Some Notes On Cell Collapsing And Its Effect On Replicate Variance Estimates With The Delete-A-Group Jackknife Variance Estimator

Samson A. Adeshiyan[1] and Katherine J. Thompson[2]
[1]Demographic Statistical Methods Division, U.S. Census Bureau, Washington, DC 20233
[2]Office of Statistical Methods and Research for Economic Programs, U.S. Census Bureau

1. Introduction

Surveys often employ weight adjustment procedures to compensate for unit non-response. Such procedures are designed to reduce the bias of the survey estimate constructed from respondent data (Little and Rubin, 2002). Additionally, often survey methodologists frame unit non-response as a random second-phase sample of respondents from the originally sampled units. Given either a uniform (missing- completely-at random (MCAR)) response mechanism or a missing-at-random (MAR) response mechanism in each weighting class h, then the second phase sample comprises a Bernoulli sample with a random sample of respondents -- each with response probability p_h -- from the first-phase sampled units in each weighting class (Sarndal, C., Swensson, B., and Wretman, J., 1992, pp. 62-65 and Kott, 1994).

In this note, we consider a commonly-used adjustment-to-sample model described in Kalton and Flores-Cervantes (2003): the unit non-response weighting adjustment factor for a weighting cell is calculated as the sample-weighted number of sampled units in the weighting cell divided by the sample-weighted number of responding units in the weighting cell, i.e., the "quasi-randomization" estimator (Oh and Scheuren, 1983). This adjustment is valid under a MAR or covariate-dependent response model. If the weighting cells correspond to sampling strata, then this is equivalent to weighting by the unweighted inverse response rate, as recommended by Little and Vartivarian (2005).

In practice, weighting cells are designed to contain at least an expected **minimum number** of respondents, where this minimum is required to be greater than or equal to one. When this condition is not satisfied, then weighting cells are collapsed according to predetermined criteria, usually determined by a response propensity model or cell mean model. Ideally, the probability of response should differ by weighting cell as should the cell mean for key data items. Under these conditions, collapsing weighting cells induces an additional bias in the survey **estimates**. Again, this additional estimation bias is reduced by judiciously grouping weighting cells with either similar response propensities or similar cell means (e.g., collapsed strata). Hereafter, we refer to cell collapsing rules based on a specified minimum number of respondents as the **counts only** collapsing criteria.

In addition to defining a minimum number of weighting cell respondents, surveys often place a "cut-off" on the weight adjustment factor, i.e., collapse weighting cells if the adjustment factor is "too large" (Lohr, p. 267). This cut-off is designed to reduce the estimated variance at the risk of increasing the bias of the estimator. Such collapsing cut-off criteria are combined with minimum cell count criteria and are hereafter referred to in this paper as the **counts plus factor** collapsing criteria. To reduce estimated variances, some surveys further place a "cap" on the size of the collapsed adjustment factors, accepting the increase in estimation bias in favor of the reduced variance component. We do not consider an estimation procedure that places size restrictions on **final** estimated response adjustment factors.

This note addresses the question of appropriate **replicate variance estimation procedures** when using weighting to account for unit non-response. Under the **counts only** and **counts plus factor** collapsing criteria, we consider

- Using the full sample cell weighting cells in each replicate (**full sample cells)**; or
- Determining the cell collapsing pattern individually for each replicate (**replicate cells)**

via a simulation study modeled along the lines of a typical business survey. We compare the variance estimation effects of **replicating** the weighting cell collapsing procedure in terms of relative bias of the variance estimates

This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the Census Bureau.

and the relative stability of the variance estimates, testing the sensitivity of our results to response model-assumptions by employing a variety of uniform and MAR response propensity models in our simulation study.

For operational reasons, any replication procedure should perform cell collapsing when the number of respondents in a replicate weighting cell is zero. If not, the replicate estimates are biased. This requirement often motivates fully replicating the weighting cells when using the **counts only** collapsing criterion.

