

Practical Considerations in Raking Survey Data

Michael P Battaglia*, David C Hoaglin†, Martin R Frankel‡

Tags: survey practice

DOI: [10.29115/SP-2009-0019](https://doi.org/10.29115/SP-2009-0019)

Survey Practice

Vol. 2, Issue 5, 2009

Practical Considerations in Raking Survey Data

INTRODUCTION

A survey sample may cover segments of the target population in proportions that do not match the proportions of those segments in the population itself. The differences may arise, for example, from sampling fluctuations, from nonresponse, or because the sample design was not able to cover the entire target population. In such situations one can often improve the relation between the sample and the population by adjusting the sampling weights of the cases in the sample so that the marginal totals of the adjusted weights on specified characteristics, referred to as control variables, agree with the corresponding totals for the population. This operation is known as raking ratio estimation (Deville and Särndal 1992; Kalton 1983), raking, or sample-balancing, and the population totals are usually referred to as control totals. Raking is most often used to reduce biases from nonresponse and noncoverage in sample surveys.

Raking usually proceeds one variable at a time, applying a proportional adjustment to the weights of the cases that belong to the same category of the control variable. The initial design weights in the raking process are often equal to the inverse of the selection probabilities and may have undergone some adjustments for unit nonresponse and noncoverage. The weights from the raking process are used in estimation and analysis.

The adjustment to control totals is sometimes achieved by creating a cross-classification of the categorical control variables (e.g., age categories×gender×race×household-income categories) and then matching the total of the weights in each cell to the control total. This approach, however, can spread the sample thinly over a large number of adjustment cells. It also requires control totals for all cells of the cross-classification. Often this is not feasible (e.g., control totals may be available for age×gender×race but not when

* **Institution:** David Izrael, Abt Associates, Inc.

† **Institution:** Abt Bio-Pharma Solutions, Inc.

‡ **Institution:** Baruch College, City University of New York

those cells are subdivided by household income).

The use of marginal control totals for single variables (i.e., each margin involves only one control variable) often avoids many of these difficulties. In return, of course, the two-variable (and higher-order) weighted distributions of the sample are not required to mimic those of the population.

The next two sections discuss the raking algorithm and its convergence. Subsequent sections discuss control totals and several issues that arise in practical applications: two-variable margins, raking at the state level in national surveys, maintaining adjustments for nonresponse and noncoverage, surveys that involve screening, and weight trimming.

BASIC RAKING ALGORITHM

The procedure known as raking adjusts a set of data so that its marginal totals match control totals on a specified set of variables. The term “raking” suggests an analogy with the process of smoothing the soil in a garden plot by alternately working it back and forth with a rake in two perpendicular directions.

In a simple 2-variable example the marginal totals in various categories for the two control variables are known from the entire population, but the joint distribution of the two variables is known only from a sample. In the cross-classification of the sample, arranged in rows and columns, one might begin with the rows, taking each row in turn and multiplying each entry in the row by the ratio of the population total to the weighted sample total for that category, so that the row totals of the adjusted data agree with the population totals for that variable. The weighted column totals of the adjusted data, however, may not yet agree with the population totals for the column variable. Thus the next step, taking each column in turn, multiplies each entry in the column by the ratio of the population total to the current total for that category. Now the weighted column totals of the adjusted data agree with the population totals for that variable, but the new weighted row totals may no longer match the corresponding population totals.

This process continues, alternating between the rows and the columns, and close agreement on both rows and columns is usually achieved after a small number of iterations. The result is a tabulation for the population that reflects the relation of the two control variables in the sample. Raking can also adjust a set of data to control totals on three or more variables. In such situations the control totals often involve single variables, but they may involve two or more variables.

Ideally, one should rake on variables that exhibit an association with the key survey outcome variables and that are related to nonresponse and/or noncoverage. This strategy will reduce bias in the key outcome variables. In practice, other considerations may enter. A variable such as gender may not be strongly related to key outcome variables or to nonresponse, but raking on it

may be desirable to preserve the “face validity” of the sample.

CONVERGENCE

Convergence of the raking algorithm has received considerable attention in the statistical literature, especially in the context of iterative proportional fitting for log-linear models (Bishop, Fienberg, and Holland 1975), where the number of variables is at least 3 and the process begins with a different set of initial values in the fitted table (often 1 in each cell). For raking survey data the iterative raking algorithm generally converges after a small number of iterations, say 3 to 10.

