Multiple Sources of Nonobservation Error in Telephone Surveys: Coverage and Nonresponse

Andy Peytchev†, Lisa Carley-Baxter†, and Michele Lynberg Black‡

† RTI, International
‡ Centers for Disease Control and Prevention

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.

Abstract

Unit nonresponse has been increasing in Random Digit Dialed (RDD) telephone surveys over the past three decades, creating large potential for nonresponse bias. In addition to obtaining interviews from fewer adults in sampled households, the RDD sampling frame is also including a smaller proportion of the eligible population. The increasing undercoverage is directly related to cell phone substitution for landline phone service. While the proportion of adults without any telephone in the household has remained somewhat low and constant, the proportion of adults with a cell phone but without a landline is relatively large and increasing rapidly. Both nonresponse and undercoverage rates provide potential for bias in survey statistics.

Surveys have limited resources in reducing error due to nonresponse and undercoverage. When an error can not be addressed through design, adjustments are often attempted. An optimum survey design focuses reduction techniques on errors that can least be minimized through adjustment. This requires: (1) the separation and estimation of relative magnitude of different sources of error, and (2) the evaluation of the degree to which each source of error can be statistically adjusted.

We conducted two studies to obtain estimates of nonresponse and coverage bias in a landline RDD survey. A double sample of landline nonrespondents and a RDD sample of cell phone numbers were selected. In addition, postsurvey adjustments were computed for different combinations of studies. Procedures were constructed to isolate coverage from nonresponse bias, and bias was computed for point estimates, variances, and associations. We find significant differences for both nonrespondents and eligible population excluded from the landline frame, that are in opposite direction. Furthermore, differences were found not only for point estimates, but also for estimates of variability and associations. However, conditional on postsurvey

adjustments, nonresponse and coverage bias were in the same direction and compounded. Adjustments were relatively effective in decreasing both sources of bias, although conducting at least one of the additional studies led to less bias in the adjusted estimates compared to the landline study alone. We conclude with a discussion about need for future work on multiple sources of survey error and tracking of statistics beyond point estimates.

1. Introduction

Inference from probability-based surveys relies on the ability to select any member of the eligible population and to obtain interviews from all selected sample members. Failure to include all eligible population in the sampling frame (*undercoverage*) and failure to obtain interviews from all selected (*unit nonresponse*) result in zero probability of inclusion for some and unknown probability of inclusion for others in the population. This makes simple estimators of means, variances, and associations biased, and when those omitted or less likely to be interviewed are systematically different, *bias* in these survey estimates arises.

The potential for coverage and nonresponse bias has been relatively low in random digit dialed (RDD) telephone surveys. In 1996, approximately 95% of the adult U.S. population had access to landline telephones (Belinfante, 1997), and the response rate, for example, in the largest centralized RDD survey in the U.S., the National Immunization Survey (NIS), was 87%[1] (NORC, 2007).

By the end of 2006, only 87% of adults had a landline (Blumberg and Luke, 2007c), and the response rate in the NIS had declined by 25% (NORC, 2007). The decline in response rates is not unique to one survey—the response rates have steadily declined in other major national

---

[1] CASRO (www.casro.org) response rate. Note that surveys and survey organizations code cases differently and that can account for vast differences in response rates. This response rate is unexpectedly high—it is more typical for face to face surveys at that time.

RDD surveys, declining by 19% in the Behavioral Risk Factor Surveillance Study (BRFSS, 1999; BRFSS, 2007), and surveys in other developed countries (de Leeuw and de Heer, 2002), as have coverage rates (Kuusela, Vehovar and Callegaro, 2007). Furthermore, both response and coverage rates continue to decline at rates of around 1.5 and 2 percentage points per year, respectively, (e.g., Curtin, Presser and Singer, 2005; Blumberg and Luke, 2007b), creating an increasingly large potential for bias in survey estimates.

While focus in studying errors of nonobservation is placed on means and proportions (e.g., Groves, 2006), undercoverage and nonresponse can affect not only different estimates, but different *statistics*; absence of bias in the mean of a survey variable does not imply lack of bias in its variance estimate or its association with another variable. In many circumstances, absence of such bias is assumed in the use of adjustments.

Survey statisticians commonly address both undercoverage and nonresponse bias through postsurvey adjustments. Adjustments require assumptions that are rarely testable in any given survey, and are usually based on aggregate information in RDD surveys. The high and increasing undercoverage and nonresponse rates leave great potential for bias in estimates that adjustments may not be effective in minimizing. Indeed, adjustments may even increase bias in estimates when the assumptions do not hold true (e.g., Lin and Schaeffer, 1995). The degree to which adjustment methods are able to decrease bias in estimates needs to be continuously evaluated.

Specifically, postratification of survey weights to demographic totals from a census are commonly employed for adjustment for both undercoverage and nonresponse in RDD surveys. Adjustment of each source of error relies on the association between the vector of census variables (X) and the vector of survey variables (Y), represented by the left arrow in Figure 1.

However, the ability to reduce coverage bias (a function of the covariance between Y and P_C) and nonresponse bias (Y and P_R) also relies on the associations between X and the propensity that a sample member is included in the frame (P_C) and between X and the propensity that a sample member responds to the survey request (P_R). Poststratification subsumes adjustment for the two sources of error, while it may only be effective for undercoverage or for nonresponse, if at all. Knowing for which source of error it is more effective could direct more resources to reducing the other.

<Figure 1 about here>

Not all bias due to nonobservation is equal. Factors leading to not having a landline telephone in the household may be very different from factors influencing cooperation among selected sample members. Hence bias due to undercoverage may be very different in magnitude and even direction from bias due to nonresponse. Since survey methodologists have different tools to address each source of nonobservation error but have limited resources, it is critically important to gauge the impact of undercoverage and nonresponse on survey estimates separately.

While the reasons for exclusion from the frame and nonresponse can be different, the two also differ in demographic characteristics (e.g., Groves and Lyberg, 1988; Blumberg and Luke, 2007a). Since reduction of both sources of error rely on postsurvey adjustments that commonly use demographic characteristics, of interest is the degree to which each bias is reduced through poststratification; it could be that one bias is much larger across estimates, but is much more effectively reduced through adjustment. The degree to which this approach is effective in adjusting error of nonobservation relies on the associations of demographic characteristics with

5

coverage and nonresponse, the association of demographic characteristics with survey variables, and the extent to which the latter associations are the same among those included/excluded from the frame and among respondents/nonrespondents.

Rather than relying on model-based adjustments, bias due to undercoverage and nonresponse can be measured and reduced through additional studies. Because most adults without a landline have a cell phone, a study can be launched on a RDD sample of cell phone numbers. Among nonrespondents, a subsample could be drawn and subjected to a different protocol to gain their cooperation (e.g., by increasing incentives, shortening the instrument). Such methods can be costly and limited available resources may allow only one such additional study, if any.

