Coverage Error in Telephone Surveys: Bias in Estimates of Intimate Partner Violence, Variances, Associations, and Total Error from Exclusion of the Cell Phone-Only Population

Andy Peytchev[1], Lisa R. Carley-Baxter[1], and Michele Lynberg Black[2]
[1] Survey Research Division, RTI International, Research Triangle Park, NC
[2] Division of Violence Prevention, Centers for Disease Control and Prevention, Atlanta, GA

Abstract

While landline telephone household surveys often draw inference about the U.S. population, a proportion with only cell phones is excluded. This proportion is substantial and increasing, providing potential for coverage bias. Improved understanding of the resulting coverage bias and the ability to adjust for it is needed. Studies have looked at bias in means and proportions, but undercoverage can affect other essential statistics. The precision of point estimates can be biased, leading to erroneous conclusions. In addition, research examining multivariate relationships will be further affected by bias in associations. Coverage bias is suspected as the cell-only population is different on demographic characteristics. These characteristics are commonly related to survey measures, creating conditions for bias. Yet bias in adjusted estimates, such as victimization rates, occurs only when differences on survey variables remain within demographic groups.

A national landline telephone survey was conducted, followed by a survey of adults with only cell phones. Differences between samples were found in estimates of not only means and proportions, but also variances and associations. Bias in some point estimates was reduced through poststratification, but became larger and in opposite direction for others; cell phone respondents were more likely to report victimization, but conditional on demographic characteristics, were less likely to report it. Different uses of survey data can be affected by omitting the cell-only population, while reliance on postsurvey adjustments can be misleading.

Key Words: Coverage bias, Coverage error, MSE, Landline surveys, Cell phone surveys

1. Introduction

Official government statistics and social science research are continually making use of telephone survey data collection. Key crime and victimization estimates are derived from national random digit dialed (RDD) telephone surveys (e.g., the Developmental Victimization Survey). A large and rapidly increasing proportion of adults in the U.S. do not have a landline telephone in the household, but have a cell phone. The National Health Interview Survey (NHIS), an area probability face to face survey, estimated the percentage to be almost 13% (Blumberg and Luke, 2007b) for the first half of 2007. Complexities and cost involved in the design and implementation of dual frame surveys that include both landline and cell phone samples prevent many studies from changing to such designs. Yet exclusion of adults with only cell phones from RDD studies can create coverage bias in survey estimates.

Coverage bias can be relatively large for some *estimates* but not for others, just as nonresponse bias has been shown to be variable-specific (e.g., Groves, 2006). Similarly, coverage bias may affect particular types of *statistics* more than others. For example, numerous studies demonstrate nonresponse bias in means and proportions, but studies that have looked at nonresponse bias in associations have found little to no evidence for it (e.g., Lepkowski and Couper, 2002). Nothing is known about coverage bias in associations. Furthermore, the effects of undercoverage can be completely different on variance estimates. If the exclusion of cell phone only adults does not necessarily bias point estimates, it may be that element variance is biased, still leading to erroneous precision and biased significance tests, and affecting total error in survey point estimates. Therefore, official government statistics and social research alike may be affected by biased confidence intervals, significance tests, and covariance structures that need evaluation.

Effort has been expended on tracking the demographic characteristics of those with only cell phones who are excluded from landline RDD surveys (e.g., Blumberg and Luke, 2007b; Tucker, Brick and Meekins, 2007). Coverage bias in point estimates due to exclusion of the cell phone-only population can be predicted based on identified demographic characteristics, and likely used by many to speculate about bias in estimates in a particular RDD survey. Yet this may be extremely misleading. Young adults may have higher prevalence rates for some behaviors, leading one to assert that a landline survey will *underestimate* these rates. Surveys commonly use postsurvey adjustments that include age. If young adults in the cell phone sample have lower prevalence rates than their counterparts in the landline population, the population-weighted landline survey will actually *overestimate* these rates. Research into coverage bias within subgroups is needed, as the observed bias in reported survey estimates is in weighted estimates.

An alternative to landline and cell phone dual frame survey designs is to create postsurvey adjustments to census demographic totals, and assert that most of the coverage bias in survey estimates is removed through the use of the poststratified weights. While this assumption has to be made for landline RDD telephone surveys in order to make inferences about the entire adult population in the U.S., it remains untestable in a single frame design. Even if adjustments reduce coverage bias in some estimates, they may not reduce bias in other estimates in the same survey. Keeter and colleagues (2007) find that postsurvey adjustments removed coverage bias in point estimates of political measures, albeit confounded with nonresponse bias (discussed below).

Finally, it is very difficult to separate coverage bias from other sources of survey error, such as unit nonresponse. Even when multiple frames can be used to reduce coverage bias, coverage bias can be reintroduced when samples from each frame are differentially affected by nonresponse, and nonresponse is not addressed separately. Due to lack of information available separately for those included and those excluded in a landline survey, nonresponse adjustments could not be made before coverage adjustments, and even studies focusing on the problem are unable to address this confounding of errors (e.g., Brick et al., 2006). Such nonresponse adjustments require the assumption that respondents from each frame are interchangeable, the implausibility of which is the very motivation for the multi frame survey design.

2. Coverage Bias

Coverage bias in a population mean is the difference between the mean for the target population that is included in the sampling frame and the mean for the entire target population. It can be decomposed as the product of the proportion excluded, and the difference between the target population included in the sampling frame and those excluded, and estimated from a sample:

$$\bar{y}_C - \bar{y} = \frac{U}{N}(\bar{y}_C - \bar{y}_U) \qquad [1]$$

Where $\bar{y}$ is the sample mean for the eligible population, $\bar{y}_C$ is the sample mean for the eligible in the frame, $\bar{y}_U$ is the estimated mean for the eligible excluded from the frame, *N* is the number eligible in the population, and *U* is the number eligible in the population not included in the frame (undercovered).

