will equal the population control totals (Izrael, 2000). Raking techniques can be carried out if reliable benchmarks are available. Auxiliary variables (i.e., gender, age, marital status, region of the country) must have been measured in the survey and their population distribution must be available.

The weighting process is more complex and more costly if a target population without reliable pre-existing benchmarks is being studied. In these situations, it is recommended that the study be conducted on a larger known general population with available control totals for the auxiliary variables. A statewide study, for example, should be conducted and the survey should include a screening question that identifies the target population members. It is best practice to collect demographic data on all respondents, including those who are not of interest in the study. After computing the design weights for all respondents, a calibration to the demographic distribution of the larger known general population control totals is necessary. This is a key step because the control totals are known, but the demographic distribution of the target population of interest is unknown (Battaglia, 2009). Once the final weights are calculated, the weight of respondents not of interest is set to missing and the analysis is conducted on the target sample of interest. This process is more expensive because data are collected on a larger than needed population, but it is an option used to achieve reasonable estimates when benchmarks are unavailable for a target population.

After raking, trimming is often used. Trimming is a sampling procedure that reduces extreme weights to cutoffs, thereby improving variance properties while potentially introducing bias (Rizzo, 2014). Guidelines suggest trimming weights greater than 3.5, 4.0, or 4.5 and weights smaller than 1/3.5, 1/4.0, or 1/4.5 (Rizzo, 2014). Another approach is to trim at the 5th and 95th percentiles, which was the method chosen in this project due to the nature of the survey topic.

4. Implementation of the Weighting Process

4.1 Calculating Design Weights.

Given the dual frame used in this study, each of the design weights had to be computed for both the landline and cell phone samples using equations 1 and 2. For illustration purposes, table 1 shows the equations of the design weight components of landline users

only. These equations, explained in the Kennedy AAPOR Webinar (2012), were used to compute the design weight components for selection probability (**PS**), undetermined eligibility (**Elig**) and nonresponse (**NR**) for landline users. Although not illustrated here, the design weight components were also computed for cell phone users.

Table 1. Design Weight Components of Landline Users[1]	
Component	*Equation*
Adjustment for sample selection probability LL[2]	$W_{PS\ LLi} = \frac{N_{LLi}}{n_{LLi}} = \frac{\textit{total count of landline numbers in the frame}}{\textit{total count of landline numbers used in the survey}}$
Adjustment for Undetermined Eligibility LL	$W_{Elig\ LLi} = \frac{\textit{count of known eligible cases} + \textit{count of estimated eligible cases}}{\textit{count of known eligible cases}}$
Adjustment for Nonresponse LL[3, 4]	$W_{NR\ LLi} = \frac{(\textit{count of known eligible nonrespondents} + \textit{count of eligible respondents})}{\textit{count of eligible respondents}}$
1. In this table we show the equations calculated for the landline frame only. However, in practice, we calculated these equations and weights for both land and cell phone frames. 2. NLL = Total size of landline frame. This can be obtained from the sample vendor, though it is commonly mistaken for the total phone numbers purchased. 3. Count of known eligible cases = number of eligible interviews + number of eligible non-interviews + number of eligible refusals. 4. Count of estimated eligible cases = {[(number of eligible respondents)/(number of respondents with a known case)]× (number of respondents with unknown cases)} where: eligible = eligible interview, eligible non-interview, eligible refusal; known case= eligible, ineligible, not a working number; unknown cases = no contact, unknown eligibility (AAPOR, 2015)	

After calculating design weight components for *PS*, *Elig* and *NR* for every stratum within each mode, the components of the design weights were merged with the survey data based on an identifier that explicitly indicated the combination of strata and mode. In this project there were 12 total combinations as a result of 6 strata and 2 modes. When all 12 combinations of *PS*, *Elig*, and *NR* were merged with the survey responses, *Num*, *NumAdult* and *IDual* could then be accessed to calculate the base weights using equation 3.

$$W_i = W_{PS} * W_{Elig} * W_{NR} * 1/Num_j * NumAdult_j * .5^{IDual} \qquad (3)$$

To accommodate the expectations of users from the social sciences, in figure 1 a portion of the SAS code is presented showing the rescaling of the design weights to reflect the sample size.

```
*population size base weight*;
BaseWeight_pop=wgt_Elig*wgt_NR*wgt_PS*(1/Num)*NumAdults*.5**(dual_ind);
*Sample base weight*;
BaseWeight_sam=(BaseWeight_pop)*(1062)/Z;
*Where sample size=1062 and Z is sum of BaseWeight_pop The BaseWeight_sam
variable are design weights reflecting the sample size;
```

Figure 1. Equation for rescaling weights to sample size

4.2 Raking and Trimming

To accomplish raking and trimming, the *Raking SAS Macro with weight trimming* was used. First developed in 2000, this macro has been revised to help improve the user's ability to find a balance between the bias of the key outcome variable and sampling variability. Specifically, the SAS Macro starts with design weights and adjusts them to match the external population distribution across multiple variables while controlling the highest and lowest extreme weights. The 2009 version of the SAS Macro has two trimming procedures (IGCV and MCV) that achieve coverage while controlling the extreme weights (Izrael, 2009).

