Social Science Computer Review

http://ssc.sagepub.com

An Evaluation of Nonresponse and Coverage Errors in a Prerecruited Probability Web Panel Survey

Sunghee Lee

Social Science Computer Review 2006; 24; 460

DOI: 10.1177/0894439306288085

Social Science Computer Review
Volume 24 Number 4
Winter 2006 460-475

10.1177/0894439306288085
http://ssc.sagepub.com
hosted at
http://online.sagepub.com

An Evaluation of Nonresponse and Coverage Errors in a Prerecruited Probability Web Panel Survey

Sunghee Lee
University of California, Los Angeles

This study examines nonresponse and coverage errors separately in a probability Web panel survey by applying traditional postsurvey adjustments. This was done by using variables whose estimates were obtainable at both the survey respondent and the full survey sample levels and whose values were known for both the full survey sample and the target population. Nonresponse error measured by the differences between the estimates from the respondents and the known full sample values was not found to be large, implying that nonresponse error in this Web survey data may not be critical. However, coverage properties of the full survey sample show some problems, and traditional postsurvey adjustments were limited in alleviating the unequal coverage of the survey sample. This coverage problem was more evident for the subpopulation-level estimates.

Keywords: *Web surveys; nonresponse error; coverage error; postsurvey adjustments*

The Web is becoming popular as a mode of survey data collection, not only in the survey practice but in academia (e.g., Ballard & Prine, 2002; Bosnjak & Tuten, 2003; Braithwaite, Emery, De Lusignan, & Sutton, 2003; Couper, 2000a; Couper, Tourangeau, Conrad, & Crawford, 2004; Harewood, Yacavone, Locke, & Wiersema, 2001; Jones & Pitt, 1999; Schoen & Faas, 2005). Surveys can be conducted on the Web at any time in any place, with many types of graphics and multimedia features at virtually no cost. These advantages have attracted an enormous amount of attention from survey practitioners, which is directly related to the growth of Web surveys.

In spite of their popularity, the quality of Web surveys for scientific data collection is open to discussion. Whether the data collected from a set of Web survey respondents can be used to make inferences about the general population raises a basic statistical question because Web surveys mistakenly exclude non-Web users. Although survey organizations may hope that their Web surveys represent the general population (e.g., `http://www.insightexpress.com/audiences/methodology.asp`), statistical properties of most Web surveys are likely to deviate from those in traditional surveys. Generalizability

Author's Note: The author wishes to acknowledge supervision by Dr. Richard Valliant, Dr. Robert Mislevy, and Dr. Carla Maffeo; financial support from Mathematica Policy Research, Inc. with their 2003 summer fellowship; and permission from the 2002 practicum class at the Joint Program in Survey Methodology to use the survey data.

may not be an issue for some special populations where all persons or households have Web access (e.g., employees of a large corporation).

Prerecruited Probability Web Panel Surveys

There are several protocols from which to choose when conducting Web surveys (for the taxonomy of Web surveys, refer to Couper, 2000b; Manfreda, 2001; Schonlau, Fricker, & Elliott, 2002). One distinctive method described in Huggins and Eyerman (2001) and used in Thaliji, Langer, Pulliam, and Wiebe (2002), Berrens, Bohara, Jenkins-Smith, Silva, and Weimer (2003), Bandilla, Bosnjak, Altdorfer, and Lohmann (2004), and Smith, Li, and Pulliam (2005) uses prerecruited probability-based panel for Web surveys (see Figure 1 for the illustration). This prerecruited probability Web panel survey method is currently practiced by Knowledge Networks (KN) in the United States. KN recruits a controlled panel via random digit dialing (RDD) and equips the entire panel with a Web-accessing medium (Web TV), regardless of its Web usage status prior to the recruitment. During the first Web survey, the panel members take a panel profile survey, which collects a range of background information. The idea behind the profile survey is that for any subsequent surveys, the profile data are used to build sampling frames for any given subsequent survey samples. In addition, the data are available both for respondents and nonrespondents of actual surveys. This provides valuable opportunities for employing postsurvey adjustments that compensate for nonresponse and coverage errors.

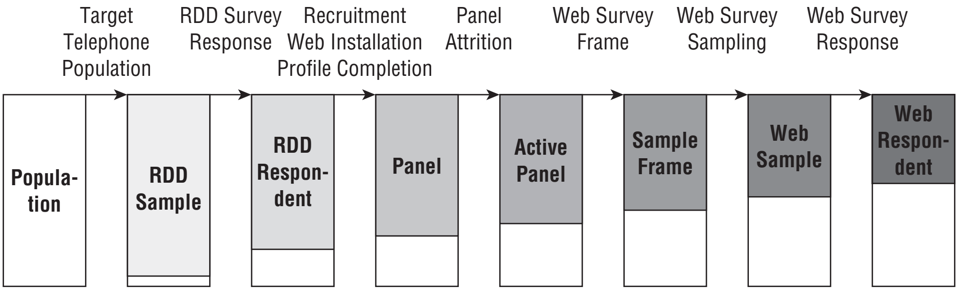
Ideally, the recruited Web panel described above represents the population of households or persons who have telephones because the panel members have a known probability of selection into the Web panel, and the samples drawn from the panel also have a known probability of selection into the sample. This protocol may diminish unequal coverage and nonprobabilistic sampling problems, which are inherent in other types of Web surveys. Although the calculability of the selection probability makes this type of Web survey the most scientific among existing Web surveys, there are significant complications.

As shown in part in Figure 1 and discussed briefly above, potential respondents go through four general stages before being sampled to participate in surveys: (a) initial RDD panel recruitment, (b) Web device installation, (c) profile survey completion, and (d) post-profile panel retention. All these stages, and actual survey participation, are susceptible to some type of loss in the potential respondent pool. The coverage and nonresponse errors are intertwined in this protocol.