However, in practice, the population mean is not estimated by $\bar{y}_C$ or $\bar{y}$, but by a weighted estimate that may include adjustments for selection probability, nonresponse, and coverage (often through poststratification). Of practical importance is the bias in weighted estimates, as the actual bias that would be in reported survey estimates:

$$\bar{y}_{C,ps} - \bar{y}_{ps} \qquad [2]$$

Where *ps* denotes (poststratification) weighting to the adult U.S. population.

Inference about parameter estimates for means and proportions, and comparisons between them, can be affected by undercoverage even in the absence of bias in means and associations. If the eligible population that is not in the frame is more variable on a survey measure, the estimate of the variance of a mean based on a sample will be understated. Such an understatement can be further exacerbated when there is also a bias in the estimate of the mean. Whether testing differences in means or differences in associations, significance tests will be affected by biased variance estimates. Apart from the influence of the proportion undercovered (*U/N*), bias in the standard deviation of the mean arises from two sources, the difference in the variability within each group (eligible population included and not included in the frame), and the difference in the means of each group. The bias in the standard deviation can be estimated by:

$$s_{y_C} - s_y \qquad [3]$$

While the difference between the two standard deviations is simply:

$$s_{y_C} - s_{y_U} \qquad [4]$$

Despite predominant attention, coverage bias is not limited to descriptive statistics, such as means and proportions. A common use of survey data is the estimation of bivariate and multivariate relationships between survey measures. The means of two variables may be subjected to coverage bias, but the association between them may remain unbiased—and vice versus. The relationship between the two measures can be the same for the eligible population regardless of inclusion in the sampling frame. This would not be surprising—the factors related to bias in means may not be the same as the factors affecting bias in associations. If bias affects point estimates but not associations for a set of variables used in a regression model, bias would not necessarily affect regression coefficients apart from the intercept. The coverage bias in a covariance σ_{y_1,y_2} between two variables, y_1 and y_2, can be estimated by:

$$s_{y_{1C},y_{2C}} - s_{y_1,y_2} \qquad [5]$$

Apart from the influence of the proportion undercovered (*U/N*), bias in the standard deviation of the mean arises from two sources, the difference in the variability within each group (eligible population included and not included in the frame), and the difference in the means of each group. Of interest to many users that employ survey data in multivariate models or compare associations across groups, would be the difference in covariances between the two groups:

$$s_{y_{1C},y_{2C}} - s_{y_{1U},y_{2U}} \qquad [6]$$

Postsurvey adjustments are often used to reduce bias; poststratification to population characteristics obtained from the census is assumed to address undercoverage. Such adjustments can affect statistics and their sampling variances, each influencing inference from the survey to population parameters. To evaluate the total effect of coverage bias on survey estimates, the concept of total error can be useful (Deming and Birge, 1934). Total error in a statistic can be estimated by the Mean Square Error (MSE), the sum of the squared bias and the variance. For the mean in the presence of undercoverage, it can be estimated as:

$$MSE_C = (\bar{y}_{C,ps} - \bar{y}_{ps})^2 + \frac{s^2_{y_{ps}}}{n_C} \qquad [7]$$

In the near absence of undercoverage, the expected value of the squared bias term $(\bar{y}_C - \bar{y})^2$ approaches zero and MSE can be estimated by the sampling variance, s^2_y / n_C.

3. Challenges in Addressing Coverage Bias

Undercoverage of the U.S. adult population in (landline) RDD surveys, i.e., the *U/N* proportion in the formulas above, was 12.6% in the first half of 2007 and has been increasing by almost 3-percentage points per year since 2004, based on estimates from NHIS (Blumberg and Luke, 2007b). This provides substantial potential for coverage bias in survey estimates based on landline sampling frames. There are three approaches to addressing coverage bias: by changing the definition for the target population, through adjustment, and through the use of multiple sampling frames. The target population can be redefined, removing the mismatch between the sampling frame and the target population. In most RDD telephone studies, this is not feasible–inference needs to be made about the U.S. population, not about the U.S. adult population with a landline phone in the household, particularly with the relatively large proportion of adults without a landline.

Alternatively, estimates could be adjusted for known population demographic distribution from an external source, such as the census, that is not subjected to the same undercoverage. This approach depends on assumptions, and conditions leading to its effectiveness are discussed below.

Finally, a dual (or multi) frame sample design could include eligible population that is not covered by a single frame. An obvious additional frame in telephone surveys are cell phone numbers. The large increase in undercoverage in landline telephone surveys is entirely parallel to an increase in the proportion of adults with only cell phones; the proportion without any type of telephone has remained constant at less than 2% since 2004 (Blumberg and Luke, 2007b). Apart from cost challenges in conducting interviews with a sample of cell phone numbers, there are statistical problems in combining the samples from the two frames, hindered by problems in identification of selection and response probabilities.

3.1. Adjustment for Coverage Bias

The effectiveness of poststratification weights for coverage error adjustment relies on the association of adjustment variables with (1) whether a population element is included in the frame, and (2) the survey measures of interest. The intention, but also an assumption, is that conditional on the vector of covariates used in poststratification, those included and those not included in the sampling frame have the same expected values on survey measures. This assumption may not hold and can only be tested through multiple frame samples.