The *Raking SAS Macro with Weight Trimming* (**rake_and_trim**) is available on the ABT Associates website and the macro is described in the 2009 paper by Izrael, Battaglia, & Frankel.

Although the **rake_and_trim** macro allows multiple auxiliary variables, for simplicity only two auxiliary variables (age and gender) are presented in this paper and for confidentiality reasons the source used for the auxiliary data is left undisclosed. The population benchmarks are displayed in table 2.

Table 2. Auxiliary Variable Percentages for the Population	
Age	
18 to 24	5.74%
25 to 44	34.02%
45 to 64	36.82%
65 to 74	12.67%
75+	10.90%
Gender	
Male	47.57%
Female	52.58%
Percentages may not add to 100 due to rounding.	

The **rake_and_trim** macro requires separate datasets for each auxiliary variable and the name of the dataset should correspond to the name of the respective variable in the dataset (Izrael, 2000). As shown below in figure 2, these datasets were named **agecat_w_rt** and **dem_sex**, corresponding to the auxiliary variables displayed in the macro call (figure 3). To refer to those who refused to fill out 'age' on the survey, **agecat_w_rt** was rescaled to compute a 6th category. When implementing the **rake_and_trim** macro, the marginal totals of the auxiliary variables were calculated to add up to the sample size (n=1062).

```
   DATA dem_sex;
     INPUT dem_sex percent mrgtotal;
     DATALINES;
1 47.57  504
2 52.58  558
;

DATA agecat_w_rt;
     INPUT agecat_w_rt  percent mrgtotal;
     DATALINES;
1 5.65  60
2 33.24 353
3 35.97 382
4 12.34 131
5 10.64 113
6 2.17   23
;
**agecat_w_rt is rescaled to compute a 6th category for those who refused
to fill out "age" on the survey;
```

Figure 2: SAS code for datasets containing marginal totals

A closer examination of the **rake_and_trim** macro call presented in figure 3 reveals some important details. The starting weights (**inwt=BaseWeight_sam**) are the design weights calculated earlier using equation 3. The auxiliary variables used to rake are defined in the **varlist** statement and **cntotal** is equal to the sample size. The IGCV (Individual and Global Cap Value) method is based on the specification of global low and high weight cap factors and individual low and high weight cap values (Izrael, 2009). In this project the individual high cap (**A**), was set to be greater than the global high weight cap (**C**), and the individual low cap (**B**) was chosen to be smaller than the global low weight cap (**D**). Because of this, in the weighting process, the global cap values were used exclusively and the individual cap values were ignored. This method was chosen because the MCV (Margin Cap Value) trimming methods were not a natural choice for trimming at the 5th and 95th percentiles. Secondly, **A** and **B** were chosen to be extremes of **C** and **D** because convergence was an issue. The values for **C** and **D** reflect the 95th and 5th percentiles of the untrimmed weights. These were obtained from first running the macro with **MethTrim** set to blank, thus performing the raking routine with no trimming.

```
%rake_and_trim
(
inds=BaseWeight,    /* name of input data set*/
outds=RakeWeight_3,  /* name of output data set*/
inwt=BaseWeight_sam,  /* input raking weight being adjusted; if there
                         is no weight, 1 is assigned*/
freqlist=  dem_sex agecat_w_rt, /* list of data sets with marginal
control totals or percents*/
outwt=RAKED_WGT_3,     /* name of raked and trimmed weight*/
varlist=dem_sex agecat_w_rt,  /* list of raking variables*/
numvar=2,   /* number of raking variables*/
cntotal=1062,  /* general control total*/
trmprec=100,   /* termination criterion based on marginal totals*/
trmpct= 0.025, /* macro will terminate based on this criterion */
numiter=75,    /* number of iterations; default is 75*/
prdiag=N,   /* print detailed diagnostics; default is N - condensed
diagnostics, Y - full printout */
MethTrim = IGCV,  /**** trimming method  - IGCV or MCV***/
            A       = 10000,   /* factor by which the respondent's
                                  weight is multiplied to get IHCV */
            B       = 0.00001, /* factor by which the respondent's
                                  weight is multiplied to get ILCV */
            C       = 2.3745913,   /* factor by which mean input
                                      weight is multiplied to get
                                      GHCV*/
            D       = 0.1967033,   /* factor by which mean input
                                      weight is multiplied to get GLCV
                                      */
            INOC    = 15     /* iteration from which to start checking
                                backward on signs of non-convergence */
  );
```