Postsurvey Adjustment for Web Surveys

Traditional postsurvey adjustments, such as poststratification, are used as a one-shot remedy for both errors in practice. The application of these adjustments implicitly assumes that the error mechanism is ignorable in the sense used by Little and Rubin (1987). Because this Web survey employs a multistep protocol not found in other surveys, it may not be reasonable to assume this ignorability. Therefore, traditional adjustments may be ineffective in

Figure 1
Protocol of Prerecruited Probability Web Panel Surveys



Note: Figure not drawn to scale.
*$p < .05$.

correcting for the errors. Moreover, because these errors are adjusted simultaneously, the respective error evaluation is difficult to disentangle. One study by Vehovar and Manfreda (1999) examined the effect of poststratification for a Web survey. However, its findings are somewhat limited because the study sample was considered self-selected because of ambiguous eligibility of the units in the sample frame. The comparison standards came from a telephone survey, which may not be a reliable source for adjustment in that it is also subject to coverage and nonresponse errors.

This study attempts to evaluate the magnitude of nonresponse and coverage errors in a prerecruited probability Web panel survey. This is possible only for this particular type of Web survey because it has profile data collected prior to the actual surveys. The profile data allow a nonresponse error assessment by comparing survey respondents and nonrespondents. If reliable population estimates on profile characteristics are obtainable externally, the magnitude of coverage error may be studied by comparing the target population and the full sample that includes respondents and nonrespondents. This makes it possible to carry out a separate examination on nonresponse and coverage error. In addition, having auxiliary information from profile data makes it possible to employ different statistical adjustments for the errors. This may shed light on how these adjustments are applied and what may be expected from the adjustments.

The following sections provide a detailed description of the data sources and the variables used in the adjustments. First, nonresponse properties will be evaluated using the full sample that includes both respondents and nonrespondents as the sources of true values. Two adjustment approaches, ratio-raking and multiple imputation, will be applied to correct for nonresponse error. Unadjusted and two types of adjusted estimates from the respondent data will be compared with the true values. Second, coverage error will be examined with the assumption that population estimates from a large government survey are true. Ratio-raking will be used to compensate for coverage error. The deviations of unadjusted and adjusted full sample estimates from the true values will be examined as coverage error. The last section will summarize findings and raise considerations for future research.

Data Source

This study entails a two-stage adjustment requiring three types of data sets: (a) at the respondent level, (b) at the full sample level, and (c) at the population level. The first two data sets will come from a KN' Web survey and the last from the Current Population Survey (CPS).

Web Survey Data

Web survey data came from the 2002 Survey Practicum class at the Joint Program in Survey Methodology. The Web survey was conducted by KN through its prerecruited Web panel method from August to November 2002. To build the sampling frame, KN first selected all adolescents between the ages of 14 and 19 in the panel living with at least one parental figure (i.e., parent, grandparent, or guardian). Using this frame, KN drew a random sample of 2,501 households with a parental figure and his or her adolescent. Because later comparisons will be made between the Web survey and the CPS data and the closest possible teenage category identifiable in the CPS was 14 to 17, households with 18- and 19-year-olds were dropped from the analysis to make the two stages of error compensation comparable. This decreased the full sample size to 1,700. Among the sampled units, 978 households completed the Web survey, resulting in a completion rate of 57.4%. To qualify as a complete case, both the parent and the teen were to complete the survey. This might have played a negative role in the completion rate. The final response rate cumulated all nonresponse and noncooperation arising in the pre-Web survey stages and the survey completion and the parental consent for the teen's involvement. After incorporating the nominal completion rate (57.4%) with the panel recruitment rate (36.0%), the Web TV connectability rate (67.0%), the profile completion rate (98.0%), the postprofile survey retention rate (47.0%), and the parental consent rate for teen's participation (86.0%), the cumulative final response rate became 5.5%.

Two data sets were created by combining the profile data and the Web survey response status. Recall that the panel profile data are available for both Web survey respondents and nonrespondents. The first data set was the profile data of the full survey sample units ($n = 1{,}700$). Second, the respondent data ($r = 978$) were constructed by keeping Web survey respondents in the profile data. The teen profile data were not used because of a large amount of items missing. Thus, the study target population was the parents living with at least one adolescent between the ages of 14 and 17.

CPS Data

The population estimates come from the September 2001 CPS. This particular wave of CPS contained the Computer and Internet Use Supplement, which collected information about Internet and computer usage by the eligible members of the sampled households (for methodological documents, refer to `http://www.bls.census.gov/cps/computer/2001/smethdocz.htm`). When restricting the September 2001 CPS sample to the scope of the target population defined above, the eligible sample size decreased from 143,300 to 11,290.

The CPS target population and its samples included persons living in nontelephone households, whereas KN Web survey started with the telephone population. This is a source of noncomparability between the coverage of the KN data set and the CPS, despite the fact that only 3.5% of persons in the United States were found to live in nontelephone households based on the 2001 CPS data. Nonetheless, Web survey organizations often claim that their surveys represent the full population living in both telephone and nontelephone households. To incorporate this error in the examination, we have used estimates based on both telephone and nontelephone households in CPS.