Demographic variables commonly used in poststratification are associated with whether an adult has a landline telephone in the household (e.g., Blumberg and Luke, 2007b; Tucker et al., 2007)–age, gender, race/ethnicity, income, and education. Demographic characteristics may also be associated with survey measures, for example in a survey on sexual violence, as the association between demographic characteristics and violence outcomes has been shown in previous research (e.g., Tjaden and Thoennes, 2000; Tjaden and Thoennes, 2006). Both sets of associations are necessary but not sufficient for effective adjustment for coverage bias. The degree to which coverage bias in survey estimates is adjusted using the census demographic variables needs to be directly evaluated through obtaining survey measures from members of the eligible population that is not included in the frame. Furthermore, in a single frame design associations between census variables and survey measures can be calculated only for population elements that are in the frame – these relationships may be different for those not in the frame. When the relationships are different, coverage adjustments may not be effective and can even further bias estimates, just as adjustments for nonresponse can increase nonresponse bias when the relationship between the adjustment variable and the survey variable is different among respondents and nonrespondents (e.g., Lin and Schaeffer, 1995). This assumption can also be evaluated with data from another frame.

3.2. Reduction in Coverage Bias

The use of additional frames can reduce coverage bias by achieving nonzero selection probability to a larger part of the eligible population, but can introduce vast complexities into the survey design. One complexity is the estimation of selection probability for sample members in each frame, including those who can be selected from either frame. While an estimator has been developed by Hartley (1962), the estimation of parameters in the model is problematic; for example, the probability that an individual with a landline is interviewed on a cell phone is unknown. This probability can be estimated with a Politz-Simmons estimator (Politz and Simmons, 1949) based on the respondent-reported proportion of calls received on each device, or assumed to be .5 as done in some studies (Brick et al., 2006; Kennedy, 2007). Alternatively, screening can be done so that sample members identified to be in both frames are interviewed only if they have been selected in one of the frames. Indeed, this is the only available option if membership in both frames can not be obtained for sample members in one of the frames, e.g., not knowing cell phone ownership for landline respondents.

Arguably, a more significant problem is the interaction of the implementation of a dual frame design for reducing coverage error with other sources of survey error. Sample members may have different expected values on a survey measure depending on which frame they have been selected through. Despite evidence in the literature of differences in reporting due to mode of survey administration, the extent of measurement differences in the two telephone methods of data collection remains to be evaluated; such differences may be small and more circumstantial, as in both modes surveys are administered by an interviewer over a phone. However, samples in both frames are subjected to nonresponse. How nonresponse adjustments are created has been shown to affect dual frame survey estimates, at least for demographic variables (Brick et al., 2006), and homogeneity of landline and cell phone populations within demographic groups is assumed.

Indeed, combining samples from multiple frames in the presence of nonresponse implicitly makes the assumption that respondents in each frame have the same expected values. For example, males typically respond at a lower rate in the landline sample (e.g., Groves and Kahn, 1979; Keeter et al., 2000), but at a higher rate in the cell phone sample (e.g., Kennedy, 2007; Link et al., 2007). Merging the samples using selection probability and then creating poststratification adjustment to the population totals, including gender, will allow the correct proportion of males in the population to be reflected in weighted survey estimates. However, an assumption is made that males in the two samples are interchangeable, over-representing males with cell phones due to survey nonresponse. If males in each sample have different expected values on survey estimates, bias will arise. While the census does not provide demographic characteristics for the landline and cell phone frames separately, a method is needed to adjust each sample to demographic totals prior to combining the samples.

From the multiple uses of survey data and possible bias in different types of statistics, difficulty in creating adjustments for coverage bias, and assumptions made in the use of the adjustments, several essential questions need to be addressed to help understand and attend to coverage bias in RDD telephone surveys:

1) Does the exclusion of the adult cell phone only population from a RDD telephone study induce coverage bias in means and proportions of substantive variables?
2) Is the element variance different for respondents in the cell phone frame?
3) Are associations susceptible to coverage bias from the exclusion of adults with only cell phones?
4) Can poststratification adjustments effectively reduce coverage bias in survey estimates?
5) Does a dual frame sample design reduce the mean squared error of survey estimates, compared to only a landline frame sample design?

In addition, methodology is needed for creating nonresponse adjustments prior to combing the frames, to avoid interference of nonresponse error with the reduction of coverage error.

4. Data and Methods

The questions outlined above are addressed with data from the National Intimate Partner and Sexual Violence Survey (NISVS) Pilot Study, an RDD telephone survey that includes measures on stalking, sexual violence, physical and psychological aggression (see the Appendix). To improve the efficiency of the study, the sample design included a sample of RDD landline numbers and a sample of listed landline numbers; selection probabilities were accounted for. These landline frames are labeled as the *landline frame* and the sampled numbers from both are referred to as the *landline study* throughout the rest of this paper. The instrument was 30 minutes long on average, and respondents were randomly assigned to receive a promised incentive of either $10 or $20. The landline study was fielded January to April 2007, collecting completed or partial interviews from 5,296 adults, with an AAPOR Response Rate 4 (AAPOR, 2006) of 28.5%. Analyses of embedded experiments show that the response rate was affected by the announced topic of the survey (Lynberg Black, Carley-Baxter and Twiddy, 2007) and the incentive amount (Carley-Baxter, Lynberg Black and Twiddy, 2007), in addition to the selection of males within households at a higher rate, due to their lower response rate.

In addition, a RDD sample of 6,254 cell phone numbers was selected by Survey Sampling International, stratified by region using the same allocation as in the landline study. Cell phone sample members with landlines were screened out, as cell phone service among landline respondents was unknown, needed for estimation of selection probabilities. This resulted in a dual-frame design where the cell phone frame is used to augment only the omitted part of the population, allowing a comparison of practical importance, but the reader is referred to other work on the effect of including cell phone respondents who also have landlines (Kennedy, 2007).