Our research utilizes a delete-a-group jackknife (DAGJ) variance estimator found in Wolter (1985). This is a variation of the estimator presented in Kott (2001). This variance estimator has excellent statistical properties in terms of consistency (for estimates of totals) while allowing sufficient "dropped sample" in each replicate for observation (c.f., the stratified jackknife). We discuss this variation in detail in Section 2.1, along with our justification for employing it in this research.

This paper explores the cell collapsing effects on DAGJ variance estimates via a simulation study, considering one non-response weight adjustment procedure. Our simulation study is modeled along a typical business survey but is representative of no particular program. The primary goal of this study was to explore empirically results that we obtained theoretically. Using Taylor linearization approximations, it can be shown that the DAGJ replicates are model-unbiased estimates of the full sample total when the full sample weighting cells and replicate weighting cells are the same, assuming a MAR response mechanism with the weighting cells corresponding to sampling strata and a **counts only** criterion is used. Under the same conditions, it can be further demonstrated that the DAGJ replicate estimates are biased when the replicate weighting cells differ from the full sample weighting cells (i.e., when replicate weighting cells are collapsed whereas corresponding full sample weighting cells are not). This collapsing bias is referenced in Rao and Shao (1999) in the context of modified half-sample and balanced repeated replication. When replicate weighting cells each have respondents, and they **do not** correspond to the full sample weighting cells, the resultant replicate variance estimates tend to be overestimates. However, when the replicates have zero respondent cells, it can be demonstrated that replicating the weighting collapsing procedure yields less-biased variance estimates under conditions that are often met by business surveys. To save space, all proofs have been omitted, but are available upon request to authors.

2. Methodology

2.1 Adjustment-to-Sample Estimation Procedures for Unit Non-response

For a particular weighting cell h, let n_h be the number of sampled cases, and r_h be the number of respondents. If y is observed for all r_h observations and a highly correlated covariate value x is known for all n_h observations then the non-response adjusted total Y_h is estimated by $\hat{Y}_h = \left(\frac{\sum_{i=1}^{n_h} w_i x_i}{\sum_{i=1}^{r_h} w_i x_i} \right) \sum_{i=1}^{r_h} w_i y_i$. So,

$\hat{Y}_h = \sum_{i=1}^{r_h} f_h w_i y_i$ where $f_h = \left(\frac{\sum_{i=1}^{n_h} w_i x_i}{\sum_{i=1}^{r_h} w_i x_i} \right)$ is the non-response adjustment factor and $f_h w_i$ is the adjustment weight for all units i in weighting cell h. In our analysis, x_i=1 for all observations, so that

$$f_h = \frac{\sum_{i=1}^{n_h} w_i}{\sum_{i=1}^{r_h} w_i}$$ (adjustment factor in the "quasi-randomization" estimator).

If a weighting cell u contains fewer than the pre-specified **count** threshold or the factor f_u is larger than the pre-specified **factor** threshold, then the weighting cell u is combined with (predetermined) weighting cell v. In this case, the adjustment factor applied to all respondent units in weighting cells u and v is $f_{uv} = \frac{\sum_{i=1}^{n_u+n_v} w_i}{\sum_{i=1}^{r_u+r_v} w_i}$.

2.2 Delete-a-group Jackknife (DAGJ) Variance Estimator

To perform delete-a-group jackknife (DAGJ) variance estimation, the non-certainty portion of the parent sample is systematically divided in K random groups. The k^{th} delete-a-group-jackknife replicate estimate is obtained by summing the weighted data from all but the k^{th} random group after increasing the design weights by a factor that accounts for the random group that was dropped, then adding the survey total from the certainty portion of the sample to each replicate. In the simulation study described below, each DAGJ replicate $\hat{Y}_{(k)}$ is given by $\hat{Y}_{(k)} = \frac{K}{K-1}(\hat{Y} - \hat{Y}_k)$ where $\hat{Y}_k$ is the estimate for the k^{th} random group $= \sum_{i \in k} w_i y_i$ and $k = 1, \ldots, K$. Employing *K/(K-1)* yields unconditionally unbiased replicate estimates as described in Wolter (1985, p. 155 and 183). Kott (2001) recommends using $n/(n - n_k)$ or (for stratified samples) $n_h / (n_h - n_{hk})$, where n_h represent the number of sampled units in stratum h and n_{hk} represents the number of sampled units in stratum h and random group k, which yields conditionally unbiased replicate estimates (conditional on the realized sample).