Variables of Interest and Covariates

All variables used in the analyses are available from both data sources. There are four dependent variables to be estimated: (a) number of owned computers in the household (none, one, or more), (b) prior Web usage experience (no, yes), (c) employment status (unemployed, employed), and (d) household size (number of household members), denoted as y_1, y_2, y_3, and y_4. Error adjustment will be made with respect to the following covariates: age (20-40, 41-45, 46-50, 51 or older), education (less than high school, high school, some college, college or above), *ethnicity* (White non-Hispanics, Black non-Hispanics, other non-Hispanics, Hispanics), region (Northeast, Midwest, South, West), and gender (male, female), denoted as $x_1, \ldots, x_5$ in ratio-raking adjustment or $x_1, \ldots, x_9$ in multiple imputation. In multiple imputation, x_1, x_2, and x_9 are assigned to age, education, and gender, as the first two are treated as continuous and the last as dichotomous. Ethnicity and region are polytomous variables with four categories, which require three binary response variables each. Thus, x_3, x_4, x_5 are assigned to ethnicity and x_6, x_7, x_8 to region. These covariates are selected to mimic KN's ratio-raking procedure. Although KN's original adjustment includes one additional covariate, household income, this was excluded because there were many missing cases in the CPS.

The covariates served another function: All categories in these variables were the units of subgroup estimation. The reasons for estimating at the subgroup level are twofold. First, existing studies make comparisons between Web surveys and traditional surveys typically at the total population level. Postsurvey adjustments may correct the errors in the total population estimates but not necessarily in the subgroup estimates. The second reason reflects the realistic analytical interests: Analyses are often performed at the subgroup level to obtain more insightful conclusions than those simply at the population level. For these reasons, this study includes the subgroup-level estimation.

Nonresponse Error Adjustment

Nonresponse error examined in this section focuses on the noncompletion of the actual Web survey, not the cumulative nonresponse for the entire panel. The full survey sample is drawn from the frame that represents the target population and therefore is treated as a simple random sample (SRS) of the target population. The estimates from the full survey sample are assumed to be true values because they contain both respondents and nonrespondents.

Consequently, deviations from the true values in the estimates from the respondents, whether adjusted or unadjusted, are regarded as nonresponse error. This corresponds to the differences between the last two boxes in Figure 1. Comparisons between the two provide the degree in nonresponse error reduction because of adjustments.

Table 1 provides the distribution of estimates of the variables of interest from both the unadjusted respondents and the full sample. Because full sample estimates (y_F) are the benchmarks of the nonresponse adjustment, the deviations of the unadjusted respondent estimates ($\hat{y}_{UR}$) are treated as the initial biases. Standard errors of these deviations are calculated as follows, where y_F and $\hat{y}_{UR}$ are not independent:

$$y_F = \frac{r}{n}\hat{y}_{UR} + \frac{n-r}{n}\hat{y}_{UN},$$

and, therefore,

$$\mathrm{se}(\hat{y}_{UR} - y_F) = \sqrt{\mathrm{var}\left[\frac{n-r}{n}(\hat{y}_{UR} - \hat{y}_{UN})\right]} = \left(\frac{n-r}{n}\right)\sqrt{[\mathrm{var}(\hat{y}_{UR}) + \mathrm{var}(\hat{y}_{UN})]}, \tag{1}$$

where there are n units in the full sample and r respondents, and $\hat{y}_{UN}$ is the estimate from nonrespondents, assuming $\mathrm{cov}(\hat{y}_{UR}, \hat{y}_{UN}) = 0$. The variability of y_F, $\hat{y}_{UR}$, and $\hat{y}_{UN}$ and is computed using variance formula for SRS because SRS was assumed previously. The standard error calculation in Equation 1 is possible because information on nonrespondents (i.e., $\hat{y}_{UN}$) is available from the profile data set. Contrary to the initial speculation, the deviations of unadjusted statistics in Table 1 are small. The deviations for computer ownership and household size, although statistically significant, do not appear to be meaningful.

The distributions of the five covariates are shown in Figure 2. The two comparison groups are distributed almost identically. Based on a χ^2 test for equality of the two groups, only ethnicity shows significant differences. There are more Whites but fewer Blacks and Hispanics among the respondents than in the full sample, but these gaps are not large. Almost perfect comparability between the two data sources shown in Table 1 and Figure 2 may suggest that the respondents are representative of the full sample (i.e., nonresponse occurred completely at random). One important implication of the identical covariate distributions in Figure 2 is that the statistical adjustments using these covariates will not correct for any biases in variables not examined in this article because the benchmark distributions are the same as the unadjusted ones.

Sample Level Ratio-Raking Adjustment

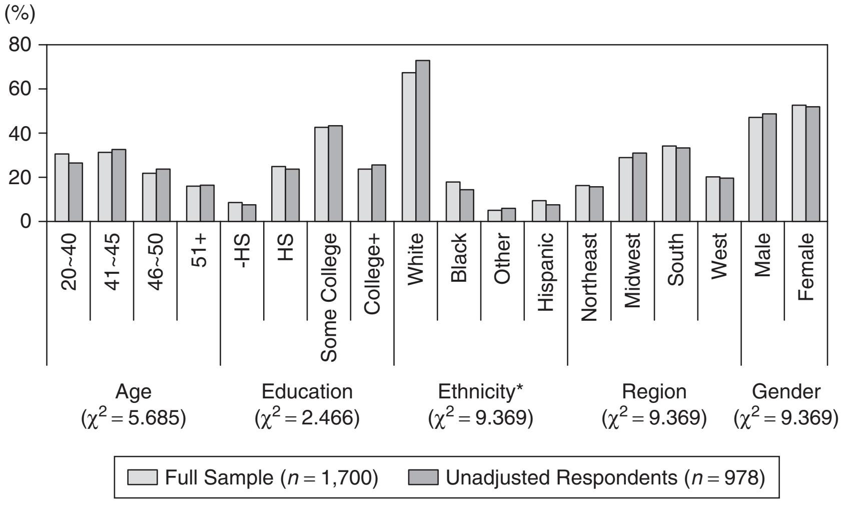
Ratio-raking adjustment is a popular modification of poststratification, which follows the iterative steps described in Deming and Stephan (1940). Unlike the cell weighting method, ratio-raking controls the marginal distributions of covariates, which decreases difficulties arising because of zero observations or unknown benchmarks in cross-classified cells. With the marginal counts of the five covariates from the full sample, ratio-raking was performed

