Enhancing Sampling Weights Using Raking Method

Yan Wang[1], Yilan Huang[2], Di Xiong[2], Kathy Phipps[3], Steve Y. Lee[4], Jie Shen[4],
Abrey Daniel[3], Maritza Cabezas[3], James J. Crall[4], Honghu Liu[1,2,4],

[1]David Geffen School of Medicine, UCLA, 10833 Le Conte Ave. Los Angeles, CA 90095
[2]Fielding School of Public Health, UCLA, 10833 Le Conte Ave. Los Angeles, CA 90095
[3]Oral Health Program, Los Angeles County, 3530 Wilshire Blvd. Suite 1010, Los Angeles, CA 90010
[4]School of Dentistry, UCLA, 10833 Le Conte Ave. Los Angeles, CA 90095

Abstract

Background: Survey samples are usually designed based on available information from population. Designs based on systematic sampling rarely consider variables related to non-response and non-coverage. However, sample weights can be modified when population-level variables are available such as race/ethnicity and socioeconomic status.

Methods: Raking method or sampling balancing is used to adjust design weights according to available marginal distribution of demographic variables at the population level. The weights are modified to estimate ratio variables (e.g. prevalence). We will show that the raking method improved the estimation.

Results: This method is applied to a 2018-2019 children's oral health needs assessment for kindergarten and third grade students in schools throughout Los Angeles County. The design weight is calculated from systematic sampling. We select variables that are related to non-response and non-coverage to balance the sample. The estimations by both methods are compared. We will report the estimation from both methods.

Conclusion: Raking survey data can modify sampling weights based on the information available at the population level, i.e. marginal population-level distribution. By accounting for non-response and non-coverage, the estimation is more accurate.

Key Words: sampling weight; survey design; raking; non-response; population balance

1. Background

To monitor the health of a population and estimate the prevalence of disease, a survey of the population is often conducted. It is important that the selected samples are representative of the targeted population [1]. Estimation will be more accurate and reliable if the distribution of the selected sample is similar to the distribution of population. However, due to the non-response and non-coverage, the sample may not represent the target population. In such a situation, we can improve the similarity between the sample and the population by adjusting sampling weight so that the weighted marginal totals of the sample agree with the corresponding marginal total of the population in terms of specific criteria [2]. This operation is known as Raking ratio estimation [3], or raking, or iterative proportional fitting [4], or sample-balancing. It is a post-stratification procedure used to reduce nonresponse and non-coverage biases [2] by adjust the sample weights. Raking is often used to reduce the sampling variability to match the population statistics. The term "Raking" comes from the process of smoothing the soils in the garden back and forth in two perpendicular directions [5], with respect to multiplying the weights in the cell by the ratio of population marginal total and weighted sample total back and forth until no further improvement. It was proposed by Deming and

Stephan in 1940 [6] to ensure consistency between complete counts and sample data from the US Census. It is widely used by adding external sources of information to generate estimates with good properties, e.g. adjust for under-coverage and response bias, improve the bias and variance of the estimates, and improve the reliability of survey estimates [7].

Raking method is often overlooked in survey analysis [4]. The goal of the survey is always to estimate something about the population from a sample [4] with the hope that the sample could represent the population. There are usually two methods to generalize the statistical inference from the sample to an population, model-based inference from assumed independent and identically distributed data, and design-based inference from data of known sampling scheme [4]. The latter is more useful when there is no information for the distribution of the population data.

The motivation of the study is to provide an efficient and effective way to estimate the population parameter by sample statistics. In the survey example of this paper, the parameters are the prevalence of dental caries (tooth decay), number of teeth with untreated caries, and urgent need of dental treatment. To establish the prevalence of dental caries among kindergarteners and third graders in Los Angeles County, children from a probability sample of public schools were screened. Kindergarteners is the youngest group in public grade schools. Third graders are at the transition from primary to permanent dentition.