Wolter (1982, p.31) provides guidelines for dividing the parent sample of either a two-stage or two-phase sample into random groups for variance estimation, advocating allocation of entire first stage units (not allocation of ultimate sampling units or respondents) to random groups. The simulation applications described below follow these guidelines, i.e., sampled units are assigned to random groups.

The DAGJ variance estimate for characteristic y is given by $\hat{V}(\hat{Y}) = \frac{K}{K-1}\sum_{k=1}^{K}(\hat{Y}_{(k)} - \hat{Y}_0)^2$, where $\hat{Y}_0$ represents the full-sample estimate. The applications discussed below use a stratified simple random sample selected without replacement and have negligible sampling fractions in all strata. Furthermore in our applications, each stratum contains at least K units, so that each stratum is represented in each random group. This avoids the variance estimation bias issue discussed in Kott (2001).

For our simulation study, we employ the simple DAGJ replicate estimation method (replicate factors of *K/(K-1)*) instead of the strata-specific factors recommended in Kott (2001). This was not our original intent because Kott (2001) provides compelling evidence for his proposed adjustment under the model assumption that $y_h \sim \mu_h, \sigma_h^2$. In fact, our theoretical derivations considered both forms of DAGJ replicate estimators. In developing our theory, a distinct advantage of Kott's method for DAGJ replicate factors over our simple factors is that fewer approximation assumptions are required when no cell collapsing is performed. However, derivations with the strata-specific replicate factors are more difficult to interpret when replicate cell collapsing is required.

This aside, in our simulation study we obtained nearly unbiased DAGJ variance estimates of the population total under 100% response using fifteen random groups and our simple DAGJ variance estimator. We were unable to achieve this in our simulation with the strata-specific replicate factors, regardless of the number of random groups (we attempted 10, 14, 15, and 16 random groups). This could be an artifact of our simulation study, for which we generated a finite population with fixed characteristics (y_i), then stratified this population for sampling. This simulated population does not meet the criteria required in Kott (2001) for nearly model-unbiased variance estimation, which requires a separate model for the characteristics within each stratum.

Weighting cells in the replicates could have fewer respondents than in the full sample. Thus, certain weighting cells might need to be collapsed for a given replicate whereas they are uncollapsed for the full sample. On the other hand, cells that are collapsed in the full sample will be collapsed in the replicates as well. This situation can occur more frequently with a replication method such as the DAGJ (as opposed to the stratified jackknife) if the number of sampled units in each full sample weighting cell is fairly small.

3. Simulation Study

3.1. Design

To conduct our Monte Carlo simulation study, we generated a finite population of bivariate lognormal data for 15,000 establishments: variable *Y* is both the stratification variable and characteristic of interest variable (an "ideal" stratification scenario); and *Z* is a highly correlated variable used for forming weight adjustment cells. We stratified this population using the Dalenius-Hodges cum-root-f rule (Cochran, 1977, p. 129) into six non-certainty strata, then independently selected 1,000 stratified SRS-WOR samples of size n = 300 using the Neyman allocation with strata allocations as follows: $n_1 = 37$, $n_2 = 44$, $n_3 = 43$, $n_4 = 45$, $n_5 = 46$, and $n_6 = 85$.