The instrument in the cell phone study was reduced to 14 minutes, retaining questions by modules to avoid differences in question comprehension. Questions in the screener were added to ensure safety (i.e., whether the respondent is driving) and to inform selection probability for cell phones (e.g., number of cell phones instead of number of landlines and number of adults in the household). Additional questions were added to learn about usage patterns, sharing, and number of cell phones. All sample members were offered a promised incentive of $25. The study was fielded in September and October 2007. The study was stopped a week into data collection and resumed three days later, to address difficulties in obtaining interviews. Changes included slightly decreasing the length of the introduction and leaving voice mail messages. Interviewer debriefings were used to identify common reasons for refusals by sample members and interviewer refusal aversion and conversion trainings were conducted. Additional changes were made to call scheduling throughout the study, based on concurrent analysis of call history data, although the effectiveness of these changes was limited by a relatively low level of interviewing hours available. One-hundred and thirty-two interviews were conducted, with a 32.9% AAPOR Response Rate 4. Seven calls on average were made to numbers finalized as noncontacts, with less than a percent called more than ten times; this level of effort may be sufficient to stabilize survey estimates (i.e., that they do not change if additional call attempts are made), but the response rate could have been improved if time and resources had permitted. We believe that additional calls under the same survey protocol would have likely led to similar survey estimates, as restricting studies and analysis to as few as five call attempts has not been found to bias estimates (Curtin, Presser and Singer, 2000; Keeter et al., 2000).

4.1. Weighting in the Dual Frame Design

After selection probabilities within each frame are calculated based on number of phones and the number of adults in the household (landline), one method of combining data from landline and cell phone studies is to obtain the proportion of the target population that falls in each frame and adjust the selection weights to match these proportions. If this is done prior to poststratification, this last adjustment can distort the proportion with only cell phones in the weighted sample. An assumption is made implicitly that nonrespondents in the landline surveys are the same in terms of demographic characteristics, and bias will be induced if the landline and cell phone populations are different on survey variables, conditional on demographic characteristics. It is the uncertainty in this assumption that motivates the need for sampling cell phone numbers, yet this adjustment method would assume it holds true.

Ideally, the demographic characteristics of the landline and cell-only populations would be available from updated census data. This would allow the respondents in the cell phone sample and the landline sample to be adjusted to the respective populations they represent. Currently, such census data are not available. However, the National Health Interview Survey (NHIS) is an area probability survey with relatively high response rates (household level response rates around 90%) that includes questions about landline and cell phone service. The NHIS has released unweighted percentages for demographic characteristics for July-December 2006, for adults with only cell phones (Blumberg and Luke, 2007c). However, NHIS reports do not provide the demographic composition of adults with landlines, but rather present the proportion with only cell phones within each demographic group. Estimates for demographic totals for adults with landlines and adults with only cell phones were obtained using census adult population estimates for 2006 for each demographic category and the proportion with only cell phones in that category (e.g., Non-Hispanic White) from NHIS.

This approach yields population totals by age, sex, race/ethnicity, and region, separately for those with only cell phones and for the rest of the population. These counts are not cross-classified, i.e., the number of adults with only cell phones who are 65 or older, male, Hispanic, and reside in the Northeast region, is unknown. The selection weights are therefore raked to the univariate totals (marginal distributions) in each sample, an approach that iteratively allocates the sample cases in the cross-classifications while preserving the known marginal distributions using a SAS macro program (Izrael, Hoaglin and Battaglia, 2004). This results in landline and cell phone samples that each represent their own population, and have been weighted to match their known (albeit estimated through different sources) distributions on age, sex, race/ethnicity, and region.

Some cases had missing values for race and age, which would not have allowed the calculation of weights for these respondents. Sequential regression imputation was used to impute values in IVEware, a SAS macro program for multiple imputation and variance estimation (Raghunathan, Lepkowski, Van Hoewyk and Solenberger, 2001). In the landline sample, 17 values for race/ethnicity and 34 values for age were imputed; in the cell phone sample 2 values for race/ethnicity and 1 for age were imputed.

The combined dual frame study did not allow for overlap of the sampling frames, i.e., adults contacted on cell phones in the landline study and adults with landlines in the cell phone study were excluded. Since the previous adjustment step weighted the landline and cell-only respondents to their respective sizes in the U.S. adult population, the necessary adjustment to combine the samples was incorporated in the raking adjustment. As the samples were also adjusted to additional variables, the combined samples approximate the population proportion of those with only cell phones, as well as population proportions of groups such as males or adults 18-24 old with only cell phones.

The poststratification for the landline and for the cell phone respondents employed population estimates based on a combination of census and a face-to-face survey (NHIS), and only univariate demographic totals for each population could be constructed. After the landline and cell phone samples were combined, they were poststratified to the adult U.S. population based on census estimates for 2006. Raking was not necessary, as the cross-classifications of age, sex, race/ethnicity, and region were available and employed through an iterative fitting process (Folsom and Singh, 2000). However, extreme weights were trimmed for each sex (separately, as they were sampled at different rates) to reduce mean squared error of estimates, and weights were re-poststratified to only two-way cross classifications of the demographic variables.

4.2. Weighting in the Single Frame Design

Poststratification weights were created also for only the landline study using the same census demographic totals under the assumption that there is no difference between landline and cell phone respondents conditional on these variables. Raking was not necessary here, but the same procedure was used in implementing the full cross-classification, weight trimming, and re-poststratification.