Figure 3. Macro call for **rake_and_trim**

4.3 Examining the Macro Output

A condensed output of the macro using the IGCV trimming method is displayed in figure 4 and tables 3-6. Figure 4 summarises the input weight and the trimming specifications of the **rake_and_trim** macro call.

```
Sample size of completed interviews: 1062
Raking input weight adjusted to population total: BASEWEIGHT_SAM_ATPT
Mean value of raking input weight adjusted to population total: 1.00
Minimum value of raking input weight: 0.05
Maximum value of raking input weight: 2.42
Coefficient of variation of raking input weight: 0.69
Global low weight cap value (GLCV): 0.20
Global low weight cap value factor: Mean input weight times 0.1967033
Global high weight cap value (GHCV): 2.37
Global high weight cap value factor: Mean input weight times 2.3745913
Individual low weight cap value (ILCV) factor: Respondent's weight times 0.00001
Individual high weight cap value (IHCV) factor: Respondent's weight times 10000
Number of respondents who have an individual high weight cap value less than the global low weight cap value
(GLCV used in weight trimming): 0
Number of respondents who have an individual low weight cap value greater than the global high weight cap
value  (GHCV used in weight trimming): 0
```

Figure 4. Summary of input weights and trimming parameters

Table 3 illustrates the weighted distribution before raking.

Table 3. Weighted Distribution of Sex and Age Before Raking

	Input Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Input Weights	Target % of Weights	Difference in %
			DEM_SEX			
1	620.88	504	116.88	58.463	47.458	11.005
2	441.12	558	-116.88	41.537	52.542	-11.005
			agecat_w_rt			
1	80.74	60	20.74	7.603	5.650	1.953
2	345.89	353	-7.11	32.569	33.239	-0.670
3	390.72	382	8.72	36.791	35.970	0.821
4	172.17	131	41.17	16.212	12.335	3.877
5	53.14	113	-59.86	5.004	10.640	-5.637
6	19.34	23	-3.66	1.821	2.166	-0.344

The termination message from the **rake_and_trim** macro is displayed below:

```
**** Program terminated at iteration 5 because all current percents differ from
target percents by less than 0.025 ****
```

 In 5 iterations the convergence criterion of 0.025% was met as displayed in table 4. As evidenced by the low percentage difference, the weighted sample aligned simultaneously for sex and age.

Table 4. Weighted Distribution of Sex and Age After Raking

	Input Weight Sum of Weights	Target Total	Sum of Weights Difference	% of Input Weights	Target % of Weights	Difference in %
			DEM_SEX			
1	503.98	504	-0.02	47.456	47.458	-0.002
2	558.02	558	0.02	52.544	52.542	0.002
			agecat_w_rt			
1	59.99	60	-0.01	5.649	5.650	-0.001
2	352.94	353	-0.06	33.233	33.239	-0.006
3	381.99	382	-0.01	35.969	35.970	-0.001
4	131.04	131	0.04	12.339	12.335	0.004
5	113.05	113	0.05	10.645	10.640	0.004
6	23.00	23	0.00	2.166	2.166	0.000

Selected statistics on the final raked and trimmed weights are shown in table 5. The number of respondents who had their weights decreased by the trimming was 83, while the number of respondents who had their weights increased by trimming was 86. A verification was made that trimming did not predominantly trim a particular stratum. If this were to occur, raking and trimming by each stratum separately would be recommended.

Table 5. Selected Statistics on Raked and Trimmed Weights

Iteration Number	Maximum Absolute Value of Difference in Sum of Weights	Maximum Absolute Value of Difference in %	Coefficient of Variation of Weights at the Completion of the Iteration
1	17.4192	1.6402	0.63113
2	4.3206	0.4068	0.62800
3	1.0181	0.0959	0.62861
4	0.2833	0.0267	0.62846
5	0.0615	0.0058	0.62855

Number of Respondents Who Had Their Weights Decreased by the Trimming: 83
Number of Respondents Who Had Their Weights Increased by the Trimming: 86
Raking output weight: RAKED_WGT_3

Table 6 shows comparisons of the base weights to the raked and trimmed weights. As expected, the mean input weight was equal to 1 due to rescaling the weights from the population to the sample size. The coefficient of variation for the raked and trimmed weights was 0.629.