Thus, an optimal design for a survey in terms of cost and error is one that uses a combination of more expensive additional studies and less expensive adjustments to minimize total error. To do this, relative magnitudes of bias in different statistics, as well as magnitudes of total error need to be compared for undercoverage and nonresponse. In addition, the degree to which each source of error can be effectively adjusted, under different combinations of error reduction studies, needs to be evaluated.

1.1. Nonresponse Reduction

After multiple contact and persuasion attempts under a survey protocol, the remaining nonrespondents are very difficult to interview, for example, in terms of interviewer hours per completed interview. Ideally, the survey protocol will be changed to one in which the nonrespondents would be much more likely to participate. Experiments in surveys have identified features that produce this effect. For example, decreasing the length of the survey,

while limiting the amount of information collected, increases cooperation by more sample members in RDD surveys (McCarty, House, Harman and Richards, 2006). Offering higher incentives increases cooperation (Singer, 2002), particularly among those less interested in the topic who are less likely to cooperate (Groves, Presser and Dipko, 2004). Offering prepaid incentives can also increase cooperation (Singer, Van Hoewyk and Maher, 2000). These can be powerful but costly survey tools to gain respondent cooperation.

When the response rates are relatively low, sending a prepaid incentive and increasing the promised incentive for the remaining nonrespondents may be infeasible. Instead, the survey protocol can be changed and a subsample of nonrespondents can be approached in a nonresponse follow-up (NRFU) study. Drawing a *double sample* allows survey effort to be focused in order to achieve the maximum possible response rate among selected nonrespondents. The inverse of the selection probabilities are then used to combine the data from the NRFU respondents with the original study.

1.2. Undercoverage Reduction

Undercoverage, the failure of a sampling frame to include eligible population, can be reduced through the use of multiple frames. While undercoverage has been rapidly increasing in RDD telephone surveys at a rate of 2% per year (Blumberg and Luke, 2007b), the proportion of adults without any telephone in the household has remained somewhat low and constant with estimates between 1.5-4.9% (Blumberg and Luke, 2007b; Tucker, Brick and Meekins, 2007). The largest and growing proportion of adult U.S. population that is missing from the RDD frame are adults with only cell phones, estimated to be 13.6% in the first half of 2007 (Blumberg and

Luke, 2007b). Launching an additional study on a RDD sample of cell phone numbers reduces most of the undercoverage of the adult U.S. population.

Methods of obtaining dual frame survey estimates have been developed for the case when frame membership of sample members is known (Hartley, 1962). An RDD study that is designed as dual frame would include questions about cell phone and landline use from respondents in both frames, if respondents with both types of phone service are interviewed in both samples. In a *post-hoc* dual frame design in which the cell phone sample has not been planned and cell phone use was not asked in the landline sample, screening for adults with only cell phones will be necessary to avoid further assumptions. Additionally, some studies elect to screen for adults with only cell phones even in planned dual frame sample designs, due to currently higher cost per completed interview from this frame and lack of population level information to mix groups in the overlap of the frames.

The lack of demographic information on the landline and cell phone populations hinders dual frame estimates even in the simple case when those with both types of service are included only from only one of the frames. Nonresponse adjustments need to be created prior to the combining of the samples in order to avoid the tenuous assumption that landline and cell phone respondents have the same expected values within demographic groups.

1.3. Nonresponse and Undercoverage Adjustment

Arguably a much less costly alternative to reduction of nonobservation error is adjustment. Unfortunately, very little information is available on RDD samples for adjustments and data users have to rely heavily on the ability of poststratification to population demographic

totals to effectively reduce nonresponse and coverage bias in estimates. Common variables used in poststratification include census region, age, sex, race and ethnicity, income, and/or education.

1.4. Nonresponse and Coverage Bias in Multiple Statistics

Research on bias in survey estimates, the difference between a realized estimate and the intended estimate, has focused on *means and proportions*. However, bias can also affect the simple response variance of means and proportions. It can be understated through bias in means, and it can also reflect different *variability* among those omitted from the survey estimates. It is the latter that may be of interest to those seeking unbiased significance tests, as well as to researchers engaged in explaining differences between respondents and nonrespondents. It may also be of interest to researchers engaged in explaining differences between those included and those excluded from the survey frame—e.g., why would nonrespondents be more variable in their responses to a particular question?

Bias can occur for other statistics of interest. Some uses of survey data may be indifferent to bias in means. However, bias in *associations* can be present even in the absence of bias in means, and affecting estimates such as regression coefficients. Lepkowski and Couper (2002), for example, examined nonresponse bias in associations among a handful of variables in a face to face survey, failing to detect large differences. Such findings can vary across variables and modes of data collection and need to be examined in RDD surveys. To date, no examinations of coverage bias in associations could be found in the literature.

The effect of nonresponse and undercoverage is not constrained to *bias* in a particular statistic. Total error, as captured by the *mean square error* (MSE), reflects both bias and variance. Debatably, an optimum survey design is one that minimizes MSE for multiple

estimates, increasing the likelihood that a survey estimate is closer to the population parameter across replications of the survey.

2. Methods

We use data from the National Intimate Partner and Sexual Violence Survey (NISVS) Pilot Study, a national RDD survey of adults in the U.S. measuring perpetration and victimization of stalking, sexual violence, physical aggression, and psychological aggression. The sample included both RDD and listed numbers, and males were selected at a higher rate to achieve desired number of respondents by sex, due to males responding at lower rates (while this is reflected in weighting adjustments, it leads to lower overall unweighted response rate). Interviews were conducted January-April, 2007. The NISVS Pilot Study achieved a response rate of 21.5%[2] and did not include cell phone numbers. Embedded experiments showed that the response rate was slightly affected by the announced topic (Lynberg, Carley-Baxter and Twiddy, 2007) and incentive amount (Carley-Baxter, Black and Twiddy, 2007), but the entire sample had received substantial effort, with most active cases at the end of the study having received at least 20 call attempts.

While nonresponse accounts for a larger proportion of the sample than the proportion undercoverage in the population due to cell phone only (i.e., nonresponse rate is higher than the undercoverage rate), the cell phone only population shares characteristics that are related to survey variables: living with unrelated roommates, young, single, and low income, (e.g., Tjaden and Thoennes, 2000; Tjaden and Thoennes, 2006). This could lead to coverage bias that is as large as, or even larger than, nonresponse bias, with implications that are contrary to the common

---

[2] American Association for Public Opinion Research, Response Rate 4 (AAPOR (2006). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. Lenexa, Kansas, AAPOR. **4th edition**.).

focus on nonresponse in RDD studies. Furthermore, if biases run in opposite directions, conducting a study only for one source of error can actually increase bias in survey estimates.