Convergence can, however, sometimes require a large number of iterations. Oh and Scheuren (1978) note that the available convergence proofs make strong assumptions about the cell counts in the cross-classification of the raking variables—that no cells are empty or that some particular combination of nonempty cells is present. They recommend setting up the raking problem in a “sensible” manner to avoid: 1) imposing too many marginal constraints on the sample, 2) defining marginal categories that contain a very small percentage of the sample, and 3) imposing contradictory constraints on the sample.

Our experience indicates that, in general, raking on a large number of variables can slow the convergence process. However, other factors also affect convergence. One is the number of categories of the raking variables. Convergence will typically be slower for raking on 10 variables each with 5 categories than for 10 variables each with only 2 categories. A second factor is the number of sample cases in each category of the raking variables. Convergence may be slow if any categories contain fewer than 1% of the sample cases. A third factor is the size of the difference between each control total and the corresponding weighted sample total prior to raking. If some differences are large, the number of iterations will typically be higher.

One simple definition of convergence requires that each marginal total of the raked weights be within a specified tolerance of the corresponding control total. Tolerance can be defined in absolute terms (e.g., maximum difference less than 10) or in relative terms (e.g., maximum difference less than 0.1 percentage point). As noted above, in practice, when a number of raking variables are involved, one must check for the possibility that the iterations do not converge (e.g., because of sparseness or some other feature in the full cross-classification of the sample). One can guard against this possibility by also setting an upper limit on the number of iterations (e.g., 75). As elsewhere in data analysis, it is sensible to examine the sample (including its joint distribution with respect to all the raking variables) *before* doing any raking. For example, if the sample contains no cases in a category of one of the raking variables, it will be necessary to revise the set of categories and their control totals (say, by combining categories). We recommend, at a minimum, checking the unweighted percentage of sample cases and the percentage of control cases in each category

of each raking variable. Small categories in the sample or in the control totals (say under 2%) are potential candidates for collapsing. This step will also reduce the chance of creating very unequal weights in raking. Category collapsing always needs to be done carefully, and in some instances it may be important to retain a small category in the raking.

SOURCES AND CHOICES OF CONTROL TOTALS

Surveys that use demographic and socioeconomic variables for raking must locate a source for the control totals. Examples of sources of control totals available in the United States include the 2000 U.S. Census short-form data, the 2000 U.S. Census long-form data, the 2000 U.S. Census 5-Percent Public Use Microdata Sample (PUMS) files, the annual March Current Population Survey (CPS), U.S. Census Bureau population estimates, the American Community Survey (ACS) published estimates, the ACS 2005–2007 PUMS, and private-sector sources such as Claritas, Inc. The ACS is a rolling sample of housing units consisting of around 1.94 million housing units per year. We have used the 2005–2007 ACS PUMS to develop control totals at the state and sub-state level.

If control totals come from more than one source, it is important to make sure that control totals from all sources add to the same population total. If not, the raking will not converge if one is using a maximum-absolute-difference convergence criterion.

One must also consider how the variables are measured. For example, a telephone survey may ask a single question to obtain household income. The source for the control totals, however, may have an income variable that is constructed from a series of questions about income from several sources (wages, cash-assistance programs, interest, dividends, etc.). One needs to consider carefully whether using income as a raking variable makes sense. If the sample is thought to substantially underrepresent low-income persons, then raking on income may be preferred. If, on the other hand, there is concern that the survey is measuring income very differently from the source of the control totals, then consideration should be given to raking on a proxy variable such as educational attainment or even a dichotomous poverty-status variable.

Control totals usually do not come with a “missing” category. The same variable in the survey may have a nontrivial percentage of cases that fall in a DK or Refused category. In this situation it may be possible to impute for item nonresponse in the survey before the raking takes place. When imputation is not feasible, the following procedure can be used to adjust the control totals. Run a weighted frequency distribution on the raking variable in order to determine the percentage of sample cases that have a missing value (e.g., 4.3%). Allocate 4.3% of the overall control total to a newly created missing category (e.g., 4.3% of 1,500,000=64,500). Reapportion the control totals in the other categories so that they add to the reduced control total (1,500,000 –

64,500=1,435,500). After raking, the weighted distribution of the sample will agree with the revised control totals and will reflect a 4.3% missing-data rate in weighted frequencies and tabulations.