This paper is trying to provide the best estimation of prevalence variables and count variables using weighted sample statistics. The design weight is described in detail in methods section. The rationale and process of generalizing raking weight is described as well. In the results section, the two weighting schemes are compared. In the discussion, we summarize the advantages and rationales for using raking method in survey sampling. We also describe the requirements for using raking method.

2. Methods

2.1 Survey design and sampling strategy

The 2018-2019 Los Angeles County oral health survey was designed to estimate the prevalence of tooth decay in a representative sample of kindergarten (K) and third grade (3^{rd}) children. Some school districts have different buildings for K and 3^{rd} and if both schools were included in the sampling frame, those schools would have a higher probability of selection Because of this, the sampling frame was limited to 3^{rd} grade schools. The inclusion criteria was non-virtual public schools, both traditional and charter, with 25 or more 3^{rd} grade children. The total enrollment at the time of sampling was 110,614 3^{rd} grade children in 1,262 schools. We made the following assumptions to determine the sample size.

2.1.1 Assumptions

We assume the frequency of the outcome as 50% to assume the large variance. The average 3^{rd} grade enrollment per schools is 88 (110,614/1,263). We assume the response rate is 80%. The acceptable margin of error is 2%. We also assume the design effect is 2 (based on data from other state surveys). The design effect is a correction factor that is to account for the heterogeneity between clusters with regard to the measured indicator. The required sample size is estimated by assuming a random sample, and then multiply by the design effect. This accounts for the loss of information inherent in the clustered design [8].

2.1.2 Consent and IRB

The survey was approved by the Los Angeles County Public Health and Health Services Institutional Review Board (IRB). The activity was classified as public health surveillance rather than human subjects research and was classified as exempt. The consent process varied by school district. The Los Angeles

Unified School District (LAUSD) required active consent while all other school districts used passive or opt-out consent.

2.1.3 Survey Screening Procedure

The survey consisted of a non-invasive dental screening completed by a licensed dental professional. No x-rays were taken and no dental treatment was provided.

2.1.4 Sample size

The number of schools was determined by following sample size formula.

$$\mathrm{n} = \frac{\mathrm{N\,Z^2 p(1-p)}}{\mathrm{ME^2(N-1) + Z^2 p(1-p)}} \times \mathrm{DEFF}$$

Here n is the number of schools selected to represent LA County. N is the population size. Z is determined by the 95% CI of the estimation. Here p is the expected frequency (assume 0.5 for largest sample size). ME is the margin of error. DEFF is the design effect. The number of schools is

$$n_{cluster} = \frac{4802}{88\times 80\%} = 68.21 \approx 70$$

In total, we selected 70 schools with 3rd grade. Two of the 70 schools did not include K; therefore, the two feeder schools with kindergarten were added to the sample for a total of 72 schools. In this way, the total weights is the population size as, $N = \sum_{i=1}^{70} n_i w_i = 70 \times 1580.2 = 110{,}614$.

2.1.5 Sampling intervals and weights

The sampling interval was defined by the number of 3rd graders divided by the number of schools. The sampling interval was 1580.2. We used a random number from 0 to 1 to start sampling procedure. The number was 0.454. The schools were sampled from the ordered list of schools using 0.454*1580.2 = 717.4 as the starting number. The next number was 1580.2+717.4. Each additional school in the list is to add 1580.2. The sampling frame was stratified by school district (LAUSD vs non-LAUSD). Next, the schools were ordered by geographic region (LA County's service plan areas (SPA)) and percent of children in the school eligible for the National School Lunch Program (NSLP). A systematic probability proportional to size sampling scheme was used to select 70 schools. We illustrated the sampling stage in Figure 1.

Figure 1: Survey design and sampling stage

If a school refused to participate, a school from the same sampling interval was selected as a replacement. The three-stage sampling (Figure 1) helps to assure that the sample is representative of the County in terms of geographic location and socioeconomic status.