To evaluate the potential undercoverage and nonresponse biases in the NISVS Pilot, we conducted two concurrent follow-up studies four months after the pilot was completed. For the nonresponse study, changes were made to the survey protocol to increase the likelihood of participation and to gain information on nonrespondents. Three changes were implemented: (1) a $5 prepaid incentive was mailed in prenotification letters to the 66% of cases for which an address could be matched, (2) a promised incentive of $20 offered to all sample members instead of the $10 vs. $20 incentive experiment in the Pilot Study, and (3) a substantially shorter instrument that maintained the same context for key survey questions but decreased the average interview length from 30 to 14 minutes (by asking only about victimization experiences and omitting questions related to perpetration). A subsample of 7,768 nonrespondents from the NISVS Pilot Study was selected for the nonresponse study, and interviewing was conducted September-October, 2007, with most interviewing ceased during the first half of October due to call center constraints. Among these nonrespondents, a response rate of 6.4% was achieved, or a combined (NISVS Pilot and nonresponse studies) response rate of 26.5%.

Concurrently, a study was conducted to measure coverage error. An RDD sample of 6,254 cell phone numbers was selected. Because lists are not available for cell phone numbers, prenotification letters were not mailed. However, the protocol was modified a few days into data collection to begin leaving voicemail messages with a similar intent. After the interview, a promised incentive of $25 was sent by mail, following a protocol designed to insure confidentiality. Personally identifiable information were kept in separate files and not linked to individual responses. The main part of the cell phone interview was the same shortened

instrument used in the nonresponse study. However, the screening questions differed from those used in the NISVS Pilot and nonresponse studies for three reasons: (1) for safety concerns, interviews were stopped with respondents who were driving, (2) different questions were needed to determine eligibility and selection probability, and (3) questions were asked about cell phone use to inform the design of future studies. The second set of differences includes important design features. Cell phones were treated as personal devices rather than household devices; therefore, within household selection was not conducted. Since cell phone use among NISVS Pilot respondents is unknown, selection probabilities for adults with both types of telephone service from each frame can not be estimated. Instead, adults with only cell phones were interviewed from the cell phone frame. The length of this instrument was similar to the nonresponse study, taking 15 minutes on average. Problems with obtaining sufficient interviewing hours, having mostly new interviewers on staff—the same constraints present for the nonresponse study as they were conducted concurrently by the same staff and interviewers— as well as difficulty in estimating optimum days and times for call attempts and lack of knowledge about cell phone respondent concerns, made interviewing difficult, despite efforts to optimize call scheduling based on daily call record data and additional interviewer training. However, additional calls under the same survey design could have very likely led to similar survey estimates (e.g., Curtin, Presser and Singer, 2000; Keeter et al., 2000). One-hundred and thirty-two interviews were collected with a response rate of 4.3%, or a combined (NISVS Pilot, nonresponse, and cell phone) response rate of 23.9%.

Combining of the NISVS Pilot, nonresponse and cell phone samples requires special attention. Different selection probabilities of nonrespondents can be estimated under alternate assumptions about the remaining nonrespondents. Simply accounting for selection probabilities

and concatenating the landline and cell phone data confounds multiple sources of error; each one can be affected by nonresponse in different ways. However, addressing problems of confounding nonresponse and coverage error is hindered by lack of information on telephone numbers in RDD surveys and even more so by the unavailability of census demographic information on the cell phone and landline frames separately.

In creating postsurvey adjustments, the strategy was to apply the least stringent assumptions about nonrespondents and the relationship between nonresponse and coverage bias. There are two justifiable methods of adjusting the base weight for respondents in a nonresponse study with a double sample (i.e., subsample of nonrespondents). Under a deterministic approach, weights for respondents in the nonresponse study would be increased to account for *all* remaining nonrespondents from the original sample. This makes the assumption that the respondents under the modified survey protocol are representative of all nonrespondents; that they are a group predetermined to be nonrespondents and it was only the new protocol that gained their cooperation. Instead, under a far more plausible stochastic model, remaining nonrespondents vary in the probability that they could have been respondents in the original survey, in the nonresponse follow-up, or remained nonrespondents. Under this approach, respondents in the nonresponse study are assigned another selection weight to account for the double sample, but nonresponse adjustments are made only after pooling together the data from the initial data collection and the nonresponse follow-up. An additional benefit from this approach is an avoidance of a large increase in loss of efficiency from a nonresponse weight adjustment. This is the approach used in this study, presented in the context of the entire adjustment design in Figure 2.

<Figure 2 about here>

A much more difficult problem is the combining of frames (i.e., landline and cell phone) in the inevitable presence of unit nonresponse. If nonrespondents in one sample are different from respondents in the other frame, nonresponse will create bias in dual-frame estimators. That is, an assumption is made that the nonrespondents in one frame can be substituted by respondents in either frame, asserting a lack of differences between frames—the implausibility of which is the very motivation to conduct a complex multi-frame survey that addresses coverage bias. Therefore, nonresponse adjustments need to be created prior to merging the samples from the two frames. While the Census Bureau does not provide demographic information separately for each frame, the National Health Interview Survey provides the proportion of adults in the U.S. with only cell phones in categories for demographic variables. Combined with census population estimates for each of these categories, separate demographic totals for each frame can be computed. Unfortunately, this yields marginal distributions rather than full cross classification. To create the adjustments within each frame, any missing values for age, sex, race/ethnicity, and region were imputed with IVEware (Raghunathan, Lepkowski, Van Hoewyk and Solenberger, 2001) and raked to marginal population distributions for each frame using a SAS macro program (Izrael, Hoaglin and Battaglia, 2004). Only then were the samples from the two frames combined, and poststratified to the full cross-classification of the demographic variables using a SAS macro for weight calibration using general exponential modeling (Folsom and Singh, 2000). Extreme values were trimmed within each study and sex, and weights recalibrated to two-way cross-classifications to match the population distribution.

Parts of the weighting procedures, presented in Figure 2, were omitted depending on the purpose; multiple poststratified weights were calculated to allow population estimates using only the NISVS Pilot Study, the Pilot and Nonresponse Studies, Pilot and Cell Phone Studies, and Pilot, Nonresponse, and Cell Phone Studies.

Key measures in the survey are victimization from stalking, sexual violence, physical aggression, and psychological aggression. The items measuring each construct (10, 2, 13, and 12 items, respectively), presented in the Appendix, were combined into scales and since an objective in the NISVS survey is to obtain prevalence rates for these victimizations, dichotomous indicators were also created. This provides the ability to evaluate bias in different statistics based on the same measures. Because of the differences between male and female experiences with sexual and intimate partner violence, analyses are conducted separately for each sex.

We first contrast respondents from each of the three studies on demographic characteristics to gauge the potential for coverage and nonresponse bias. We also look at the association between the demographic variables and survey measures to evaluate the degree to which the demographic variables will be effective in adjustments for each bias.