Table 1
Initial Nonresponse Deviation Measured by Estimates of Computer Ownership, Web Experience, Unemployment Rates, and Household Size and Their Standard Errors From the Unweighted Full Survey Sample and Unadjusted Survey Respondents

	Full Sample[a]		Unadjusted Respondents[b]		
	Estimate (y_F)	*SE*	Estimate ($\hat{y}_{UR}$)	*SE*	Deviation[c]
Computer ownership (%)	79.6	0.98	81.4	1.25	1.8*
Prior Web experience (%)	72.0	1.09	71.2	1.45	–0.8
Unemployment (%)	3.9	0.47	4.1	0.63	0.2
Household size	4.2	0.03	4.1	0.04	–0.1**

a. $n = 1{,}700$.
b. $n = 978$.
c. Deviation = $\hat{y}_{UR} - y_F$.
*$p < .05$. **$p < .01$. ***$p < .001$.

Figure 2
Distributions of Covariates From the Full Survey Sample and the Unadjusted Respondents



using WesVar 4.0. Initial design weights were adjusted for sample-level nonresponse through ratio-raking, and the adjusted weights were used in the estimation.

Cell-based poststratification was not use in this study because of its infeasibility. Cross-classification of the four adjustment variables required 128 (= 4 × 4 × 4 × 2) adjustment cells, and the small sample size was highly likely to result in many zero cells.

Multiple Imputation

Multiple imputation was first suggested by Rubin (1978) for item nonresponse. Although this section does not examine item nonresponse, unit nonresponse in this Web survey may be regarded as item nonresponse in that there is sufficient background information for survey respondents and nonrespondents. Through multiple imputation, values for the missing observations could be imputed by specifying an explicit model that produces posterior predictive distributions of the missing data, conditional on the distribution of the observed data.

The models for the three dichotomous variables in this study (y_1, y_2, y_3) are specified in the following way:

$$y_i \sim \text{Bernoulli}\,(\theta_i), \quad \text{logit}\,(\theta_i) = \alpha_i + \sum_{j=1}^{9} \beta_{ij} x_j + \varepsilon_i,$$

where α_i, β_{ij} ~ Normal(0,1) and ε_is are random errors with a mean of 0 for $j = 1, \ldots, 9$ and $i = 1, 2, 3$. Because dichotomous variables are categorical, they are modeled as having Bernoulli distributions determined by the parameters, θ_i. The θ_is are predicted by the covariates known for both respondents and nonrespondents. The model parameters, α_is and β_is, have normal prior distributions, with mean 0 and variance 1. Similarly, the continuous variable, y_4, is modeled as follows:

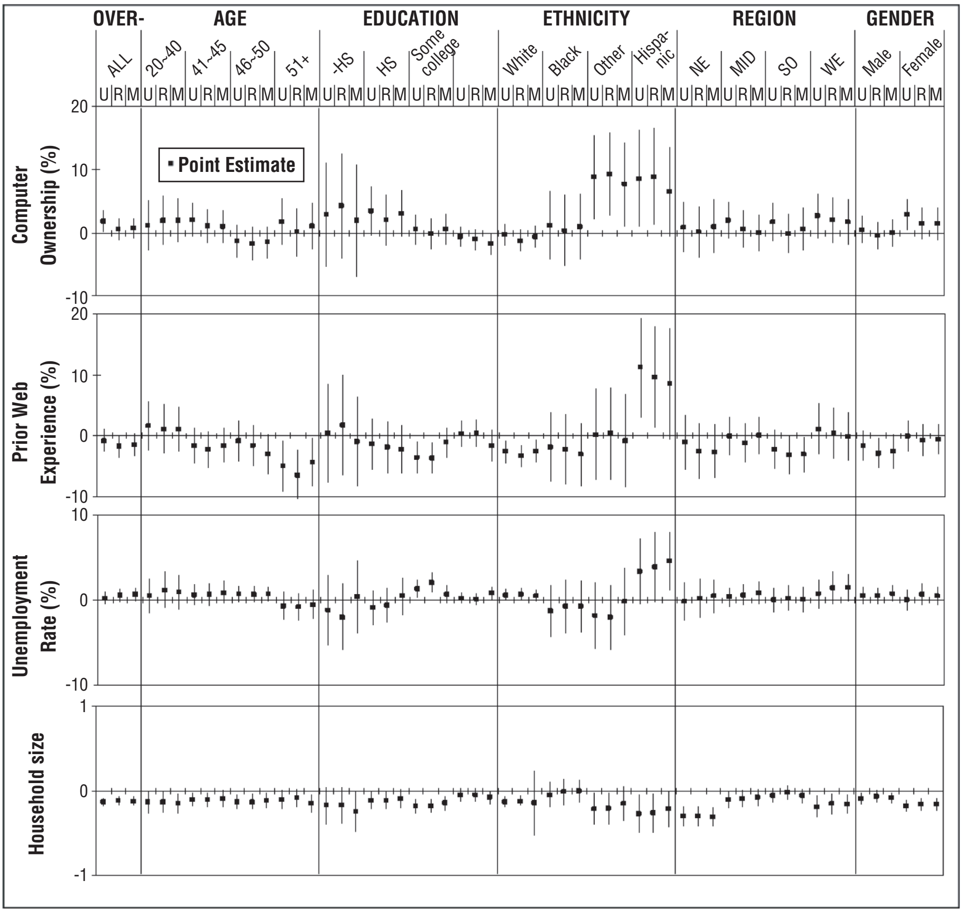
$$y_4 \sim \text{Normal}\,(\theta_4, \upsilon), \quad \theta_4 = \alpha_4 + \sum_{j=1}^{9} \beta_{4j} x_j + \varepsilon_4,$$

where θ_4 is the prior of y_4 predicted in a linear function of the same series of covariates, using prior information, υ ~ Gamma(0.5, 1), α_4, β_{4j} ~ Normal(0,1) for $j = 1, \ldots, 9$, and ε_4, a random error. Note that the same set of covariates was utilized here as in the ratio-raking procedure. The imputation model fit and modification are not considered here because the purpose of this article is to compare the two adjustment methods, given the same auxiliary information.