2.2 Raking method

Based on the survey design, the variables that are both highly associated with oral health outcomes and coverage and response rate of participants are LAUSD/on-LAUSD, SPAs of LAC, and the social economic disadvantages based on NSLP. We included the additional information available for the population in LAC. The final raking weights were calculated based on these four variables with marginal distribution in population. The example of raking algorithms is shown in Figure 2. The example applied raking method on the cell (j, k) of tables formed with race/ethnicity by gender.

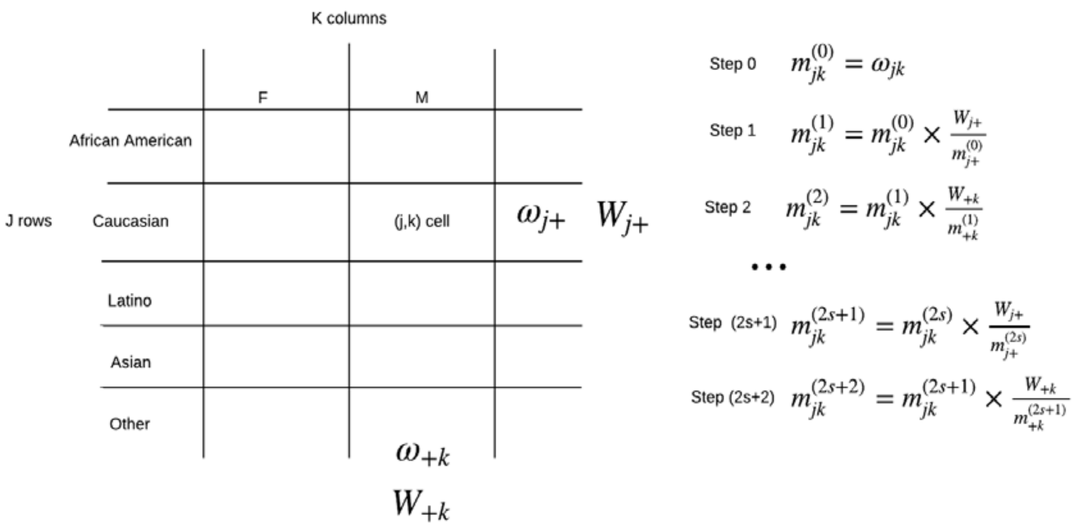

Figure 2: Example of raking algorithm using race/ethnicity and gender

The convergence of raking methods in this paper is achieved by increasing the number of iterations. Usually, if the number of variables is large or number of categories for each variable is large, the convergence rate is slow [2]. Convergence rate is affected by number of samples in each category. If the variable selected in the raking method does not converge, we will select a different variable. In the following analysis, we used the LAUSD vs non-LAUSD, SPA information, NSLP together with the race/ethnicity distribution in the population. The selection criteria for these raking variables include [2, 4], (1) the control totals of these variables add to the same population total; (2) usually no missing category; (3) the raking variables are associated with the outcome measure and nonresponse and noncoverage rate.

In this paper, we use the R package [9] *survey* to calculate the raking weight based on the population distribution. The procedure is also available in STATA [10] and SAS [1, 11].

3. Results

3.1 Sample characteristics

We screened children from 73 schools. One school was excluded from the results due to low response rate. We visited 33 LAUSD schools and 29 non-LAUSD schools. The size of school ranged from 71 to 395 students. The final sample size included 10,489 students, with response rates ranging from 32% to 95%. There were 5,897 K students and 4,592 3rd students in the sample. We excluded 68 K students and 29 3rd students because of missing values in the raking variables. The characteristics of the sample are presented in Table 1.

