

ONLINE SUPPLEMENT TO:

Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with
Probability and Non-Probability Samples

David S. Yeager and Jon A. Krosnick

Stanford University

LinChiat Chang

Kantar Health

Harold S. Javitz

SRI International, Inc.

Matthew S. Levindusky

University of Pennsylvania

Alberto Simpser

University of Chicago

Rui Wang

Stanford University

August, 2009

Overview

This Online Appendix provides the following:

- Descriptions of the firms' methods for collecting data;
- All question wordings and response options;
- Benchmark data sources and calculations;
- Missing data techniques;
- All t-tests comparing the firms' average errors (Tables 1-3);
- The variability of accuracy across the seven telephone surveys, seven probability sample Internet surveys, and seven non-probability sample Internet surveys (Tables 4-6);
- Results obtained when dropping health status as a benchmark (Table 7);
- Results obtained when using weights provided by the firms (Table 8);
- Results obtained when capping weights (Table 9);
- T-tests assessing whether post-stratification improved each survey's average absolute error (Table 10)
- Targets used to build post-stratification weights (Table 11);
- The weighting program (Appendix 1);
- A description and copy of the bootstrapping procedure used for statistical testing (Appendix 1).

Methods

SAMPLES

Telephone survey. A national sample of 966 American adults was interviewed by telephone via Random Digit Dialing (AAPOR Response Rate 3 = 35.6%). For this survey, an RDD sample of 6,990 phone numbers was generated, and pre-notification letters were sent to

2,518 (36% of the sample for which addresses could be obtained) to notify them that an interviewer would be calling them. Up to 12 call attempts were made to each telephone number, and one refusal conversion was attempted for each number if needed. A total of 879 households with known addresses that had not been reached partway through the field period were mailed a different letter encouraging participation and offering \$10. Refusal conversion letters were also sent to 95 households that had initially refused to participate and for whom addresses were available, also offering \$10. Interviews were conducted between June 15, 2004, and November 7, 2004.

Probability sample Internet survey. The probability sample Internet survey firm maintains a panel of about 40,000 people, from which a sample of individuals are drawn for each survey. The full panel was recruited via RDD telephone calls. Before panel recruiting calls were made to these randomly generated telephone numbers, sampled households for which a valid postal address could be obtained via reverse listings (for about half of the RDD sample) were sent pre-notification letters. Of the households for which no postal address could be obtained, a randomly selected 70% were called for recruitment. Phone numbers were not called if they were located in areas where MSN-TV (which provided Internet access to potential panelists) did not offer service.

During the initial RDD recruitment calls, potential respondents were told they had been selected to participate in an “important national study.” People who had access to the Internet from home were invited to participate in surveys using their own equipment and were told that by doing so, they would earn points that could be redeemed for cash. Households without Internet access were offered equipment and Internet access at no cost in exchange for survey participation. Information was obtained on all members of each participating household,

including names, ages, and genders. Once a household had Internet access, all household members (ages 18 or over) were asked to complete profile surveys and were then invited to complete subsequent surveys via email invitations.

Each household member had his or her own e-mail account (with separate log-in names and passwords). For each survey, selected respondents were sent an e-mail inviting them to participate and providing a hyperlink that took them directly to the questionnaire.

Panel members were asked to complete about one questionnaire per week, usually not exceeding 15 minutes. Panelists were entered into regular prize drawings in return for staying in the panel. When panel members were asked to complete a longer questionnaire, they were given a week off or offered some other form of incentive or compensation. Respondents could complete each questionnaire whenever they liked, and people could stop before completing a questionnaire and return to it later. Respondents who failed to complete eight consecutive questionnaires were dropped from the panel, and the Internet access equipment was removed from their homes if one had been placed there.

Some members of the probability sample Internet survey firm's panel were randomly selected to be invited to complete our survey, with unequal probabilities. These probabilities were directly proportional to the number of adults living in the panel member's household and were inversely proportional to the number of non-business telephone landlines that could reach the panel member's household at the time of recruitment. During recruitment, telephone numbers were over-sampled if a mailing address could be obtained for them (to allow mailing an advance letter before recruitment) and if they were located in areas served by MSN-TV or were located in areas with high densities of racial minorities, so these households had a proportionally lower probability of selection to participate in our survey. Some panelists were recruited during

a pilot phase that took place in only a few metropolitan areas, and these panelists had proportionally lower probabilities of selection than other panelists recruited later. Panelists who had completed more than the allowed number of surveys in a month were assigned a selection probability of zero. The probability of selection was also adjusted to eliminate discrepancies between the full panel and the population in terms of sex, race, age, education, and Census region (as gauged by comparison with the Current Population Survey). Therefore, no additional weighting was needed to correct for unequal probabilities of selection during the recruitment phase of building the panel.

A total of 1,533 panelists were drawn from the panel and were invited to complete our survey on June 18, 2004; 1,137 did so before the survey closed on July 2, 2004, yielding a completion rate of 74.2%. Taking into account non-response during the initial recruitment phone calls that yielded our sample, as well as the rate at which recruited panelists completed the profile survey and eventually completed our survey, the AAPOR Cumulative Response Rate 1 (Callegaro & DiSogra, 2008) was 15.3%. Respondents were offered points to be redeemed for cash in return for completing our survey.