Winbugs 1.4 (Spiegelhalter, Thomas, & Best, 1999) was used for the multiple imputation. The prior distributions of the model parameters were updated by the profile data. Missing values were predicted by the updated values of model parameters. Each missing value for each nonrespondent was imputed using five different initial values (or chains), which resulted in five different predicted values. Each model stated above was run in 10,000 iterations using the Markov Chain Monte Carlo method (details in Gelman, Carlin, Stern, & Rubin, 1995, chap. 1), among which the first 2,999 iterations were regarded as burn-in to use samples that produced convergent statistics among different initial values. For each chain, imputed values for nonrespondents were combined with observed values from respondents. Estimates were calculated following the procedure in Rubin (1987, chap. 3).

Figure 3 displays the 95% confidence intervals for the deviation of unadjusted (U), ratio-raking adjusted (R), and multiple imputation (M) estimates ($\hat{y}_{\text{UR}}$, $\hat{y}_{\text{RR}}$, and, $\hat{y}_{\text{MI}}$, respectively) from the true full survey sample values for the whole sample and different subpopulations formed by the covariates. Estimation for standard errors followed Equation 1, replacing

Figure 3
95% Confidence Intervals of Deviations Between Different Respondent Estimates and Full Survey Sample Estimates



Note: Different respondent estimates $n = 978$; full survey sample estimates $n = 1{,}700$. U = unadjusted estimates; R = ratio-raking adjusted estimates; M = multiple imputation estimates from the respondent data.
$*p < .05$. $**p < .01$. $***p < .001$.

var($\hat{y}_{UR}$) with var($\hat{y}_{RR}$) and var($\hat{y}_{MI}$) for adjusted estimates. More specifically, var($\hat{y}_{RR}$) was obtained from WesVar 4.0 and var($\hat{y}_{MI}$) using the procedure described in Rubin (1987, chap. 3). Confidence intervals placed around zero indicate that those deviations are not statistically significant, leading to a conclusion that the nonresponse error is negligible.

Most deviations in Figure 3 are not significant for the total population. Household size appears to have a statistically significant deviation that is not meaningful. When examined by subgroup, only the estimates for different ethnic subgroups are likely to diverge from the benchmarks. Overall, U, R, and M estimates do not differ from one another, and the performance of the two adjustments is equivalent. This result is not surprising when recalling that the preliminary analysis showed that the unadjusted estimates for all variables match the full sample values well. Nonresponse adjustments on these variables might have been unnecessary after all.

Coverage Error Adjustment

Coverage error is defined as the failure to include all eligible people in the sampling frame. Because the Web survey sample frames are created from the prerecruited panel members, coverage error pertinent to this type of Web survey may be examined by comparing the full sample, which, by theory, represents the frame, and the target population. An alternative view of coverage error may also include coverage and response or cooperation status in the presurvey stages shown in Figure 1, through which potentially eligible units may become lost systematically. Web survey frames built only on the active panel members may be biased to begin with, unlike traditional surveys, in which full samples represent the target populations through their sampling frames. Although it may be debatable whether to view this error as a coverage or coverage and nonresponse problem, the answer may depend on whether to focus on the RDD frame or the Web survey frame. If the latter, the difference between the population and the sample frame will compose coverage error. Because the Web sample mirrors the frame perfectly, the difference between the population and the Web sample becomes a reliable proxy measure for coverage error. (Please refer to the first and seventh boxes in Figure 1.) This study will place an emphasis on the Web survey frame and consider the difference between the population and the full sample (i.e., frame) as coverage error.

Table 2 provides estimates from the CPS and the unadjusted full sample data (y_{CPS} and $\hat{y}_{\mathrm{UF}}$, respectively) and the differences between the two. The population values used for comparison were calculated by applying the final weights provided in the CPS public use data and will be assumed as true values. The full Web survey sample estimates are calculated by applying the base design weights to the 1,700 cases provided by KN. Similar to nonresponse adjustment, the discrepancies between the full sample estimates and the true values will be considered as coverage error.

Direct calculation of variances for y_{CPS} is not possible because information about the design variables is unavailable in the public use CPS data. Instead, the following ad hoc approach is used for standard error of CPS estimates:

$$\mathrm{se}(y_{\mathrm{CPS}}) = k \times \mathrm{se}(\hat{y}_{\mathrm{UF}}), \tag{2}$$

where $SE(\hat{y}_{\mathrm{UF}})$ is the standard error of the unadjusted full sample estimate and k is a constant based on the ratio of the full Web survey sample size to the CPS size. For the same reason, the standard error of the deviation is calculated as in Table 2:

$$\begin{aligned} \mathrm{se}(y_{\mathrm{CPS}} - \hat{y}_{\mathrm{FS}}) &= \sqrt{\mathrm{var}(Y_{\mathrm{CPS}}) + \mathrm{var}(\hat{y}_{\mathrm{UF}})} \\ &= \sqrt{k\,\mathrm{var}(Y_{\mathrm{CPS}}) + \mathrm{var}(\hat{y}_{\mathrm{UF}})} \\ &= \sqrt{k+1} \times \mathrm{se}(\hat{y}_{\mathrm{UF}}), \end{aligned} \tag{3}$$