Table 1: Characteristics of the sample and weighted population

Characteristic	Number of Children (Unweighted)	Weighted Percent （95% CI)
Grade		
Kindergarten	5829	54.4 (53.7,55.2)
3rd	4563	45.6 (44.8,46.3)
Sex		
Female	5121	49.4 (48.4,50.5)
Male	5239	50.3 (49.2,51.3)
Missing/Unknown	32	0.3 (0.1,0.5)
Race/Ethnicity		
Asian	1393	9.7 (6.3,13.1)
Black/African American	783	7.3 (4.6,10)
Hispanic/Latino	6682	65.7 (59.1,72.4)
Other	402	3.2 (2.2,4.1)
White	1132	14.1 (9.1,19.1)
Socioeconomically disadvantaged (SES) by NSLP		
No	3101	29.6 (22.8,36.4)
Yes	7291	70.4 (63.6,77.2)
Service Planning Area		
SPA 1	824	5.3 (0.2,10.4)
SPA 2	1874	20.7 (11.1,30.4)
SPA 3	1499	16.7 (8.4,25)
SPA 4	614	9 (2.3,15.7)
SPA 5	513	4.4 (0,9.3)
SPA 6	1299	13.5 (5.6,21.4)
SPA 7	1464	14.9 (6.2,23.7)
SPA 8	2305	15.5 (6.9,24.1)

We had three oral health outcomes, the prevalence of decay experience, the prevalence of untreated decay and treatment urgency. The prevalence of untreated decay was measured by the percentage of children who had at least one tooth with untreated decay and the number of teeth with untreated decay. The prevalence of decay experience was measured by the percentage of children who had at least one tooth with untreated decay or treated decay and the number of teeth with untreated decay or treated decay. Treatment urgency was categorized as urgent, early and none. We estimated the percentage of children who needed urgent or early dental care. For those categorical variables, we reported percentage with standard error in Table 2. For count variables of number of teeth, we reported the mean and standard error in Table 2. The analysis for K and 3rd were separated as we expected that their oral health outcomes were different. For example, third should have a higher number of teeth with decay experience.

Table 2: Compare outcomes using design weight and raking weight

	Grade	Kindergarteners		Third Graders	
Oral Health Related Outcomes	**Weighting method**	**mean or %**	**Std error**	**mean or %**	**Std error**
Active Caries					
Prevalence (%)	Design	18.86	0.91	20.9	1.19
	Raking	18.84	0.69	20.71	1.02
Number of untreated decay	Design	0.48	0.03	0.42	0.03
	Raking	0.48	0.03	0.42	0.02
Caries Experiences					
Prevalence (%)	Design	46.89	1.92	66.36	2.02
	Raking	46.83	1.08	64.7	1.05
Number of untreated/treated decay	Design	2.27	0.12	3.15	0.14
	Raking	2.26	0.07	3.04	0.08
Treatment urgency					
Urgent (%)	Design	1.77	0.25	2.08	0.29
	Raking	1.73	0.23	2.07	0.26
Early (%)	Design	16.6	0.8	18.16	1.11
	Raking	16.62	0.61	17.99	1.01

3.3 Sensitivity analysis of variable selection

We used mean standard error to compare different selection of raking variables. The results indicated the best selection is to use all four variables (Table 3). We included two extra variables, gender and primary language to compare the selection results. We concluded the selection of raking variables, LAUSD vs Non-LAUSD, SPA, SES (NSLP), and race/ethnicity, generated the smallest average standard error among all outcomes.

Table 3: Compare average standard error using different set of raking variables

LAUSD/Non-LAUSD	SPA	SES	Race/Ethnicity	Average std error	
				Design	**Raking**
x	x	x	x	0.734	0.513
x		x	x	0.734	0.540
	x	x	x	0.734	0.533
x	x	x		0.734	0.521
LAUSD/Non-LAUSD + SPA + SES + Race/Ethnicity + Gender				0.736	0.515
LAUSD/Non-LAUSD + SPA + SES + Race/Ethnicity + Primary Language				0.735	0.523

4. Discussion

Raking weight was used to estimate the prevalence of the outcomes. Compared with the design weight based on systematic sampling, raking method provided similar point estimates but with better precision (i.e., smaller standard errors). Raking can be done with known population marginal totals. The algorithm involves repeatedly estimating weights as shown in Figure 2. Raking algorithms match the totals in the sample to the population total marginally. Convergence is achieved by increasing the default number of iterations in the software procedure. In this paper, the raking algorithm is applied to only four variables in population, the convergence was not a concern.