INCLUSION OF TWO-VARIABLE RAKING MARGINS

Raking can be viewed as analogous to fitting a main-effects-only model. Because of sample size limitations and/or availability of only one-variable (factor or dimension) control totals, many raking applications follow this approach. In some situations it may be important to fit a two-variable interaction to the data. For example, one is planning to rake on Variables A, B, C, and D. However, control totals for Variable C crossed with Variable D are available and exhibit a strong interaction (e.g., persons aged 0–17 years are more likely to be Hispanic than persons aged 65+ years). If the cell counts in the C×D margin of the sample are large enough to support fitting a C×D interaction, one would rake on three margins: A, B, and C×D. It is not necessary also to rake on separate margins for Variables C and D. If, however, the C×D raking margin involved collapsing, one could consider also raking on one-variable margins for Variables C and D without any collapsing of their categories.

RAKING AT THE STATE LEVEL IN A LARGE NATIONAL SURVEY

Some large national surveys stratify by state and are designed to yield state estimates. The resulting total national sample is usually very large. The survey analysts seek to provide national estimates as well as state estimates. Often one sets up raking control totals at the state level and carries out 51 individual rakings. Assume those rakings use Variables A, B, and C; but the number of categories of each variable is limited because of the state sample sizes.

For example, one might collapse Variables A, B, and C differently by state. If Variable A were race/ethnicity, one might be able to use Hispanic as a separate category in California, but not in Vermont because of the small sample size. After the 51 rakings one might compare the weighted distributions of Variables A, B, and C with national control totals and observe some differences that are caused by the state-level collapsing of categories.

If having precise weighted distributions at the national level is important for analytic or “face validity” reasons, one can use the following raking technique. Set up a single raking that includes margins for State × A, State × B, and State × C (i.e., combine the 51 individual state rakings into a single raking). Then add detailed national margins for Variables A, B, and C.

Another, similar example would involve adding Variable D as a national raking margin because its control total is available only at the national level (e.g., household income).

This type of raking can also be applied to a state sample that has been stratified into sub-state areas.

MAINTAINING PRIOR NONRESPONSE AND NONCOVERAGE ADJUSTMENTS IN THE FINAL WEIGHTS

Frankel et al. (2003) have discussed methods based on data on interruptions in telephone service (of a week or longer in the past 12 months) to compensate for the exclusion of persons in nontelephone households in random-digit-dialing surveys. One typically adjusts the base sampling weights of persons with versus without an interruption in telephone service. The resulting interruption-based weight adjusts for the noncoverage of nontelephone households. If one then rakes the sample on age, sex, and race, the impact of the nontelephone adjustment may be diluted somewhat, even though the raking starts with the interruption-based weight.

In that situation it generally makes sense to create weighted control totals (using the interruption-based weight) from the sample for persons residing in households with versus without an interruption in telephone service. These weighted control totals should be ratio-adjusted so that they have the same sum as the age, sex, and race control totals. For example, if the age, sex, and race margins sum to 180,000,000 persons, then the interruption margin must also sum to 180,000,000. The raking would use the four variables instead of just three and would ensure that the nontelephone adjustment is fully reflected in the final weights. This approach would be appropriate where the interruption-in-telephone-service category could be small (e.g., in states where telephone coverage is very high), but one still wants to maintain that small category in the raking.

RAKING SURVEYS THAT SCREEN FOR A SPECIFIC TARGET POPULATION

A common survey model for obtaining interviews with a specific target population is to screen a sample of households for the presence of members of the target population. An example is children with special health care needs. The screening interview collects a roster of children with, say, their age, sex, and race, and determines whether each child has special health care needs. If the household contains one child with special health care needs, a detailed interview is conducted for that child. If the household has two or more such children, one is selected at random for the detailed interview. Of course, the interview response rate will be less than 100%, because some parents will not agree to do the detailed interview.

Assume that, in a national telephone survey, the survey analysts need to look at the prevalence of children with special health care needs, and they will also be analyzing the detailed-interview data. In this situation one would calculate the usual base sampling weights, make adjustments for unit nonresponse, and possibly make a noncoverage adjustment to compensate for the exclusion of children living in nontelephone households if warranted. One first obtains control totals for age, sex, and race in the U.S. population aged 0–17 years. One

then rakes the entire sample of children in the screened households to those control totals, because that sample is a sample of children aged 0–17 in the U.S. The resulting screener weights can then be used to estimate the prevalence of children with special health care needs in the U.S.

That screener weight would typically serve as the input weight in the calculation of weights for the children with completed detailed interviews. As part of that calculation process one also seeks to weight the detailed-interview sample by age, sex, and race. Of course, control totals are unlikely to be available for children with special health care needs. One can, however, use the screener weight and the sample of children with special health care needs identified in the screened households to form weighted control totals for age, sex, and race and then use those totals in raking the detailed-interview weights. This method ensures that the age distribution of children with special health care needs from the screener sample will agree exactly with the distribution in the detailed-interview data.