Table 2
Initial Coverage Deviation Measured by Estimates of Computer Ownership, Web Experience, Unemployment Rates, and Household Size and Their Standard Errors From the Current Population Survey and Unadjusted Full Survey Sample

	Current Population Survey[a]		Unadjusted Full Sample[b]		
	Estimate (y_{CPS})	*SE*	Estimate ($\hat{y}_{\text{UF}}$)[c]	*SE*	Deviation[d]
Computer ownership (%)	80.9	0.57	77.5	1.29	–3.4**
Prior Web experience (%)	65.8	0.62	70.9	1.39	5.1***
Unemployment (%)	2.6	0.19	4.1	0.59	1.5**
Household size	4.3	0.02	4.2	0.04	–0.2***

a. $n = 11,290$.
b. $n = 1,700$.
c. Estimates reflect the base design weights.
d. Deviation = $y_{\text{CPS}} - \hat{y}_{\text{UF}}$.
*$p < .05$. **$p < .01$. ***$p < .001$.

It should be noted that Equations 2 and 3 follow a crude approach to derive variance estimates because they assume that the variability of an estimate is a function of sample sizes.

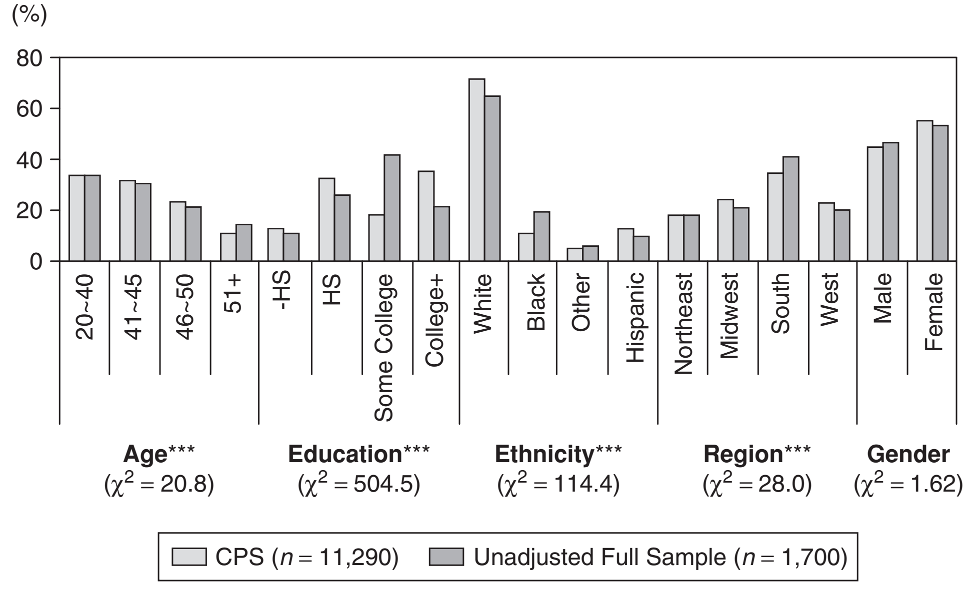
Unlike the nonresponse section, Table 2 suggests potential coverage problems in that the comparisons between the true values and the initial estimates reveal significant deviations. Computer ownership and prior Web experience show rather meaningful differences. People in this Web sample are less likely to own computers but more likely to have Web experience. Figure 4 illustrates remarkable inconsistencies in covariates between two sources of data: Four out of five covariates show significant differences. This is more obvious for education and ethnicity, as the Web sample contains disproportionately more people with some college education and more Blacks compared to the population. It becomes imperative to remedy these discrepancies through adjustments.

Coverage properties are examined by replicating the same procedure used in ratio-raking for nonresponse. The adjustment weights were computed by projecting the covariate marginal distributions in the Web sample to those in the CPS through ratio-raking. This procedure adjusted the base design weights in the Web sample so that they may correct for coverage error, and these weights were used in estimation using WesVar 4.0.

Multiple imputation was not used for coverage error compensation. For imputation, we need a full frame of the target population. Although we assumed that the CPS represents the population reasonably better than the Web survey sample, it does not provide the full frame in that it is a realized sample of the target population. To carry out model-based imputation, we also need auxiliary information on all units that were not sampled. Therefore, it was impossible to apply multiple imputation for coverage adjustment.

The 95% confidence intervals of the deviations of the unadjusted (U) and ratio-raking adjusted (R) estimates from the population values are shown in Figure 5. If the ratio-raking procedure is effective in reducing coverage bias, Figure 5 would show more confidence