4.3. Analytic Approach

Using selection weights, demographic characteristics are first compared for the landline and cell phone respondents. Apart from informing potential for coverage bias, differences in these demographic characteristics show any potential in using them in poststratification to adjust for coverage bias, under the assumptions discussed earlier. Next, differences in key survey estimates are presented, for four types of statistics: proportions, means, element variances, and associations. We then evaluated whether any differences found in means and proportions are effectively adjusted through poststratification. Finally, the effect of implementing a single frame design versus a dual frame design on the mean squared error of survey estimates is computed. Variance estimation was performed with Taylor Series approximation in SUDAAN.

5. Results

Consistent with other studies (e.g., Blumberg and Luke, 2007b; Tucker et al., 2007), there were significant differences between landline and cell phone respondents by age, sex, census region, and race/ethnicity. Cell phone respondents were younger, more likely to be from the Midwest and South census regions, much more likely to be of Hispanic origin, and more likely to be Black. For example, 91% of landline respondents were 30 years of age or older, while only 52% among the cell phone respondents were 30 years of age or older. Since all these characteristics are used in poststratification, this provides the potential that poststratified estimates from the landline survey alone can be unbiased relative to dual frame landline and cell phone estimates. Another condition for the effectiveness of such adjustments is an association between the demographic characteristics and the key survey variables – the four proportions and four scale measures for stalking, sexual violence, physical aggression, and psychological aggression. Those product-moment correlations were relatively low and similar for males and females, ranging from less than 0.001 to 0.156, with significant associations with age and race (results not shown). An expected exception was the association between sexual victimization and being female, which was 0.259.

Table 1 presents survey estimates for landline and cell phone respondents, weighted for selection probability due to the differences in within household selection and the effect of number of phones in order to infer about each of the two populations. As expected, means and proportions tend to be higher for the cell phone sample, but only one of the differences in point estimates is significant at the .05 level and two at the .10 level, with large standard errors in the cell phone sample.

Table 1. Key Survey Estimates for Landline and Cell Phone-Only Respondents Weighted for Selection Probability and Unweighted Estimates of Standard Deviations.

	Landline[a]			Cell Phone Only[b]			Difference (Cell-Landline)			Landline & Cell		(Naïve) ‡ Percent Rel. Bias
Variable (scale)	Mean	StdErr	StdDev	Mean	StdErr	StdDev	Mean	StdErr	StdDev †	Mean	StdDev	of the Mean
	$\bar{y}_C$	$s.e.(\bar{y}_C)$	s_{y_C}	$\bar{y}_U$	$s.e.(\bar{y}_U)$	s_{y_U}	$\bar{y}_U - \bar{y}_C$	$s.e.(\bar{y}_U - \bar{y}_C)$	$s_{y_U} - s_{y_C}$	$\bar{y}$	s_y	$\frac{\frac{U}{N}(\bar{y}_C - \bar{y}_U)}{\bar{y}}$
Females												
Stalking (1=yes)	38.8%	(1.2%)	48.8%	36.2%	(7.0%)	48.8%	-2.6%	7.1%	0.1%	38.6%	48.8%	0.8%
Sexual Violence (1=yes)	22.1%	(1.0%)	43.1%	26.9%	(6.4%)	45.1%	4.8%	6.5%	1.9%	22.5%	43.2%	-2.5%
Physical Aggr. (1=yes)	38.7%	(1.2%)	48.9%	43.9%	(7.2%)	50.4%	5.2%	7.3%	1.5%	39.1%	48.9%	-1.6%
Psych. Aggr. (1=yes)	54.7%	(1.2%)	49.7%	67.5%	(6.9%)	47.1%	12.8%*	7.0%	-2.6%	55.8%	49.7%	-2.7%
Stalking (0-10)	1.21	(0.05)	2.07	1.00	(0.26)	2.32	-0.21	0.26	0.25	1.20	2.07	2.1%
Sexual Violence (0-2)	0.31	(0.02)	0.67	0.34	(0.09)	0.63	0.03	0.09	-0.04	0.32	0.67	-1.1%
Physical Aggr. (0-13)	1.44	(0.06)	2.53	1.37	(0.31)	2.77	-0.07	0.32	0.24	1.43	2.53	0.6%
Psych. Aggr. (0-12)	2.95	(0.08)	3.07	3.63	(0.43)	3.17	0.69	0.44	0.11	3.00	3.07	-2.7%
Males												
Stalking (1=yes)	29.0%	(1.2%)	45.4%	36.8%	(5.5%)	48.2%	7.8%	5.6%	2.9%	30.4%	45.5%	-3.0%
Sexual Violence (1=yes)	5.2%	(0.5%)	23.3%	11.6%	(3.7%)	31.6%	6.4%*	3.7%	8.4%***	6.3%	23.6%	-12.0%
Physical Aggr. (1=yes)	44.2%	(1.3%)	49.7%	50.9%	(5.7%)	50.3%	6.7%	5.8%	0.6%	45.4%	49.7%	-1.7%
Psych. Aggr. (1=yes)	52.5%	(1.3%)	49.9%	55.8%	(5.9%)	50.1%	3.3%	6.0%	0.2%	53.0%	49.9%	-0.7%
Stalking (0-10)	0.69	(0.04)	1.49	1.03	(0.21)	1.94	0.33	0.21	0.44***	0.75	1.51	-5.2%

Sexual Violence (0-2)	0.07	(0.01)	0.30	0.14	(0.05)	0.41	0.07	0.05	0.11***	0.08	0.31	-10.9%
Physical Aggr. (0-13)	1.61	(0.06)	2.47	1.88	(0.30)	2.84	0.27	0.30	0.38*	1.65	2.48	-1.9%
Psych. Aggr. (0-12)	2.42	(0.07)	2.44	3.23	(0.36)	3.17	0.81**	0.37	0.73***	2.56	2.47	-3.7%

* p<.10, ** p<.05, *** p<.01; † Levene's test for homogeneity of variance; ‡ Weighted only for selection and size of sampling frame; [a] n=2814 for females, n=2482 for males; [b] n=51 for females, n=81 for males.