THE ORIGINAL IHB RAKING MACRO

Izrael et al. (2000) introduced a SAS macro for raking (sometimes referred to as the IHB raking macro) that combines simplicity and versatility. The IHB raking macro was enhanced to increase its utility and convergence diagnostics (Izrael et al. 2004). The IHB SAS macro produces diagnostic output that contains the following information: iteration number, name of variable currently being raked on, and marginal control total and calculated total weight for each level of the current raking variable, along with their difference and also percentage difference. At termination, the macro gives the iteration number at which termination occurred and the reason, which is either that the tolerance was met or that the process did not converge. The macro also writes diagnostics into the SAS LOG, from several of the checks that it makes.

Exhibit 1 illustrates the use of the macro with an example involving two raking variables, VARIABLE1 (containing 3 categories) and VARIABLE2 (containing two categories). The IHB raking macro output shows the weighted distribution of each of the two variables at each iteration (Calculated Margin and Calculated % columns), the corresponding marginal control totals (Marginal Control Total and Marginal Control % columns), and the differences between each control total and weighted sample total (Difference and Difference in % columns). At the start of the raking the sum of the design weights equals 44,850.82, compared to the control total of 62,800.30. After adjusting the design weights so that the sum of the weights in each of the 3 categories of VARIABLE1 equals the corresponding control total, the sum of the weights now equals 62,800.30. Iteration 1 involves first adjusting the weights so that the control totals for the 3 categories of VARIABLE1 (22,154.39, 16,533.88, and 24,112.03, respectively) are satisfied and then adjusting the resulting weights so that the control totals for the two categories of VARIABLE2 (30,697.33, and 32,102.97, respectively) are met. With the

convergence tolerance set to a difference of 1, the raking converged after 3 iterations.

VARIABLE1, Iteration - 1

VARIABLE1	Calculated margin	Marginal Control Total	Difference	Calculated %	Marginal Control %	Difference in %
1	15915.87	22154.39	6238.52	35.486	35.278	0.209
2	10912.05	16533.88	5621.83	24.330	26.328	-1.998
3	18022.90	24112.03	6089.13	40.184	38.395	1.789
	=====	=====		=====	=====	
	44850.82	62800.30		100.00	100.00	

VARIABLE2, Iteration - 1

VARIABLE2	Calculated margin	Marginal Control Total	Difference	Calculated %	Marginal Control %	Difference in %
1	32684.74	30697.33	-1987.40	52.046	48.881	3.165
2	30115.56	32102.97	1987.40	47.954	51.119	-3.165
	=====	=====		=====	=====	
	62800.30	62800.30		100.00	100.00	

VARIABLE1, Iteration - 2

VARIABLE1	Calculated margin	Marginal Control Total	Difference	Calculated %	Marginal Control %	Difference in %
1	22102.81	22154.39	51.586	35.195	35.278	-0.082
2	16442.32	16533.88	91.553	26.182	26.328	-0.146
3	24255.17	24112.03	-143.139	38.623	38.395	0.228
	=====	=====		=====	=====	
	62800.30	62800.30		100.00	100.00	

VARIABLE2, Iteration - 2

VARIABLE2	Calculated margin	Marginal Control Total	Difference	Calculated %	Marginal Control %	Difference in %
1	30708.98	30697.33	-11.6455	48.899	48.881	0.019
2	32091.32	32102.97	11.6455	51.101	51.119	-0.019
	=====	=====		=====	=====	
	62800.30	62800.30		100.00	100.00	

VARIABLE1, Iteration - 3

VARIABLE1	Calculated margin	Marginal Control Total	Difference	Calculated %	Marginal Control %	Difference in %
1	22154.09	22154.39	0.29992	35.277	35.278	-0.000
2	16533.34	16533.88	0.53717	26.327	26.328	-0.001
3	24112.87	24112.03	-0.83710	38.396	38.395	0.001
	=====	=====		=====	=====	
	62800.30	62800.30		100.00	100.00	

VARIABLE2, Iteration - 3

VARIABLE2	Calculated margin	Marginal Control Total	Difference	Calculated %	Marginal Control %	Difference in %
1	30697.40	30697.33	-0.068148	48.881	48.881	0.000
2	32102.90	32102.97	0.068148	51.119	51.119	-0.000
	=====	=====		=====	=====	
	62800.30	62800.30		100.00	100.00	

**** Program terminated at iteration 3 because all calculated margins differ from Marginal Control Totals by less than 1

Exhibit 1 Example of Raking On Two One-Variable Margins Using the Original IHB SAS Macro.