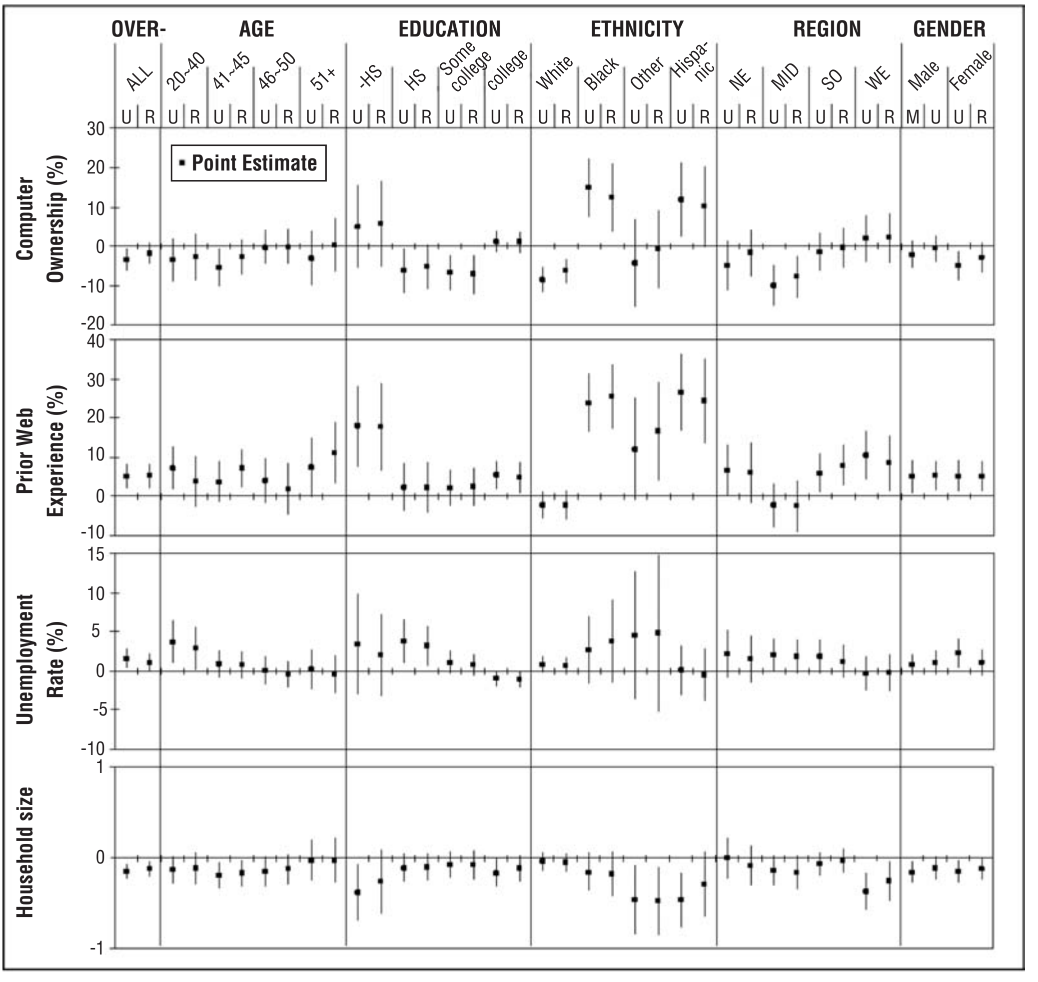
Figure 4
Distributions of Covariates From the Current Population Survey and the Unadjusted Full Survey Sample



intervals of the deviations for the R estimates containing zero than those of the U estimates. The ratio-raking adjustment seems to make improvements. However, the adjusted estimates are still closer to the unadjusted ones than they are to the benchmarks. Significant deviations remain even after the adjustment, and they become more conspicuous for the subgroup estimates. Most prevalent discrepancies are found for subgroups under education and ethnicity. This coincides with the divergence observed in Figure 4. Although the initial deviations were hoped to be corrected by ratio-raking, estimates for subgroups formed by these two covariates are still far from the true values.

Figure 5 shows that persons with a high school education or less reported having prior Web experience at a far higher rate in the Web survey than did their counterparts in the CPS. In fact, its percentage in the Web sample is about 20 percentage points higher than that in the CPS. One explanation may be a misunderstanding among persons in the Web sample of what *Web* means. Another explanation may be that people with lower education in the Web sample might have been more likely to be unemployed before joining the KN Web survey panel and to own fewer computers but to have had experience with the Internet at a higher rate than did their counterpart in the population. These people might have had more free time, thus more potential opportunities to access the Web but less ability to afford computers. This might have made their reactions to an offer of free Web TV and free Web access more positive than persons with higher education, inducing them to stay active on the panel to maintain the free Web access.

Figure 5
95% Confidence Intervals of Deviations Between Full Survey Sample Estimates and Current Population Survey Estimates



Note: Full survey sample estimates n = 1,700; Current Population Survey estimates n = 11,290. U = unadjusted estimates; R = ratio-raking adjusted estimates; M = multiple imputation estimates from the respondent data.

The discrepancies in computer ownership and Web usage by ethnicity warrant attention. The Web sample includes higher proportions of technology-savvy Blacks and Hispanics than does the CPS. Both unadjusted and adjusted Web sample estimates of the computer ownership for Blacks and Hispanics are 10 percentage points higher than the population values. These ethnic groups in the Web sample have higher levels of Web experience than their counterpart in the population—the full sample overestimates the prior Web experience by far more than 20 percentage points. Interestingly, Whites in the Web sample are somehow less technologically experienced than those in the population, as measured by computer ownership and Web experience. This suggests that the Web sampling frame coverage may differ systematically from the population with respect to ethnicity. Ratio-raking does not appear to be a sufficient solution.

Discussion

This study documents one of the first examinations of statistical adjustment approaches for Web surveys. The respondents in this particular Web survey seemed to well represent the full sample, even though the completion rate was fairly low. Consequently, the sample-level nonresponse adjustment was not even necessary for at least the variables examined in the study. This is similar to the recent findings about nonresponse (e.g., Curtin, Presser, & Singer, 2000; Keeter, Kohut, Miller, Groves, & Presser, 2000; Merkle & Edelman, 2002). The covariate distributions for the respondents and the full sample were very close. This may imply that adjustments based on these characteristics may not improve the survey estimates.

The coverage of the Web sample frame did not appear to be sufficient. Estimates for the subgroups whose population and sample covariate distributions showed inconsistencies tended to deviate significantly from the population values. Traditional postsurvey adjustment using ratio-raking had a limited effect in correcting for these deviations. Thus, this result failed to support the assumption of ignorability of the underlying coverage mechanism in ratio-raking.