In each sample, we systematically assigned sample units to fifteen random groups. Unit non-response was randomly induced as a Bernouilli trial within sample, so that the same unit could have different response status in different samples. We consider two different response mechanisms. Initially, we considered a uniform response mechanism also known as the MCAR mechanism. In this setup, let *Y* be the variable of interest and M be the missing data indicator with *M*=1 if the value of Y is missing and 0 otherwise. Under the MCAR mechanism, $P(M|Y)=P(M)$ for all *Y*. Since MCAR mechanism is often not applicable to business surveys, we then employ a more realistic MAR response mechanism. The general setup for the MAR response mechanism is that $P(M|Y)=P(M|Y_{observed})$ for all *Y*. However, we use a covariate-dependent MAR response mechanism in our simulations where given *Z*, a completely known covariate, $P(M|Y, Z)=P(M|Z)$ so that the missing indicator M is conditionally independent of *Y* given *Z*. For both the uniform response and MAR scenarios, we consider two different sets of nonresponse weighting cells: one where the unit non-response adjustment is performed within strata; the other where we employ **ten** disjoint weighting cells, determined by percentiles of *Z* in the population.

3.2. Evaluation Criteria

Let Y be the true population total of our variable of interest, and let

$\hat{Y}_s$ = estimate of the population total from sample *s* (where s indexes the sample) assuming 100% response.

Assuming some unit non-response, let

$\hat{Y}_{is}$ = non-response adjusted estimate of the population total from sample where *i* =0 when the full sample weighting cells are used from the replicates in sample *s* and is *r* otherwise.

We estimate the survey error of the two adjusted sample estimates in terms of empirical mean squared error (MSE). With complete (100%) response, the empirical MSE of the survey estimate is given by:

$$\text{MSE}\left(\hat{Y}\right)=\frac{1}{S}\sum_{s=1}^{S}\left(\hat{Y}_s-Y\right)^2 .$$

The respective empirical MSE of $\hat{Y}_{1s}$ and $\hat{Y}_{2s}$ are given by

$$\text{MSE}\left(\hat{Y}_j\right)=\frac{1}{S}\sum_{s=1}^{S}\left(\hat{Y}_{js}-Y\right)^2,$$ where *j* = 1 when the collapsing criteria is **counts only** and *j* = 2 when the collapsing criteria is **counts plus factor**.

Note that we do **not** attempt to compare the MSEs between estimates obtained from using the two procedures. These estimation properties are greatly affected by factors such the underlying distribution of the data, the survey design, and the response mechanism. Consequently, it would be unwise to make a general comparative statement about the estimation properties from either procedure from this particular simulation study.

For the three sample-based estimates above, we can calculate the following DAGJ variance estimates:

$v_s = \hat{V}(\hat{Y}_s)$ is the variance estimate from sample *s* assuming 100% response.

$v_{ijs} = \hat{V}_i(\hat{Y}_{js})$, where *i* and *j* are as defined above.

We use the following measures to evaluate the statistical properties of the variance estimators for procedure *i* (*i* = **counts only** or **counts plus factor**)

Relative bias = $\dfrac{\frac{1}{S}\sum_{s=1}^{S} v_{is}}{MSE\left(\hat{Y}_i\right)} - 1$ **Relative Stability** = $\dfrac{\left[\frac{1}{S}\sum_{s=1}^{S}\left(v_{is} - MSE\left(\hat{Y}_i\right)\right)^2\right]^{1/2}}{MSE\left(\hat{Y}_i\right)}$

With an "ideal" variance estimator, the relative bias will be near zero as will the relative stability. The relative bias and relative stability of our v_s (100% response) DAGJ variance estimates for *Y* constructed from our 1,000 samples were -0.02 and 0.11 respectively. We rely primarily on relative bias as our measurement of accuracy. We expect the stability to worsen by developing replicate-specific weighting cells, since this adds variability to the replicate estimates. Initially, we examined confidence interval coverage rates as well. However, the analysis of coverage rates is complicated since both the survey estimates and variance estimates are biased, and we were not confident that the coverage effects seen in our study would be applicable to other data sets. The coverage results are, however, available upon request from the authors.