Because Variable 2 is the last variable adjusted, there is exact agreement with its control totals. As general guidance we recommend making the most important control variable the last variable in the iteration.

In Exhibit 1 the Calculated % column shows the weighted percentage distribution of the sample, the Marginal Control % shows the percentage distribution of the control totals, and the Difference in % column shows the difference between the two preceding columns.

WEIGHT TRIMMING

One limitation of the original IHB macro for raking is that it does not place any limits on the highest and lowest weight values. In some situations the raking may converge, but the resulting weights exhibit considerable variability, as measured by the ratio of the highest weight to lowest weight and by the design effect due to weighting ($1+cv^2$, where cv is the coefficient of variation of the weights). Weight trimming increases extremely low weights and decreases extremely high weights to reduce their impact on the variance of the estimates, especially for subgroup estimates. For example, all weights that are less than L are increased to L , and all weights that are greater than U are reduced to U . One consequence of the trimming of low and high weight values is that the weights of the entire sample will not add to the population size. Although weight trimming is a separate topic from raking, they are certainly related in the sense that weight trimming typically takes place at the last step in the weight calculations, which is often raking. The objective of weight trimming is to reduce the mean squared error (MSE) of the key outcome estimates. Trimming low and high weight values generally lowers sampling variability but may incur some bias. The MSE will be lower if the reduction in variance offsets the increase in bias. We developed two alternative weight-trimming methods. Both are implemented during the raking process in order to ensure that: 1) limits are placed on low and high weight values in the final weights, and 2) the convergence criteria are satisfied, and the weights sum to the population total. The 2009 SAS Global Forum paper by Izrael, Battaglia and Frankel on weight trimming can be found at:

[http://www.abtassociates.com/attachments/Extreme_Survey_Weight_Adjustment_as_a_Component_of_Sample_Balancing_\(a.k.a.Raking\).pdf](http://www.abtassociates.com/attachments/Extreme_Survey_Weight_Adjustment_as_a_Component_of_Sample_Balancing_(a.k.a.Raking).pdf)

SUMMARY

Raking is often the last weighting step in a survey before producing estimates and analyses. It is sometimes relied on as a “black box” that will improve the quality of the sample by reducing nonresponse bias. We have given some background on how raking works, discussed the convergence process, and indicated items that should be checked before and after the raking. Brick et

al. (2003) also discuss issues that one should be aware of when using raking. Estimating standard errors is more complicated if raking has been used to develop the final weights. Readers interested in variance-estimation issues related to raking can consult Deville and Sarndal (1992) and Brick et al. (2000).

The new SAS macro is available for free at:

<http://www.abtassociates.com/Preview.cfm?PageID=40858&FamilyID=8600>

REFERENCES

- Bishop, Y.M.M., S.E. Fienberg, and P.W. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.
- Brick, J.M., J. Montaquila, and S. Roth. 2003. "Identifying Problems with Raking Estimators." In *2003 Proceedings of the Annual Meeting of the American Statistical Association [CD-ROM]*, American Statistical Association, Alexandria, VA, 710–17.
- Brick, J.M., D. Morganstein, and R. Valliant. 2000. "Analysis of Complex Survey Data Using Replication." <http://www.westat.com/wesvar/techpapers/ACS-Replication.pdf>.
- Deming, W.E. 1943. *Statistical Adjustment of Data*. New York: Wiley.
- Deville, J.C., and C.E. Särndal. 1992. "Calibration Estimators in Survey Sampling." *Journal of the American Statistical Association* 87: 376–82.
- Frankel, M.R., K.P. Srinath, D.C. Hoaglin, M.P. Battaglia, P.J. Smith, R.A. Wright, and M. Khare. 2003. "Adjustments for Non-Telephone Bias in Random-Digit-Dialling Surveys." *Statistics in Medicine* 22: 1611–26.
- Izrael, D., D.C. Hoaglin, and M.P. Battaglia. 2000. "A SAS Macro for Balancing a Weighted Sample." In *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference*, SAS Institute Inc., Cary, NC, 1350–55.
- . 2004. "To Rake or Not to Rake Is Not the Question Anymore with the Enhanced Raking Macro." In *Proceedings of the Twenty-Ninth Annual SAS Users Group International Conference*, SAS Institute Inc, Cary, NC. Paper 207.
- Kalton, G. 1983. *Compensating for Missing Survey Data*. Survey Research Center, Institute for Social Research, University of Michigan.
- Oh, H.L., and F. Scheuren. 1978. "Some Unresolved Application Issues in Raking Ratio Estimation." In *1978 Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, DC*, 723–28.