Differences between landline and cell phone respondents could also be exhibited in how variable the two populations are in terms of the survey measures. Such differences would bias the population variance estimates, bias statistical tests between groups, and may in fact be of interest to researchers – why a subset of the population is more variable on factors that have negative consequences on their lives. Indeed, the unweighted element variance, presented as standard deviations in Table 1, was higher for many of the measures, although statistically significant only for males.

Albeit by very small differences for 6 of the 16 estimates, standard deviations for the dual frame estimates were higher than the single frame estimate (i.e., landline) that is subjected to undercoverage of the cell phone only population (second and last data columns in Table 1). Much of this was driven by the higher means and proportions among the cell phone sample, yet the consequences remain the same and future research into differences in variances is needed.

Some research questions may focus on bivariate associations, regressions, or causal modeling (e.g., covariance modeling). For such analyses bias that could arise from the exclusion of the cell phone-only population is in associations. The variance-covariance matrices for the stalking, physical and psychological aggression by sample and gender are presented in Table 2. Most of the variances and covariances are higher in the cell phone sample, and significantly different among males (Box's M test (Box, 1949), p<.001). Covariances were estimated based on log transformation of the three variables with sufficient number of scale points in order to remedy extreme values. However, these variables are left censored as respondents could not have a score of zero, yet they could still vary in how likely they are to be subjected to violence. Tobit regression models were fit, in which each variable was regressed on the other, on whether the respondent was from the cell phone sample, and their interaction. A significant interaction signifies a different association between the two survey variables, across the landline and cell phone populations. The interactions were significant in all models, for both females and males (results not shown).

Table 2. Associations between (Log-Transformed) Survey Variables for Landline and Cell Phone-Only Respondents.

	Landline[a] ($s_{y_{1C},y_{2C}}$)			Cell Phone Only[b] ($s_{y_{1U},y_{2U}}$)			Diff. †
Variable (original scale)	Stalking	Phys. Aggr.	Psych. Aggr.	Stalking	Phys. Aggr.	Psych. Aggr.	Sig.
Females							
Stalking (0-10)	0.49			0.51			
Physical Aggression (0-13)	0.29	0.59		0.21	0.62		
Psych. Aggression (0-12)	0.29	0.43	0.72	0.21	0.45	0.65	n.s.
Males							
Stalking (0-10)	0.31			0.44			
Physical Aggression (0-13)	0.17	0.59		0.37	0.67		
Psych. Aggression (0-12)	0.14	0.36	0.57	0.33	0.49	0.75	***

*** p<.01; † Box's M test for equality of variance-covariance matrices; [a] n=2814 for females, n=2482 for males; [b] n=51 for females, n=81 for males.

There were small differences between the landline and the combined landline & cell phone estimates in Table 1, but we also saw that demographic characteristics used in poststratification were strongly associated with frame membership and weakly associated with survey measures, providing the potential for effective adjustments. Table 3 presents the poststratified estimates under the single and the dual frame designs.

Table 3. Estimates Based on Landline and on Landline and Cell Phone Respondents, Poststratified to Demographic Characteristics of the U.S. Adult Population.

Variable (scale)	Landline[a] Mean $\bar{y}_{C,ps}$	Landline[a] Std.Err. $s.e.(\bar{y}_{C,ps})$	Landline & Cell Phone[b] Mean $\bar{y}_{ps}$	Landline & Cell Phone[b] Std.Err. $s.e.(\bar{y}_{ps})$	(Percent) Relative Bias $\frac{\bar{y}_{C,ps}-\bar{y}_{ps}}{\bar{y}_{ps}}$
Females					
Stalking (1=yes)	40.0%	(1.4%)	38.9%	(1.5%)	2.8%
Sexual Violence (1=yes)	22.1%	(1.1%)	22.0%	(1.2%)	0.5%
Physical Aggression (1=yes)	38.7%	(1.3%)	38.6%	(1.5%)	0.3%
Psychological Aggression (1=yes)	55.8%	(1.4%)	56.0%	(1.5%)	-0.4%
Stalking (0-10)	1.28	(0.06)	1.24	(0.07)	3.3%
Sexual Violence (0-2)	0.31	(0.02)	0.31	(0.02)	1.3%
Physical Aggression (0-13)	1.45	(0.07)	1.42	(0.08)	2.3%
Psychological Aggression (0-12)	2.99	(0.09)	3.01	(0.10)	-0.6%
Males					
Stalking (1=yes)	29.6%	(1.2%)	29.9%	(1.4%)	-1.0%

Sexual Violence (1=yes)	5.8%	(0.6%)	5.6%	(0.6%)	3.6%
Physical Aggression (1=yes)	45.2%	(1.3%)	44.5%	(1.5%)	1.6%
Psychological Aggression (1=yes)	53.8%	(1.4%)	52.3%	(1.5%)	2.9%
Stalking (0-10)	0.74	(0.05)	0.73	(0.05)	1.8%
Sexual Violence (0-2)	0.08	(0.01)	0.07	(0.01)	5.6%
Physical Aggression (0-13)	1.69	(0.07)	1.62	(0.08)	4.5%
Psychological Aggression (0-12)	2.50	(0.07)	2.44	(0.08)	2.6%

[a] n=2814 for females, n=2482 for males; [b] n=51 for females, n=81 for males.