3.3 Results

We present our simulation study results below. In each scenario, we used a data-based decision rule to develop collapsing criteria, i.e., we determined cut-off values that would force the replicate-based weighting cells to differ occasionally from corresponding full-sample weighting cells. In doing this, we **did not** simulate any conditions of zero-response in any replicate-based weighting cell. This latter condition is not intrinsically interesting, since there is very little choice in replication cells.

In Tables Two and Four, the weighting cells are identical to the strata, and the thresholds for collapsing depends on data with the minimum number of respondents in a weighting cell being 33 (**count)** and the maximum allowable factor level being 1.4 (**factor).** In Tables Three and Five, the ten weighting cells **are not identical** to strata, and the thresholds for collapsing depended on data with the minimum number of respondents in a weighting cell being 14 (**count)** and the maximum allowable factor level being 1.75 (**factor).** These thresholds were chosen to guarantee some collapsing.

3.3.1 Uniform Response Mechanism

Table Two presents our simulation results when weighting cells correspond to strata.

Table Two: Variance Estimation Measures (in Percentages) Given a Uniform Response Mechanism Where Weighting Cells Correspond to Strata

Weighting Cell Response Propensity	Collapsing Criteria	Replication Weighting Cells	Relative Bias	Relative Stability
100	n/a	Full Sample	-2	11
90	Count	Full Sample	-21	23
		Replicated	-21	23
	Count Plus Factor	Full Sample	-21	23
		Replicated	-21	23
80	Count	Full Sample	-22	24
		Replicated	386	387
	Count Plus Factor	Full Sample	-22	24
		Replicated	386	387
70	Count	Full Sample	59	61
		Replicated	59	61
	Count Plus Factor	Full Sample	36	43
		Replicated	36	43

For two of the three response propensities (90 and 70), there is no evidence of improvement in relative bias or relative stability using the replicate-based weighting cells instead of the full sample weighting cells within the same collapsing criteria.

The results for the 80-percent response propensity are less encouraging, demonstrating very large increases in relative bias when using the replicate-based weighting cells. The explanation for this follows from the earlier discussion. In these two sets of simulations, the replicate-weighting cells are collapsed more frequently than the full sample weighting cells in a high proportion of our samples, leading to overestimation of MSE with the replicate-based weighting cells. This additional collapsing bias in the replicate estimates is not present when using the full sample weighting cells to construct adjusted replicate estimates. We did not see a similar pattern with the 70 percent propensity because there was so much collapsing in the full sample. Table Three presents our simulation results when the weighting cells do not correspond to strata.

Table Three: Variance Estimation Measures (In Percentages) Given a Uniform Response Mechanism Where Weighting Cells Do Not Correspond to Strata

Weighting Cell Response Propensity	Collapsing Criteria	Replication Weighting Cells	Relative Bias	Relative Bias
100	n/a	Full Sample	-2	11
90	Count	Full Sample	-28.0	34.7
		Replicated	-28.0	34.7
	Count Plus Factor	Full Sample	-28.0	34.7
		Replicated	-28.0	34.7
80	Count	Full Sample	50.6	67.2
		Replicated	50.6	67.2
	Count Plus Factor	Full Sample	50.6	67.2
		Replicated	50.6	67.2
70	Count	Full Sample	12.3	37.4
		Replicated	12.3	37.4
	Count Plus Factor	Full Sample	11.9	37.6
		Replicated	101.4	883.0

With the higher response propensities (90 and 80), there is no evidence of improvement in relative bias or relative stability for full sample based versus replicate-based weighting cells within the same collapsing criteria. This pattern changes slightly for **counts plus factor** collapsing criteria, when again we see increases in relative bias and stability by using replicate-based weighting cells.