We then turn to direct comparisons of the nonresponse and cell phone studies respondents to the NISVS Pilot Study respondents on key survey estimates, providing absolute and relative nonresponse and coverage bias. In addition, simple response variances are compared across the groups to evaluate the degree to which bias can be exhibited in the degree of variability within each group. Differences in associations are tested, a statistic for which bias is commonly ignored or assumed to be absent.

We then compare population estimates based on the full information from the three studies to estimates based on data limited to (a) only the NISVS Pilot Study, (b) the Pilot and

Nonresponse Studies, and (c) the Pilot and Cell Phone Studies. This allows us to evaluate the

degree to which each of these sources of nonobservation error contribute to bias in the NISVS

Pilot survey estimates, their relative magnitudes, and their combined effect. It also provides an

assessment of the degree to which post-survey adjustments can compensate for each source of

nonobservation error in this study. Finally, we compare the total error and its components across

each data collection scenario.

3. Results

Respondents in both nonresponse and cell phone studies were significantly different from the

NISVS Pilot Study respondents, shown in Table 1. Furthermore, more differences were found

for the cell phone study, and they were substantially larger in magnitude. For example, 91.0% of

the respondents in the NISVS Pilot Study were 30 years of age or older. This percentage was

only 2.7 percentage points higher in the nonresponse study, but it was 38.7% percentage points

lower in the cell phone study. Another noteworthy result is the direction of the differences.

Apart from sex (by which analyses have to be stratified due to the sex-specific phenomena), for

all demographic variables for which both nonresponse and cell phone respondents differed from

the NISVS Pilot, the differences were in opposite directions. This has multiple implications,

supporting the need to conduct studies for both nonresponse and coverage, and certainly for the

use of these variables in poststratification *prior* to merging samples.

<Table 1 about here>

Another condition for the effectiveness of postsurvey adjustments using these variables is an association between the demographic characteristics and the key survey variables, indicators and scales for stalking, sexual violence, physical aggression, and psychological aggression. Those correlations were relatively low and similar for males and females, ranging from <0.001 to 0.156, with significant associations with age and race (results not presented). An expected exception was the considerably higher association between sexual victimization and being female, which was 0.259.

In both the nonresponse and cell phone studies, respondents differed significantly on estimates of means and proportions for some of the key survey variables, with estimates weighted for selection probability presented in Table 2. Differences were larger for the cell phone study but reached significance for fewer variables as a substantially fewer number of interviews were collected, compared to the nonresponse study. Differences were also in opposite directions; estimates in the nonresponse study were lower, while estimates in the cell phone study were higher than the NISVS Pilot Study. This implies that failure to conduct a study targeted at either source of survey bias can lead to even greater bias.

<Table 2 about here>

Differences in means between the nonresponse and cell phone, and the NISVS Pilot Study were significant for different variables; at the .05 level of significance, nonrespondents had lower prevalence on female stalking and male physical and psychological aggression, while cell phone respondents were higher on male psychological aggression.

Significant differences were found also in estimates of variability—some of the standard deviations for both the nonresponse and cell phone studies were significantly different from those in the NISVS Pilot Study. Similar to the differences in means, although for different variables, nonresponse and coverage bias in variability affect mostly different variables; at the .05 level of significance, nonrespondents were less variable on female sexual violence and physical aggression, and on male stalking, while cell phone respondents were more variable on male stalking, sexual violence, and psychological aggression. While each source of error leads to bias in the estimates of variability, the overall standard deviations presented in the last column in Table 2 are almost identical to those based only on the NISVS Pilot data; these biases seem to cancel each other out for this study and set of variables.

Bias was also found for associations (Table 3), with a similar pattern: associations were lower in the Nonresponse Study and higher in the cell phone study (although not significant for females), compared to the NISVS Pilot Study respondents, for the three variables that had a sufficient number of scale points. The tests for equality of variance-covariance matrices were conducted on the natural log of the variables to remedy deviations from the multivariate normal distributions in kurtosis. However, the three variables were left-censored, with many respondents having reported zero, although they could still vary on the constructs of interest. The analysis was repeated using Tobit regression, regressing each variable on the other, on the nonresponse and cell phone study indicators, and on their interactions, separately for each sex. Significant interactions would show different associations between the pairs of variables across the studies. These interactions were significant in all models, for both females and males (not shown here).

<Table 3 about here>

Poststratified estimates of the means and proportions, presented in Table 4, reveal different aspects of bias. First, these adjustments seem quite effective at reducing bias in estimates from one source of error when a study has been conducted to address the other—poststratified estimates for NISVS Pilot and Nonresponse, and for Pilot and Cell Phone, come quite close to those based on the Pilot alone.

<Table 4 about here>

Second, and this underscores why differences between studies do not tell the same story as differences in population estimates, is a reversal of bias from undercoverage. Based on demographic characteristics alone, one would expect that omitting adults with only cell phones who are more likely to be young, low income, and in a minority group, will lead to underestimates of sexual, physical, and psychological violence victimization. Indeed, this was evident in the selection-weighted differences in Table 2. However, *conditional* on these characteristics achieved through poststratification, cell phone respondents were slightly *less* likely than NISVS Pilot respondents to have been victimized, i.e., 38.9% vs. 40.0% for female stalking and 52.3% vs. 53.8% for male psychological aggression. While this conditional effect is well known in statistics (i.e., Simpson's Paradox), it is of great importance here as RDD studies that exclude cell phone numbers and speculate about the direction of coverage bias, may be gravely wrong.

Thirdly, biases compound. For example, weighted female stalking based on the Pilot

Study alone was 40.0%. Adding the nonresponse study, it decreased to 38.6%, while adding the

cell phone study instead, it decreased to 38.9%. Adding both nonresponse and cell phone

studies, the estimated proportion drops further to 37.5%. This pattern occurs again for male

physical and psychological aggression. In the absence of studies measuring nonresponse and

coverage bias, researchers hope that biases cancel out, particularly in light of the differences in

the opposite directions shown in Table 2. Yet once adjusted, biases can compound in population

estimates.

An optimal design may incorporate a mix of methodological studies and statistical

adjustments that can minimize bias, variance, or both. Figure 3 presents MSE and its

components under four design scenarios, each poststratified to the adult U.S. population: (1)

NISVS Pilot Study, (2) Pilot and Nonresponse Studies, (3) Pilot and Cell Phone Studies, and (4)

Pilot, Nonresponse, and Cell Phone Studies. Examining MSE, it is apparent that the preferred

design is different for various estimates. For example, the NISVS Pilot Study alone yields the

lowest MSE for female physical and psychological aggression, but yields the highest MSE for

these estimates for males. Despite the increase in variance by the relatively small number of

respondents in the cell phone study and by the subsampling of nonrespondents in the

nonresponse study, both studies led to lower MSE for estimates for males, although neither the

nonresponse nor cell phone study is clearly superior over the other in terms of reducing MSE. If

only one additional study could be afforded by a project, the cell phone study would be preferred

if greater interest lies in male victimization from physical aggression, while the nonresponse

study would be preferred if more emphasis is placed on male victimization from psychological

aggression. Based on estimates for males, both cell phone and nonresponse studies are needed to reduce MSE.