Ideally data is collected via simple random sample, with model-based inference to be used. However, in practice, due to size of sampling frame and/or the limitation of time and budget, samples are collected from finite population, clustered respondents, and unequal selection probabilities of respondents (not random). Commonly used sampling methods include stratified sampling, cluster sampling, or both. Sampling weights (inverse of sampling probabilities) are used to account for the design-based inequalities. The sample, however, may not be representative of the population due to nonresponse rate and noncoverage issues. Potentially, the sample may not have similar distributions of many variables as the population, such as income, race and ethnicity. In this situation, raking can be used to balance the sample by matching the marginal distributions between sample and population.

In this paper, we did not perform the weight trimming process, which is used to truncate the extreme weight values to reduce their impact in the final estimation [5]. This process may leave the unequal of weighted sample total and population size. Also, by removing extreme values, it may lead to bias because the sample variability could be decreased. Currently, there is no existing rules for weight trimming procedures. The size of LA County smile survey is still small and the number of raking variables we used is four. Therefore, we did not use the weight trimming procedure. The aim of this paper is focused on the raking algorithms for smile survey in Los Angeles County. The smile survey also collected other variables, such as height, weight, and sealants and caries status of molars. Please refer to the county report for the estimate of these variables.

LA County smile survey is part of the California smile survey. The analyses of rest of smile survey of California may also apply raking algorithms to ensure the weighted estimates will be representative for its targeting populations level. Raking approach is becoming more and more important when reweighting the samples with known variable distribution of population as the algorithm only requires the marginal distribution of the variables instead of joint distributions and is more flexible.

Acknowledgements

We would like to thank all the hygienists, volunteers, participating schools, and all parents and children who participated the survey. We appreciate the great help from Los Angeles County Office of Education and the L.A. Trust for Children's Health. Research reported in this publication was supported by the National Institutes of Health under award number **T32MH080634** and **T37MD01425**. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or Los Angeles County.

References

1. Izrael D, Battaglia MP, Frankel MR: **Extreme survey weight adjustment as a component of sample balancing (aka raking)**. In: *Proceedings from the Thirty-Fourth Annual SAS Users Group International Conference: 2009*; 2009.
2. Battaglia MP, Izrael D, Hoaglin DC, Frankel MR: **Tips and tricks for raking survey data (aka sample balancing)**. *Abt Associates* 2004, **1**:4740-4744.
3. Kalton G: **Compensating for missing survey data**. 1983.
4. Anderson L, Fricker Jr RD: **Raking: An important and often overlooked survey analysis tool**. *Phalanx* 2015, **48**(3):36-42.
5. Battaglia MP, Izrael D, Hoaglin DC, Frankel MR: **Practical considerations in raking survey data**. *Survey Practice* 2009, **2**(5):1-10.
6. Deming WE, Stephan FF: **On a least squares adjustment of a sampled frequency table when the expected marginal totals are known**. *The Annals of Mathematical Statistics* 1940, **11**(4):427-444.
7. Brick JM, Montaquila J, Roth S: **Identifying problems with raking estimators**. In: *annual meeting of the American Statistical Association, San Francisco, CA: 2003*; 2003.
8. Katz J, Zeger SL: **Estimation of design effects in cluster surveys**. *Annals of epidemiology* 1994, **4**(4):295-301.
9. Lumley T, Lumley MT: **Package 'survey'**. In.; 2020.
10. Kolenikov S: **Calibrating survey data using iterative proportional fitting (raking)**. *The Stata Journal* 2014, **14**(1):22-59.
11. Izrael D, Hoaglin DC, Battaglia MP: **A SAS macro for balancing a weighted sample**. In: *Proceedings of the twenty-fifth annual SAS users group international conference: 2000*: Citeseer; 2000: 9-12.