Table 6. Comparison of Base Weights to Raked and Trimmed Weights

Weight	Mean	Min	Max	CV
BASEWEIGHT_SAM_ATPT	1.00	0.05	2.42	0.686
RAKED_WGT_3	1.00	0.20	2.38	0.629

5. Outcomes at each stage of weighting

To summarize the results of the method, a comparison of the design (without any raking and trimming), raked but not trimmed, and raked and trimmed weights are presented in table 7. The noteworthy measure in this table is the design effect due to weighting ($1 + cv^2$) which measures the expected increase in sampling variability due to unequal weights compared to a simple random sample of 1062 respondents. Generally, smaller design effects are preferred (Izrael, 2009).

Table 7. Outcomes at Each Stage of Weighting

Weight	Variable name	Min weight	Max weight	Coefficient of variation (cv) of the weights	Design effect due to weighting ($1 + cv^2$)	Change in sample variability relative to design-weighted sample
Design	**BASEWEIGHT_SAM**	0.05	2.42	0.686	1.47	-
Raked but not trimmed	Macro ran with **MethTrim= ;**	0.05	8.62	0.724	1.52	+3.4%
Raked and trimmed	**RAKED_WGT_3**	0.20	2.38	0.629	1.40	-4.8%

The design effect of the input weights was 1.47. The design effect due to raked weights was 1.52. The raked but not trimmed weight brought the sample into agreement with the population distribution for the age and gender control variables; however, the weight exhibited an increased variability of 3.4% relative to the design-weighted sample. In general, raking reduces bias but increases variance (Battaglia, 2009).

The design effect of the raked and trimmed weights was 1.40. The raked and trimmed weights not only brought the sample into agreement with the population distribution for the age and gender control variables; they also decreased the sampling variability by 4.8% relative to the design-weighted sample. In general, trimming decreases variability but increases bias (Battaglia, 2009).

6. Conclusion

In conclusion, weighted survey estimates should be compared to population estimates. Weighting complex surveys is a multiple-stage process that can be challenging, but the ultimate goal is to obtain the best possible estimates. The sampling decisions made at the planning stage are incorporated in the design weights. Raking aligns the sample with the population benchmarks but tends to increase sample variance, while trimming decreases sample variance but increases bias. In the end, the raking used in combination with trimming requires a delicate balance.

Acknowledgements

References

The American Association for Public Opinion Research. 2010. New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers. *AAPOR*.

The American Association for Public Opinion Research. 2015. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 8th edition. *AAPOR*.

Battaglia, M. P., D. Izrael, D. C. Hoaglin, and M. R. Frankel. 2009. Practical Considerations in Raking Survey Data. *Survey Practice*, 2.5.

Battaglia, M. P., D. Izrael, D. C. Hoaglin, and M. R. Frankel. 2009. Tips and Tricks for Raking Survey Data (a.k.a. Sample Balancing). *American Association for Public Opinion Research*.

Blumberg, S. J., N. Ganesh, J. V Luke, and G. Gonzales. 2013. Wireless substitution: State-level estimates from the National Health Interview Survey 2012. *National Health Statistics Reports* (70).

Dillman, D., J.D. Smyth, and, L.M. Christian. 2009. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. New York, NY: John Wiley & Sons.

Izrael, D., D. Hoaglin, and M. P. Battaglia. 2000. A SAS Macro for Balancing a Weighted Sample. *SUGI 25*, 258-25.

Izrael, D., M. P. Battaglia, and M. R. Frankel. 2009. Extreme Survey Weight Adjustment as a Component of Sample Balancing (a.k.a, Raking). *SAS Global Forum 2009,* 247-2009.

Kennedy, C. 2007. Evaluating the Effects of Screening for Telephone Service in Dual Frame RDD Surveys. *Public Opinion Quarterly* 71 (5): 750-771.
doi: 10.1093/poq/nfm050

Kennedy, C., and S. Kolenikov. 2012. Weighting Approaches for Dual Frame RDD Surveys. [Webinar]. *AAPOR*.

Lavrakas, P.J. 2008. *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: SAGE Publications, Inc.

Potter, F. 1990. A study of procedures to identify and trim extreme sampling weights. *ASA Proceedings,* Survey Research Methods Section. 225-230.

Rizzo, L. 2014. A Rake-Trim SAS Macro and Its Uses at Westat. *SAS Global Forum*. 1627-2014.

SAS, Version 9.3. 2009. *SAS Institute, Inc.* [computer software].