For some measures, poststratification reduced even the small bias due to undercoverage. Among females, for example, the selection weighted percent relative bias in the landline estimates for being sexually victimized was -2.5% in Table 1, which changed to 0.5% after poststratification, in Table 3. Among males, this difference was reduced from -12.0% to 3.6%. However, for other measures poststratification changed the bias in estimates. Among males, for example, the selection weighted percent relative bias in the landline estimates for being victimized by physical aggression was -1.7% (Table 1), which became 1.6% (Table 3) after poststratification; for psychological aggression it changed from -0.7% to 2.9%. This means that not all statistics have the same expected value in the two frames, even conditional on demographic characteristics. While male respondents in the landline survey were less likely to have been psychologically abused, they were in fact more likely to have been psychologically abused when controlling for demographic characteristics known to differentiate the landline and cell phone-only populations.

Inference from sample surveys depends on both bias and variance of survey estimates. The combined effect of bias and variance was evaluated through MSE. We find almost no difference between the single and dual frame designs, with lower MSE for the three largest differences in the dual frame design as this design does not have a coverage bias term, and the lower sampling variance overwhelms the additional variability in the weights (data not presented) and higher element variance (from Table 1) in the variance component of MSE.

6. Discussion and Conclusions

Bias in estimates of intimate partner violence due to the substantial undercoverage in RDD samples of landline numbers is not limited to means and proportions; bias in variance estimates can lead to erroneous inferences about these point estimates, while bias in associations will affect multivariate analyses. A relatively low-cost solution is postsurvey adjustment to population demographic characteristics, yet this approach will not be effective if those without landlines are different on survey measures within their demographic groups. A more costly approach is to administer the interview on a sample of cell phone numbers, which poses additional problems, such as the confounding of different sources of survey error. We attended to these issues in a survey on intimate partner violence and offer two main conclusions from this study and a dual frame estimation procedure:

1) Coverage bias in landline telephone surveys affects not only means and proportions of survey measures, but also:
   a. Element variances, and
   b. Associations between survey variables

Excluding the cell phone-only population from telephone surveys can affect not just descriptive statistics, but also their estimates of variance, thus affecting significance tests. In this study, variances would be slightly underestimated under a single frame design. Analyses of multivariate relationships between survey measures would also be biased. While researchers pursuing multivariate modeling of survey data may be more comfortable in asserting lack of bias in associations due to nonresponse based on extant literature, this seems far less tenable for bias due to undercoverage in landline telephone surveys.

2) Demographic characteristics associated with both undercoverage (no landline phone) and the survey measures can effectively adjust for coverage bias in some survey estimates, but also exacerbate bias in other estimates in unpredictable ways.

We know from face to face surveys that there are sharp differences in terms of demographic characteristics between those with and without a landline in the household. When making inference about the U.S. population from a landline RDD survey, an assumption is made that postsurvey adjustments reduce or eliminate coverage bias. To the degree that the information used in these adjustments is associated with undercoverage and with survey measures, coverage bias could be adjusted in survey estimates. In an adjustment cell approach, an additional assumption is that those included and those excluded from the sampling frame are homogenous on survey measures. In a propensity model, this assumption translates to having the same expected values conditional on the auxiliary information. This assumption held true for some survey measures but not for others, similar to findings from in-person interviewing of both landline and cell phone populations (Blumberg and Luke, 2007a). Notably, postsurvey adjustments changed the direction (analogous to Simpson's paradox) and even increased coverage bias in some survey estimates, while decreasing it for others.

In sum, coverage bias, as with other sources of survey error, is estimate and statistic specific. Survey practitioners often rely on postsurvey adjustments, but could not know when these adjustments are effective. In any given survey, adjustments effective in reducing bias in estimates of stalking may not be effective for estimates of sexual violence; when effective for females they may not be so for males; and even when effective for estimates of means, they may fail for other statistics. With a growing cell phone-only population and shrinking landline coverage, studies need to incorporate multiple frames without naïve reliance on postsurvey adjustments. Inference from probability-based surveys relies on the ability to select sample members from the entire target population. Failure to include the cell phone-only population assigns a zero probability of inclusion to a nonrandom subset of the adult population, for which models can not *always* adjust for in survey estimates.

In addition, a method is offered to combine multiple frames to minimize the interference of nonresponse in providing dual frame estimates. Creating intermediate poststratification weights prior to combining frames avoids making the unrealistic assumption that eligible population in one frame is the same as the population in the other frame, conditional on demographic characteristics; the very impetus for using a multi frame design. However, we had to use two external sources and raking to achieve this. Hopefully, in the future, large national surveys would provide demographic cross-classifications separately for different survey frames; landline and cell phone in particular.

This study had a small number of respondents in the cell phone sample, limiting the detection of bias. This limitation demonstrated an important aspect of sample design that can be overlooked—while the addition of the cell phone frame decreases coverage bias, interviewing a small number of cell phone adults relative to the landline sample can increase total error through the increase in variance due to weighting. In this case, total error was generally reduced for males due to the decrease in bias, but it was increased for females where total error was dominated by variance.

Results may be affected by nonresponse bias, particularly in the cell phone study due to the low response rate. Measurement differences between landline and cell phone mediums may also affect responses. Both nonresponse and measurement error, such as induced by the shorter field period and the shorter instrument in the cell phone interviewing, are confounded in this investigation of coverage error in landline surveys and require targeted experiments. While many differences in interviewing sample members on landlines and cell phones are necessary, such as asking the respondents whether they are driving while talking on the cell phone, they can temper the degree to which dissimilarities between landline and cell phone samples are attributable to true differences. However, as in other investigations, unless the unique features in the survey protocol have been found to substantially affect estimates through inducing nonresponse or measurement error, conclusions are tempered but not refuted.