In many of the scenarios presented above, we obtain identical results for bias and stability respectively either for **counts only** or **counts plus factor** collapsing patterns. One explanation is that the counts only collapsing pattern completely dominated the factor-based collapsing pattern, i.e., in most cases, the cells with replicate factors that exceed the threshold also have replicate counts that are lower than the threshold. This is a fairly typical scenario: minimum respondent thresholds are generally chosen to support the sample design and to prevent variance spikes caused by overly large adjustment factors.

3.3.2 Missing-at-Random (MAR) Response Mechanism

A uniform response mechanism is a bit unrealistic for many business surveys. Often, the sampled units from larger size strata have higher response rates or are targeted for intensive analyst follow-up, and the response propensity tends to decrease proportionally to unit size. Table Four presents simulation results for two different missing-at-random (MAR) response mechanisms, where the response propensities increase as the unit size-strata increases.

Again, we see the pattern exhibited in Table Two for the 80% response propensity group, namely large increases in the relative bias of the replicate-based weighting cells over using the full sample weighting cells with either collapsing criteria. We attribute the large difference in relative biases within collapsing criteria to the simulation study design. This exaggerated effect occurs when the replicates do not have cells with zero respondents. This can be explained by our theoretical results: when there are no replicates with zero respondent cells, replicating the weighting collapsing pattern tends to increase the bias in the DAGJ variance estimates. In our simulation study, there are no replicates with zero respondent cells, so we do not capture mitigating effects on the bias from replicating the weighting collapse pattern.

Table Four: Variance Estimation Measures (in Percentages) Given a MAR Response Mechanism Where Weighting Cells Correspond to Strata

Weighting Cell Response Propensity	Collapsing Criteria	Replication Weighting Cells	Relative Bias	Relative Stability
100	n/a	Full Sample	-2	11
60,60,70, 80,80,90	Count	Full Sample	-33.5	34.2
		Replicated	110.0	110.3
	Count Plus Factor	Full Sample	-33.5	34.2
		Replicated	110.0	110.3
60,65, 70,75,80, 90	Count	Full Sample	-22.7	26.8
		Replicated	744.2	745.4
	Count Plus Factor	Full Sample	-22.7	26.8
		Replicated	744.2	745.4

Finally, Table Five presents results for two different MAR response mechanisms, using ten weighting cells determined by an auxiliary variable that is highly correlated to the characteristic of interest. This is the most general of the simulation scenarios presented, but it also makes the most use of the response model in the formation of the weighting cells.

Table Five: Variance Estimation Measures (In Percentages) Given a MAR Response Mechanism Where Weighting Cells Do Not Correspond to Strata

Weighting Cell Response Propensity	Collapsing Criteria	Replication Weighting Cells	Relative Bias	Relative Stability
100	n/a	Full Sample	-2	11
50,50, 60,60,70,70, 80,80,90 90	Count	Full Sample	26.2	51.1
		Replicated	26.2	51.1
	Count Plus Factor	Full Sample	-26.9	39.2
		Replicated	40.4	102.5
70,70, 70,70,80,80, 80,80,90 90	Count	Full Sample	12.3	37.4
		Replicated	12.3	37.4
	Count Plus Factor	Full Sample	11.9	37.6
		Replicated	101.4	883.0

Again with the **counts only** collapsing criteria, neither set of variance estimates (full sample versus replicated) appears to have an advantage over the other. In contrast, incorporating the factor-thresholds into replicate weighting procedures appears to have detrimental effects on the replicate MSE with the **counts plus factor** collapsing criteria.

4. Conclusion

There is a practical advantage to full replication when minimum weighting cells sizes are based on respondent counts: it eliminates any concerns about zero-respondent replicate weighting cells. In our simulation, the replicate weighting cells always contain more than one respondent. In the majority of our simulation scenarios, there is no difference in the relative bias of the variance estimate computed with original weighting cells or replicate-based weighting cells with the **counts only** collapsing criteria. Moreover, in the case where a replicate weighting cell is empty, there are theoretical advantages to replicating the weighing cells (at least with our considered variance estimator). However, it should be noted that when there are differences between corresponding variance estimates, the relative biases of the fully replicated variances are substantially larger than their counterparts constructed from the full sample weighting cells.