This study has certain limitations to consider in interpreting the results. First, they apply only to this particular type of Web survey. Other Web surveys using different protocols rely on convenience or volunteer samples and, thus, may have completely different error structures. Second, coverage and nonresponse errors are properties of a statistic, not of a survey. The selection of statistics examined in this study was based solely on their availability at the all respondent, the full sample, and the population levels. Other statistics may provide different implications on nonresponse and coverage errors. Third, the target population was very specific—parent figures with at least one teen household member. This group of people compared with other populations may produce unique nonresponse and coverage properties. Findings in this study can serve as a window to those error mechanisms, and their generalization requires caution.

This study found that coverage errors of this Web panel survey were more severe than nonresponse errors conditional on the profile data. However, the full survey sample already included multiple stages of nonresponse prior to the actual Web survey, which were captured under the coverage error examination. Coverage errors from nonresponse or noncooperation in the procedures of recruiting and maintaining Web survey panel members may arise in a more systematic way than nonresponse in the actual survey. Further investigations on disentangling coverage and nonresponse mechanisms at each stage would be informative. If consistent evidence against ignorability of these errors is found, more innovative adjustment methods will be needed for sound inferences from this type of Web survey.

References

Ballard, C., & Prine, R. (2002). Citizen perceptions of community policing: Comparing Internet and mail survey responses. *Social Science Computer Review*, 20, 485-493.

Bandilla, W., Bosnjak, M., Altdorfer, P., & Lohmann, H. (2004, May). *Mode-effects in pre-recruited panels of full population*. Paper presented at the American Association for Public Opinion Research annual meeting, Phoenix, AZ.

Berrens, R. P., Bohara, A. K., Jenkins-Smith, H. C., Silva, C. L., & Weimer, D. L. (2003). The advent of Internet surveys for political research: Comparison of telephone and Internet samples, *Political Analysis*, *11*, 1-22.

Bosnjak, M., & Tuten, T. L. (2003). Prepaid and promised incentives in Web surveys—An experiment. *Social Science Computer Review*, *21*, 208-217.

Braithwaite, D., Emery, J., De Lusignan, S., & Sutton, S. (2003). Using the Internet to conduct surveys of health professionals: A valid alternative? *Family Practice*, *20*, 545-551.

Couper, M. P. (2000a). Usability evaluation of computer assisted survey instruments. *Social Science Computer Review*, *18*, 384-396.

Couper, M. P. (2000b). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, *64*, 464-494.

Couper, M. P., Tourangeau, R., Conrad, F., & Crawford, S. (2004). What they see is what we get: Response options for Web surveys. *Social Science Computer Review*, *22*, 111-127.

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the Index of Consumer Sentiment. *Public Opinion Quarterly*, *64*, 413-428.

Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, *11*, 427-444.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. B. (1995). *Bayesian data analysis.* Boca Raton, FL: Chapman and Hall.

Harewood, G. C., Yacavone, R. F., Locke, G. R., & Wiersema, M. J. (2001). Prospective comparison of endoscopy patient satisfaction surveys: E–mail versus standard mail versus telephone. *American Journal of Gastroenterology*, *96*, 3312-3317.

Huggins, V., & Eyerman, J. (2001, November). *Probability based Internet surveys: A synopsis of early methods and survey research results.* Paper presented at the Research Conference for the Federal Committee on Statistical Methodology, Arlington, UA.

Jones, R., & Pitt, N. (1999). Health surveys in the workplace: Comparison of postal, email and WWW methods. *Occupational Medicine*, *49*, 556-558.

Keeter, S., Kohut, A., Miller, C., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a large national telephone survey. *Public Opinion Quarterly*, *64*, 125-148.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: John Wiley.

Manfreda, K. L. (2001). *Web survey errors*. Unpublished doctoral dissertation, University of Ljubljana, Slovenia, Faculty of Social Science.

Merkle, D., & Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 243-58). New York: John Wile.

Rubin, D. A. (1978). Multiple imputation in sample surveys—A phenomenological Bayesian approach to nonresponse. In *Proceedings of the Section on Survey Research Methodology* (pp. 20-34). Alexandria, VA: American Statistical Association.

Rubin, D. A. (1987). *Multiple imputation for nonresponse in surveys.* New York: John Wiley.

Schoen, H., & Faas, T. (2005). When methodology interferes with substance: The difference of attitudes towards e-campaigning and e-voting in online and offline surveys. *Social Science Computer Review, 23*, 326-333.

Schonlau, M., Fricker, R. D., Jr., & Elliott, M. N. (2002). *Conducting research surveys via e-mail and the Web.* Santa Monica, CA: RAND.

Smith, T. W., Li, R. J., & Pulliam, P. (2005, May). *Comparing the Knowledge Networks Web-enabled panel and the in-person 2002 General Social Survey: Experiments with mode, format, and question wordings*. Paper presented at the American Association for Public Opinion Research annual meeting, Miami Beach, FL.

Spiegelhalter, D. J., Thomas, A., & Best, N. G. (1999). *WinBUGS Version 1.2 user manual.* Cambridge, UK: MRC Biostatistics Unit.

Thalji, L., Langer, M., Pulliam, P., & Wiebe, E. F. (2002, May). *Nonresponse error and mode effects in the Web-enabled survey on civic attitudes and behaviors after 9/11*. Paper presented at the American Association for Public Opinion Research annual meeting, St. Pete Beach, FL.

Vehovar, V., & Manfreda, K. L. (1999). Web surveys: Can the weighting solve the problem? In *Proceedings of the Section on Survey Research Methodology* (pp. 962-967). Alexandria, VA: American Statistical Association.

Sunghee Lee is a survey methodologist at the Center for Health Policy Research, University of California, Los Angeles. She can be reached at slee9@ucla.edu.