Further research is needed to replicate these results, just as bias is survey and statistic specific. Other types of measures, such as attitudinal questions, and other topics, such as questions on politics, would bring further insight into the properties of coverage bias in telephone surveys and the ability to adjust for it. Research is also needed into understanding the causes of bias in element variances and associations. Bias in these statistics can be due to true differences, but also due to differences in measurement error and correlated measurement error. Such biases can affect inference from survey data for both descriptive statistics and multivariate analysis, but has received very little attention.

Acknowledgments

The findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention/the Agency for Toxic Substances and Disease Registry.

References

AAPOR (2006). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. Lenexa, Kansas, AAPOR. **4th edition**.

Blumberg, S. J. and J. V. Luke (2007a). "Coverage Bias in Traditional Telephone Surveys of Low-Income and Young Adults." Public Opinion Quarterly **71**(5): 734-749.

---. (2007b). "Wireless Substitution: Early Release of Estimates Based on Data from the National Health Interview Survey, January-June 2007." Retrieved January 10, 2008, from http://www.cdc.gov/nchs/nhis.htm.

---. (2007c). "Wireless Substitution: Early Release of Estimates Based on Data from the National Health Interview Survey, July – December 2006." Retrieved December 10, 2007, from http://www.cdc.gov/nchs/nhis.htm.

Box, G. E. P. (1949). "A General Distribution Theory for a Class of Likelihood Criteria." Biometrika **36**: 317-346.

Brick, J. M., S. Dipko, S. Presser, C. Tucker and Y. Yuan (2006). "Nonresponse Bias in a Dual Fame Sample of Cell and Landline Number." Public Opinion Quarterly **70**(5): 780-793.

Carley-Baxter, L., M. Lynberg Black and S. Twiddy. 2007. The Impact of Incentives on Survey Participation and Reports of Intimate Partner and Sexual Violence. Paper presented at the American Association for Public Opinion Research annual conference, Anaheim, CA.

Curtin, R., S. Presser and E. Singer (2000). "The Effects of Response Rate Changes on the Index of Consumer Sentiment." Public Opinion Quarterly **64**(4): 413-428.

Deming, W. E. and R. T. Birge (1934). "On the Statistical Theory of Errors." Review of Modern Physics **6**: 119-161.

Folsom, R. E. and A. C. Singh (2000). The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Poststratification. Proceedings of the Joint Statistical Meetings of the American Statistical Association.

Groves, R. M. (2006). "Nonresponse Rates and Nonresponse Bias in Household Surveys." Public Opinion Quarterly **70**(5): 646-675.

Groves, R. M. and R. L. Kahn (1979). Surveys by Telephone. New York, Academic Press.

Hartley, H. O. (1962). Multiple Frame Surveys. Proceedings of the American Statistical Association, Section on Social Statistics.

Izrael, D., D. C. Hoaglin and M. P. Battaglia (2004). To Rake or Not to Rake Is Not the Question Anymore with the Enhanced Raking Macro. Proceedings of the Twenty-Ninth Annual SAS Users Group International Conference.

Keeter, S., C. Kennedy, A. Clark, T. Tompson and M. Mokrzycki (2007). "What's Missing from National Landline RDD Surveys? The Impact of the Growing Cell-Only Population." Public Opinion Quarterly **71**(5): 772-792.

Keeter, S., C. Miller, A. Kohut, R. M. Groves and S. Presser (2000). "Consequences of Reducing Nonresponse in a National Telephone Survey." Public Opinion Quarterly **64**: 125-148.

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

Lepkowski, J. M. and M. P. Couper (2002). Nonresponse in the Second Wave of Longitudinal Household Surveys. Survey Nonresponse. R. Groves, D. Dillman, J. Eltinge and R. J. A. Little. New York, Wiley**:** 259-272.

Lin, I.-F. and N. C. Schaeffer (1995). "Using Survey Participants to Estimate the Impact of Nonparticipation." Public Opinion Quarterly **59**: 236-258.

Link, M. W., M. P. Battaglia, M. R. Frankel, L. Osborn and A. H. Mokdad (2007). "Reaching the U.S. Cell Phone Generation: Comparison of Cell Phone Survey Results with an Ongoing Landline Telephone Survey." Public Opinion Quarterly **71**(5): 814-839.

Lynberg Black, M., L. Carley-Baxter and S. Twiddy. 2007. Does the Introductory Context Affect Reporting of Victimization and Perpetration in a National Survey of Intimate Partner Violence or Sexual Violence? Paper presented at the American Association for Public Opinion Research Conference, Anaheim, CA.

Politz, A. and W. Simmons (1949). "An Attempt to Get the 'Not at Homes' into the Sample without Callbacks." Journal of the American Statistical Association **44**: 9-16.

Raghunathan, T. E., J. M. Lepkowski, J. Van Hoewyk and P. Solenberger (2001). "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models." Survey Methodology **27**: 85-95.

Tjaden, P. and N. Thoennes (2000). Extent, Nature, and Consequences of Intimate Partner Violence: Findings from the National Violence against Women Survey. NCJ 181867.

--- (2006). Extent, Nature, and Consequences of Rape Victimization: Findings from the National Violence against Women Survey. NCJ 210346.

Tucker, N. C., J. M. Brick and B. Meekins (2007). "Household Telephone Service and Usage Patterns in the United States in 2004: Implications for Telephone Samples." Public Opinion Quarterly **71**(1): 3-22.