<Figure 3 about here>

Similarly, if interest is in either bias or variance rather than both, the optimal combination of studies depends on the estimate of interest; the only clear preference is one by design—there is no bias when both nonresponse and coverage studies are conducted as this serves as the gold standard.

4. Discussion and Conclusions

This study presents a rare opportunity to measure bias and total error from undercoverage and nonresponse in an RDD survey, and to evaluate the extent to which postsurvey adjustment are effective in reducing error from these sources. We use a broader definition of bias, which if present, can lead to biased significance tests and biased multivariate analyses. There are several noteworthy findings:

1. Both respondents in the nonresponse and cell phone studies were different from respondents in the NISVS Pilot Study on demographic characteristics, but differences were much larger and mostly in the opposite direction in the cell phone study.

If interest lies in bias in a demographic statistic, differences in opposite directions due to each source of survey error may be good news. However, when interest is in survey variables

for which demographic characteristics are mostly used for adjustments, this leads to great complications and likely interactions between errors. In particular, demographic information on the cell phone and landline populations are not yet available from official government statistics. Merging samples from a cell phone and a landline frame assumes that nonrespondents and respondents from the two frames are interchangeable—yet the implausibility of sample members from the two frames to be equivalent is the very reason of launching dual frame designs in RDD surveys.

This study presented a method to create nonresponse adjustments within the samples from each frame, prior to merging them together, using a combination of census and survey demographic information. While we hope that direct estimates will be available in the future, such alternative procedures are necessary to minimize the confounding of survey errors that could lead to bias in survey estimates when done out of sequential order.

2. Both nonresponse and coverage bias was found not only for means and proportions, but also for estimates of variability and associations among survey variables.

As a primary purpose of surveys is to produce descriptive statistics on the population, focus has been placed on bias in means and proportions. However, obtaining biased estimates for the degree of variability in a survey variable also leads to erroneous inference about the population and leads to error in significance tests.

Much of social research uses survey data for describing phenomena in the population through multivariate analyses. For such purposes, bias in means may not affect any substantive conclusion; it is the potential for differences in associations between those excluded and those

included in the frame and between nonrespondents and respondents that are of interest. Such differences were found with respondents in both the nonresponse and cell phone studies.

3. Differences between nonresponse and NISVS Pilot Study respondents were in opposite direction to differences between cell phone and NISVS Pilot Study respondents, and affected mostly different estimates.

This was true for differences in point estimates, variances, and associations; all three types of statistics were *lower* in the nonresponse study and *higher* in the cell phone study, compared to the NISVS Pilot respondents. In this study, an analysis not using poststratification weights can yield even more biased estimates if only nonresponse or coverage is addressed. However, these are unconditional estimates—see (5) below.

4. Adjustments were somewhat effective in reducing bias due to nonresponse and undercoverage.

Comparing poststratified estimates of proportions from the Pilot to those from the Pilot and either the Nonresponse or Cell Phone Study, most differences were below 1 percentage point. Similar results are found when comparing estimates from all three studies to those based on the Pilot and Nonresponse, or Pilot and Cell Phone Studies. However, some bias still remained despite adjustments—this is only expected as the demographic characteristics used in computing weights are, as commonly the case, not strongly associated with survey measures.

5. Biases compound. Conditional on adjustments, coverage bias changed direction and compounded to that from nonresponse.

Rather than nonresponse and coverage error cancelling out as suggested by the differences in estimates from each study (Table 2), bias compounded in the poststratified estimates (Table 4 and Figure 3). As expected based on demographic characteristics of adults with only cell phones, cell phone respondents had *higher* reports of victimization. However, conditional on demographic characteristics, they actually had *lower* reports of victimization compared to the NISVS Pilot (landline) respondents. For example, cell phone respondents are more likely to have characteristics such as being of Hispanic origin, which are associated with higher reports of victimization, but controlling for these characteristics that are typically in postsurvey adjustments, cell phone respondents were less likely to have been victimized. Respondents in the nonresponse study had *lower* unconditional and conditional reports. Hoping that errors would cancel, such as based on demographic correlates that may suggest a positive bias from nonresponse and a negative bias due to undercoverage, would be erroneous; even if only one additional study could be afforded to gain purchase on one of the sources of survey error, a study should still be implemented rather than expect that different biases would cancel out. In NISVS, bias was arguably less intuitive from undercoverage and a cell phone study may have been preferred.

6. Optimal design to reduce nonresponse and coverage error is estimate-specific.

Unfortunately, this does not provide guidance for what domains of interest a particular survey design should be used. Arguably, these variables are in somewhat the same substantive domain, yet for example, the addition of a cell phone study decreased MSE in physical aggression among males more than the nonresponse study, whereas the opposite combination was best for psychological aggression. The same finding holds if interest is only in bias rather than total error. Based on the limited evidence from this study, a likely conclusion is reached: addressing one source of survey error is beneficial, but addressing all is ideal in order to minimize survey error in different estimates and statistics. This approach may be problematic in that it is likely to be associated with higher costs and increased study design complexity, and survey design may have to be optimized for a subset of estimates.

While being able to address multiple sources of nonobservation error in the same study, interpretation should be tempered to some degree. The low response rates across the studies leave potential for estimates to be different had all sample members been interviewed, and particularly threatening is the possibility that nonresponse bias in each study is different. We attempted to control this threat by keeping survey protocols as similar as possible. While response rates could have been higher had the nonresponse and cell phone studies not been subjected to limitations such as few interviewers and relatively short and interrupted field period, they were subjected mostly to the same limitations and conducted concurrently. Ideally, such studies would be planned together with the main landline survey that would alleviate many problems of implementation, and we hope that this research helps support their routine inclusion.

Future research is needed on identification of individual error sources in the context of total survey error. We know that all sources of survey error affect estimates from a survey, yet it

is the identification of their relative magnitudes that would allow the minimizing of the total

error in the context of study constraints. Isolation of individual error sources is challenging at

best as they are usually confounded. Innovative designs, analytic approaches, and replication of

studies are needed to separate different errors. For example, effort was made in this study to

separate nonresponse and coverage error, yet an error source that could not be addressed with

these data is the confounding with measurement error. An argument could be made that lower

reports in the nonresponse study were the result of underreporting among respondents with lower

response propensities, if there are common causes for not reporting victimization and for not

participating in the survey. Similarly, respondents may also be more likely to underreport

victimization if interviewed over cell phones, from fearing disclosure over this medium, for

example, or sample members with only cell phones may be individuals who are less likely to

disclose such information.