In contrast, there are no genuine theoretical or practical advantages of incorporating factor-thresholds into replicate cell collapsing criteria. Moreover, many of our simulation results provide some evidence of detrimental effects on MSE estimates when using replicate-based weighting cells with the **counts plus factor** collapsing criteria. Thus, we recommend excluding factor thresholds when developing replicate weighting-cells from any replicate variance procedure. This recommendation is consistent with the current practice for most surveys.

The decision whether to use the full-sample weighting cells or the replicate-based weighting cells when using respondent **counts** to determine collapsing is less straightforward. We derive a series of theoretical results using the delete-a-group jackknife variance estimator. Our theoretical analysis provided supporting justification for using the same replicate-weighting cells as the full sample cells when there are no replicates with zero respondent cells. The choice to develop replicate-based weighting cells is often governed by operational considerations. Under certain conditions often met by business surveys, there are statistical arguments for choosing this approach as well for this version of the replicate estimator. As an alternative to applying one of these two approaches uniformly, one might consider a hybrid approach: use the full-sample weighting cells in each replicate except for when the weighting cell in a replicate has zero respondents. In such cases, we could apply the replicate-based weighting cells approach.

Replicate variance estimates that employ the full sample weighting cells essentially condition on the full-sample collapsing pattern. If the collapsing criteria are **counts only** based, using the full sample cells in replication will not fully capture the variability component from the unit non-response. **Factor** based collapsing criteria are used to reduce the variance of the estimate, and the collapsing factor thresholds should be set that do not overly affect the bias. In our simulation, the replicate variances that use the full sample weighting cells and the **counts plus factor** collapsing criteria appear to be very close to their respective MSEs, thus capturing the variance reduction (from adding the **factor**).

Ultimately, it seems preferable for a one-stage sample survey to sidestep the entire question by using a different variance estimator that is supported by the survey design, such as modified half-sample replication or the stratified jackknife (where one unit is dropped in each replicate). With both of these methods, the replicates weighting cell respondent counts are the same (or approximately the same) as the full sample weighting cell respondent counts. Future research will focus on cell collapsing effects in multi-stage designs, using these other variance estimators.

Acknowledgements

The authors thank Rita Petroni, Reid Rottach, and Lynn Weidman for their careful review of earlier versions of this manuscript.

References

Cochran, W.G. (1977). Sampling Techniques (3rd Edition). New York: Wiley.

Little, R.J.A. and Rubin, D.R. (2002). Statistical Analysis with Missing Data. New York: John Wiley and Sons.

Little., R.J.A. and Vartivarian, S. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means? *Survey Methodology*, **31**, pp. 161-168.

Lohr, S.L. (1999). Sampling: Design and Analysis. California: Duxbury Press.

Oh, H.L. and Scheuren, F.J. (1983). Weighting Adjustment of Unit Nonresponse. Incomplete Data in Sample Surveys (Vol.20). New York: Academic Press, 143-184.

Kalton, G. and Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics*, **19**, pp. 81-97.

Kalton, G. and Maligalig, D.S. (1991). A Comparison of Methods of Weighting Adjustment for Nonresponse. *Proceedings of the 1991 Annual Research Conference*, U.S. Bureau of the Census, 409-428.

Kott, P. (1994). A Note on Handling Nonresponse in Sample Surveys. *Journal of the American Statistical Association*, **89**, pp. 693 –696.

Kott, P. (2001). The Delete-a-Group Jackknife. *Journal of Official Statistics*, **17**, pp. 521 –526.

Rao, J.N.K. and Shao, J. (1999). Modified Balanced Repeated Replication for Complex Survey Data. *Biometrika*, **86**, pp. 403-415.

Särndal, C., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer Verlag.

Wolter, K. (1985). Introduction to Variance Estimation. New York: Springer-Verlag.