Even identification of error sources that are being targeted requires replication—both

nonresponse and cell phone studies were subjected to low response rates; the nonresponse

adjustments within each study can fail to address any remaining bias sufficiently. A possible

association between P_C and P_R (Figure 1) could not be evaluated under the current design, but

such an interaction between undercoverage and nonresponse is possible—i.e., cell phone

nonrespondents may be inducing the most bias in these survey estimates. Work is also needed in

variance estimation for adjustment of nonresponse and coverage error—variance estimates and

comparisons of MSE reported here excluded any additional variance that results from the

uncertainty in the adjustments, due to the complex adjustment procedure in these studies (i.e.,

imputation, three poststratification steps to each sample, and weight trimming). However, future

research could attempt incorporating the often overlooked variance component from uncertainty in postsurvey adjustments.

Identification of optimal combinations of reduction versus adjustment of survey error is a complex but necessary objective. Surveys currently implement both, but with little guidance on which error source requires (often) cost-intensive reduction methods more than another error source. This requires not only the relative magnitude of error sources, but also evaluations of the degree to which different adjustments are able to effectively address each error source, as a less cost-intensive alternative to reduction.

Another area that is in need of further attention is bias in different statistics. Research has already demonstrated that bias in means and proportions, such as from nonresponse, varies greatly across estimates, i.e., the variable of interest (e.g., Groves, 2006). However, survey data has multiple uses. As demonstrated here, sources of survey error such as nonresponse and coverage, can affect other statistics (e.g., biasing variances and multivariate associations).

While not among the objectives of this paper, anticipation of nonresponse and of coverage error, and improving the ability to adjust for them, can best be addressed by studying their causes. While considerable attention has been devoted to the study of reasons for nonparticipation in surveys, far less has been done on reasons for exclusion from the sampling frame, and dropping landline telephone service in particular. Understanding such reasons would allow anticipation of bias in particular survey estimates, as well as inform the collection of relevant correlates to be used in adjustments. Furthermore, causes of different sources of error may overlap—for example, social isolation has been forwarded as a reason for unit nonresponse, yet it could also be a reason for keeping only a cell phone. Such theoretical work is direly needed.

References

AAPOR (2006). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for

Surveys. Lenexa, Kansas, AAPOR. **4th edition**.

Belinfante, A. (1997). Telephone Subscribership in the United States. Washington, DC, Federal

Communications Commission.

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.

BRFSS (1999). 1999 Behavioral Risk Factor Surveillance System Summary Data Quality

Report, CDC.

--- (2007). 2006 Behavioral Risk Factor Surveillance System Summary Data Quality Report,

CDC.

Carley-Baxter, L., M. Black and S. Twiddy (2007). The Impact of Incentives on Survey

Participation and Reports of Intimate Partner and Sexual Violence. American Association

for Public Opinion Research annual conference. Annaheim, CA, American Statistical

Association.

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.

--- (2005). "Changes in Telephone Survey Nonresponse over the Past Quarter Century." Public Opinion Quarterly **69**(1): 87-98.

de Leeuw, E. and W. de Heer (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. Survey Nonresponse. R. Groves, D. Dillman, J. Eltinge and R. J. A. Little. New York, Wiley**:** 41-54.

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 L. E. Lyberg (1988). An Overview of Nonresponse Issues in Telephone Surveys. Telephone Survey Methodology. R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II and J. Waksberg. New York, NY, John Wiley & Sons**:** 191-211.

Groves, R. M., S. Presser and S. Dipko (2004). "The Role of Topic Interest in Survey Participation Decisions." Public Opinion Quarterly **68**(1): 2-31.

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. 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.

Kuusela, V., V. Vehovar and M. Callegaro (2007). Mobile Phones' Influence on Telephone Surveys. Advances in Telephone Survey Methodology. J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. D. Leeuw, L. Japec, P. J. Lavrakas, M. W. Link and R. L. Sangster. New York, Wiley & Sons.

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.

Lynberg, 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? American Association for Public Opinion Research Conference. Anaheim, CA.

McCarty, C., M. House, J. Harman and S. Richards (2006). "Effort in Phone Survey Response Rates: The Effects of Vendor and Client-Controlled Factors." Field Methods **18**(2): 172-188.

NORC (2007). National Immunization Survey: A User's Guide for the 2006 Public-Use Data File, CDC.

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.

Singer, E. (2002). The Use of Incentives to Reduce Nonresponse in Household Surveys. Survey
    Nonresponse. R. M. Groves, D. A. Dillman, J. L. Eltinge and R. J. A. Little. New York,
    Wiley**:** 163-177.

Singer, E., J. Van Hoewyk and M. P. Maher (2000). "Experiments with Incentives in Telephone
    Surveys." Public Opinion Quarterly **64**(2): 171-188.

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.

Table 1. Demographic Characteristics of Respondents in the NISVS Pilot, Nonresponse, and Cell Phone Studies.

Indicator (1=yes)	Pilot n=5,296 (%)	Std.Err.	Nonresponse n=411 (%)	Std.Err.	Cell Phone n=132 (%)	Std.Err.	Pilot minus Nonresp. (%)	Pilot minus Cell (%)
Age 30+	91.0%	(0.4%)	93.7%	(1.2%)	52.3%	(4.3%)	2.7% **	-38.7% ***
Age 50+	58.3%	(0.7%)	64.7%	(2.4%)	17.4%	(3.3%)	6.4% ***	-40.9% ***
Male	46.9%	(0.7%)	57.2%	(2.4%)	61.4%	(4.2%)	10.3% ***	14.5% ***
Midwest Region	26.1%	(0.6%)	24.6%	(2.1%)	36.4%	(4.2%)	-1.5%	10.3% **
Northeast Region	17.2%	(0.5%)	21.2%	(2.0%)	13.6%	(3.0%)	4.0% *	-3.6%
South Region	35.9%	(0.7%)	33.3%	(2.3%)	23.5%	(3.7%)	-2.6%	-12.4% ***
West Region	20.8%	(0.6%)	20.9%	(2.0%)	26.5%	(3.8%)	0.1%	5.7%
Hispanic	7.1%	(0.4%)	5.6%	(1.1%)	21.2%	(3.6%)	-1.5%	14.1% ***
Nonhispanic Black	8.5%	(0.4%)	6.3%	(1.2%)	13.6%	(3.0%)	-2.2% *	5.1% *

* p<.10, ** p<.05, *** p<.01

Table 2. Key Survey Estimates in the NISVS Pilot, Nonresponse, and Cell Phone Studies Weighted for Selection Probabilities and Unweighted Standard Deviations.

Variable (scale)	Pilot[a] Mean (Std.Err.)	Std. Dev.	Nonresponse[b] Mean (Std.Err.)	Std. Dev.	Cell Phone[c] Mean (Std.Err.)	Std. Dev.	Overall Std. Dev.
Females							
Stalking (1=yes)	38.8% (1.2%)	48.8%	26.6% ** (4.6%)	45.0%	36.2% (7.0%)	48.8%	48.6%
Sexual Violence (1=yes)	22.1% (1.0%)	43.1%	16.2% * (3.2%)	39.6%	26.9% (6.4%)	45.1%	43.0%
Physical Aggr. (1=yes)	38.7% (1.2%)	48.9%	35.7% (4.9%)	48.0%	43.9% (7.2%)	50.4%	48.9%
Psych. Aggr. (1=yes)	54.7% (1.2%)	49.7%	50.6% (5.0%)	50.0%	67.5% * (6.9%)	47.1%	49.8%
Stalking (0-10)	1.21 (0.05)	2.07	0.88 * (0.19)	1.86 *	1.00 (0.26)	2.32	2.06
Sexual Violence (0-2)	0.31 (0.02)	0.67	0.23 (0.05)	0.59 **	0.34 (0.09)	0.63	0.67
Physical Aggr. (0-13)	1.44 (0.06)	2.53	1.39 (0.26)	2.24 **	1.37 (0.31)	2.77	2.52
Psych. Aggr. (0-12)	2.95 (0.08)	3.07	2.71 (0.31)	2.96	3.63 (0.43)	3.17	3.06
Males							
Stalking (1=yes)	29.0% (1.2%)	45.4%	25.2% (3.6%)	43.3%	36.8% (5.5%)	48.2%	45.3%
Sexual Violence (1=yes)	5.2% (0.5%)	23.3%	6.3% (2.5%)	21.2% *	11.6% * (3.7%)	31.6% ***	23.4%
Physical Aggr. (1=yes)	44.2% (1.3%)	49.7%	35.5% ** (3.9%)	47.7%	50.9% (5.7%)	50.3%	49.6%
Psych. Aggr. (1=yes)	52.5% (1.3%)	49.9%	43.8% ** (4.0%)	49.6%	55.8% (5.9%)	50.1%	50.0%
Stalking (0-10)	0.69 (0.04)	1.49	0.62 (0.18)	1.08 ***	1.03 (0.21)	1.94 ***	1.48
Sexual Violence (0-2)	0.07 (0.01)	0.30	0.08 (0.03)	0.29	0.14 (0.05)	0.41 ***	0.31
Physical Aggr. (0-13)	1.61 (0.06)	2.47	1.36 (0.23)	2.26 *	1.88 (0.30)	2.84 *	2.46
Psych. Aggr. (0-12)	2.42 (0.07)	2.44	2.04 * (0.21)	2.28	3.23 ** (0.36)	3.17 ***	2.46

* p<.10, ** p<.05, *** p<.01

Note: Standard deviations were tested using Levene's test statistic for equality of variances.

[a] n=2814 for females, n=2482 for males.

[b] n=176 for females, n=235 for males.

[c] n=51 for females, n=81 for males.

Table 3. Associations among Log-Transformed Key Survey Variables in the NISVS Pilot, Nonresponse, and Cell Phone Studies.

Variable (scale)	Pilot[a]			Nonresponse[b]			Diff Sig.	Cell Phone[c]			Diff Sig.
	Stalk.	Phys. Aggr.	Psych. Aggr.	Stalk.	Phys. Aggr.	Psych. Aggr.		Stalk.	Phys. Aggr.	Psych. Aggr.	
Females											
Stalking (0-10)	0.49			0.43				0.51			
Physical Aggr. (0-13)	0.29	0.59		0.22	0.53			0.21	0.62		
Psych. Aggr. (0-12)	0.29	0.43	0.72	0.26	0.47	0.75	n.s.	0.21	0.45	0.65	n.s.
Males											
Stalking (0-10)	0.31			0.22				0.44			
Physical Aggr. (0-13)	0.17	0.59		0.13	0.52			0.37	0.67		
Psych. Aggr. (0-12)	0.14	0.36	0.57	0.14	0.31	0.59	**	0.33	0.49	0.75	***

* p<.10, ** p<.05, *** p<.01
Significance based on Box's M test, contrasting to Pilot.
[a] n=2814 for females, n=2482 for males.
[b] n=176 for females, n=235 for males.
[c] n=51 for females, n=81 for males.

Table 4. Key Survey Estimates Based on NISVS: (1) Pilot, (2) Pilot and Nonresponse, (3) Pilot and Cell Phone, and (4) Pilot, Nonresponse, and Cell Phone Studies, Weighted by Final Poststratified Weights.

Variable (scale)	Pilot[a]		Pilot & Nonresponse[b]		Pilot & Cell[c]		Pilot, Nonresponse, and Cell Phone[d]	
	Mean	Std.Err.	Mean	Std.Err.	Mean	Std.Err.	Mean	Std.Err.
Females								
Stalking (1=yes)	40.0%	(1.4%)	38.6%	(1.3%)	38.9%	(1.5%)	37.5%	(1.5%)
Sexual Violence (1=yes)	22.1%	(1.1%)	21.7%	(1.1%)	22.0%	(1.2%)	21.7%	(1.3%)
Physical Aggr. (1=yes)	38.7%	(1.3%)	38.3%	(1.3%)	38.6%	(1.5%)	38.5%	(1.5%)
Psych. Aggr. (1=yes)	55.8%	(1.4%)	55.1%	(1.4%)	56.0%	(1.5%)	55.5%	(1.5%)
Stalking (0-10)	1.28	(0.06)	1.24	(0.06)	1.24	(0.07)	1.21	(0.08)
Sexual Violence (0-2)	0.31	(0.02)	0.31	(0.02)	0.31	(0.02)	0.30	(0.02)
Physical Aggr. (0-13)	1.45	(0.07)	1.47	(0.08)	1.42	(0.08)	1.45	(0.09)
Psych. Aggr. (0-12)	2.99	(0.09)	2.98	(0.09)	3.01	(0.10)	3.01	(0.11)
Males								
Stalking (1=yes)	29.6%	(1.2%)	28.7%	(1.2%)	29.9%	(1.4%)	29.4%	(1.4%)
Sexual Violence (1=yes)	5.8%	(0.6%)	5.6%	(0.6%)	5.6%	(0.6%)	5.6%	(0.6%)
Physical Aggr. (1=yes)	45.2%	(1.3%)	43.6%	(1.4%)	44.5%	(1.5%)	43.1%	(1.5%)
Psych. Aggr. (1=yes)	53.8%	(1.4%)	51.8%	(1.4%)	52.3%	(1.5%)	50.9%	(1.5%)
Stalking (0-10)	0.74	(0.05)	0.72	(0.04)	0.73	(0.05)	0.71	(0.05)
Sexual Violence (0-2)	0.08	(0.01)	0.07	(0.01)	0.07	(0.01)	0.07	(0.01)
Physical Aggr. (0-13)	1.69	(0.07)	1.64	(0.08)	1.62	(0.08)	1.57	(0.08)
Psych. Aggr. (0-12)	2.50	(0.07)	2.41	(0.07)	2.44	(0.08)	2.38	(0.08)

[a] n=2814 for females, n=2482 for males.
[b] n=2990 for females, n=2717 for males.
[c] n=2865 for females, n=2563 for males.
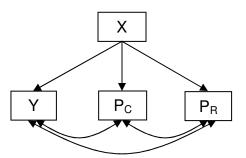[d] n=3041 for females, n=2798 for males.

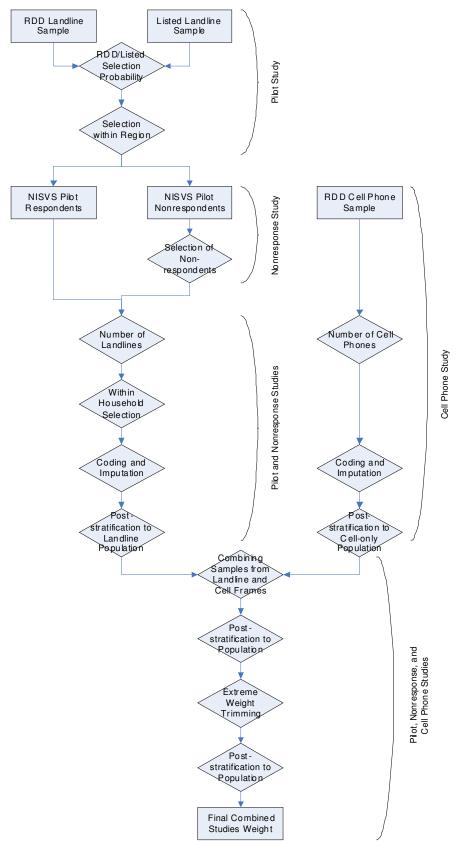Figure 1. Model for Poststratification Adjustment of Undercoverage and Nonresponse.

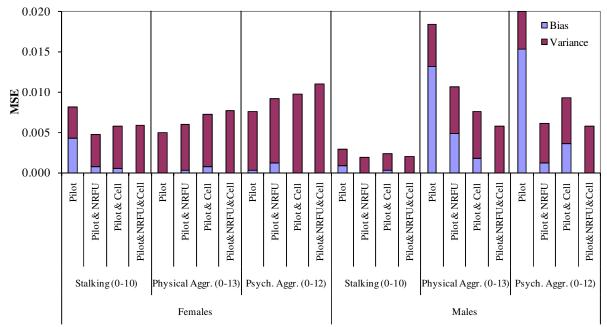Figure 2. Weighting Design for Combined NISVS Pilot, Nonresponse, and Cell Phone Studies.

Figure 3. Mean Square Error for Key Survey Measures Based on NISVS: (1) Pilot Study, (2) Pilot and Nonresponse (NRFU) Studies, (3) Pilot and Cell Phone Studies, and (4) Pilot, Nonresponse, and Cell Phone Studies, Weighted for Selection Probabilities, and by Final Poststratified Weights.

Appendix

Stalking

I'm going to ask you some questions about harassing and unwanted behaviors you may have experienced since you turned 18 that may have made you feel anxious or frightened. In answering, please think about anyone who may have done these things to you, including family members, people you knew, or strangers.

If you've worked as a law enforcement officer or served in the military, please do not include any situations that happened while you were on the job.

Since your 18th birthday has anyone made you feel anxious or frightened by repeatedly…
(code yes = 1, no = 2, don't know = 98, or refused = 99 for each question)
  1   following you or spying on you?
  2   sending you unwanted letters or written correspondence?
  3   standing outside your home, workplace, school, or place of recreation?
Since your 18th birthday has anyone made you feel anxious or frightened by…
  4   leaving unwanted items for you to find?
  5   sending you unwanted presents?
  6   monitoring your mail, e-mail, or other types of written or verbal communication?
  7   vandalizing your property?
Other than a bill collector, sales person, or law enforcement, since your 18th birthday, has anyone made you feel anxious or frightened by repeatedly. . .
  8   showing up unexpectedly at places you were, even when he or she had no reason to be there?
  9   by making unwanted telephone calls to you?
  10   or by sending electronic communications or e-mail notes to you over the Internet?

Sexual Violence

Now I'm going to ask you some questions about unwanted sexual experiences you may have had since your 18th birthday. These questions may seem disturbing, but it is important I ask them this way so that you are clear about what I mean. The information you are providing will be kept private.

Unwanted sex includes putting a penis, finger, hand, or other object in your [if female, fill: vagina or] anus when you did not want this to happen to you. It also includes unwanted contact between the mouth and the penis, vagina, or anus. It includes times when you were forced or unable to consent because you were drunk or asleep, or because you thought you would be hurt or punished if you refused.

Having sex without your consent could have been with anyone, including a spouse, partner, dates, relative, acquaintance, or stranger.

Since your 18th birthday, has anyone ever had sex with you after you said or showed that you didn't want them to or without your consent?

1 Yes
2 No
98 (volunteered) Don't know
99 (volunteered) Refused

Since your 18th birthday, has anyone *tried to force you* to have sex after you said or showed that you didn't want them to or without your consent?

1 Yes
2 No
98 (volunteered) Don't know
99 (volunteered) Refused

Physical Aggression

The next series of questions asks about other experiences you may have had since your 18th birthday and [randomly assigned:] about different types of crimes and the harm they may cause/about health issues that people may have experienced/about personal relationships and quality of life.

Since your 18th birthday did anyone, including a spouse, partner, date, relative, acquaintance, or stranger. . .
(code yes = 1, no = 2, don't know = 98, or refused = 99 for each response)

1   throw something at you that could hurt you?
2   push, grab or shove you?
3   pull your hair?
4   slap or hit you?

Since your 18th birthday, did anyone . . .

5   kick or bite you?
6   choke or strangle you?
7   hit you with some object?
8   beat you up?

Since your 18th birthday, did anyone . . .

9   threaten you with a gun?
10   threaten you with a knife or other weapon besides a gun?
11   attempt to drown you?
12   use a gun on you?
13   use a knife or other weapon on you besides a gun?

Psychological Aggression

The next questions ask about experiences you may have had with any partner. By partner, I am referring to a spouse, ex-spouse, someone you've lived with romantically as a couple, or someone you've dated.

Since your 18th birthday, has anyone who's ever been your partner . . .

(code yes = 1, no = 2, don't know = 98, or refused = 99 for each response)

1   had a hard time seeing things from your point of view?
2   been jealous or possessive?
3   tried to provoke you?
4   tried to limit your contact with family or friends?

Since your 18th birthday, has anyone who's ever been your partner . . .

5   called you names or put you down in front of others?
6   made you feel inadequate on purpose?
7   shouted or sworn at you?
8   thrown objects or broken things when angry?

Since your 18th birthday, has anyone who's ever been your partner . . .

9   intentionally frightened you?
10   prevented you from knowing about or having access to the family   income even when you asked?
11   prevented you from working outside the home?
12   insisted on knowing who you were with at all times?