Non-probability sample Internet survey 1. The 1.4 million members of non-probability Internet survey firm 1's panel were recruited in several ways. Initially, RDD phone calls were made to invite American adults to sign up to receive email invitations to participate in surveys, yielding about 2,500 panel members. Additional phone calls were made to professionals working in the information technology sector who were on professional lists; these calls yielded about 2,500 more panel members.

These initial 5,000 panel members were offered a chance to win cash or gift certificates if they would refer friends or family who might sign up to complete online surveys. Referred panel

members were offered the same incentives to refer other people. Panel members received a chance to win a prize each time they completed a survey, when someone they referred completed a survey, and each time the referral's referral completed a survey. Panel members were also recruited through online ads (on the firm's own website, news sites, blogs or search engines) and through emails from businesses or non-profit organizations with which the panelist had an affiliation.

3,249 members of a stratified random sample of non-probability Internet survey firm 1's panel received an email inviting them to participate in our survey, and 1,101 did so (34%). Probability of selection within demographic strata was specified so that invitees resembled the distribution of gender, age, and Census region in the U.S. adult population, as estimated by the 2000 Census. No quotas were used to restrict who completed the survey. 80% of the invitees had been referred, and the remaining 20% joined the panel from other sites, blogs or search engines that did not use referrals. Respondents were entered into a sweepstakes in return for completing our survey. Email invitations were sent on February 21st, 2005, and no data were collected after February 23rd, 2005.

Non-probability sample Internet survey 2. The 5 million members of non-probability Internet survey firm 2's panel were led to the firm's website through banner ads and other Internet advertisements. After signing up on firm 2's website, panelists completed a profile survey and became active panel members.

A stratified random sample of 11,530 veteran panel members (who had previously completed at least one survey for the firm) were invited to complete our survey, and 1,841 did so (16%). 8,520 panelists were randomly selected within gender, age, and region strata to match the over-18 U.S. population estimated by the 2000 Census. 1,504 additional African American

panelists and 1,506 additional Hispanic or Latino panelists were also selected within gender, age, and region strata and were invited to complete our survey. No quotas were used to restrict who completed the survey. In return for completing the survey, respondents were offered 100 points to redeem for prizes, and they were entered in a monthly drawing for \$10,000. Emails inviting the respondents to complete our survey were sent on June 11, 2004, and responses were accepted until June 21, 2004.

Non-probability sample Internet survey 3. The 1.7 million members of non-probability Internet survey firm 3's panel were recruited via the Internet in several ways. Some panelists clicked on banner advertisements or text links on news and e-commerce websites. In addition, affiliate websites, including both search engines and other websites, were offered incentives to post links on their sites inviting people to join the panel. Firm 3 occasionally sponsored newsletters that were sent via mail to potential panelists, offering invitations to join the panel.

A stratified random sample of 50,000 panel members were invited to participate in our survey in proportions matching those provided by a large marketing research firm, which conducted a survey of 80,000 U.S. offline and online individuals in 2002, in terms of gender, age, and region, and 1,223 (2%) did so. Quotas were imposed for gender, age, and region. In return for completing the survey, respondents were entered into a weekly drawing for \$3,000. E-mails invitations were sent to potential respondents on June 11, 2004, and no responses were accepted after June 14, 2004.

Non-probability sample Internet survey 4. The 1.6 million members of non-probability Internet survey firm 4's panel were recruited in several ways. A small proportion was contacted via RDD phone calls and were invited to visit the firm's website, but most clicked on banner ads or were sent emails after their email address had been given to Firm 4 by a partner website or

some other organization. Panelists became active when they completed a profile survey.

A stratified random sample of 9,921 panel members, in proportions matching 2001 Current Population Survey estimates for gender, age, and income, was invited to complete our survey, and 1103 (11%) did so. No quotas were used to restrict who could complete the survey. Respondents were entered into a drawing to win one of 114 prizes valued at \$10,000 each. Email invitations were sent on June 23, 2004, and no responses were accepted after June 30, 2004.

Non-probability sample Internet survey 5. The 2.5 million members of non-probability Internet survey firm 5's panel were recruited through more than 400 partner websites, the owners of which were paid to recruit panelists for Firm 5. Panelists visiting these partner websites, most of which provided news updates, club memberships, or other services that required website visitors to sign up, indicated their interest in joining the Firm 5 panel at the same time as they signed up for the affiliate website. They were then sent an email from Firm 5 inviting them to join the panel, but they were only considered active if they completed a profile survey after receiving a second email.

A stratified random sample of 14,000 panel members drawn in proportions matching the 2000 Census estimates of gender, age and region were invited to complete our survey, and 1,323 did so (9%). No quotas were used to limit potential respondents, and no incentives were offered for completing our survey. Invitations were sent on August 25th, 2004, and no data were collected after September 1, 2004.

Non-probability sample Internet survey 6. Non-probability Internet survey firm 6 did not maintain a panel, but instead used river methodology, which refers to recruitment through advertisements that pop-up to visitors of websites. All respondents in our study were recruited

through pop-up invitations on one popular web portal. People who accepted the invitation were asked to fill in their demographic information and were then assigned to complete one of various possible surveys.

The river methodology firm did not keep track of the number of people invited to complete the survey or whether the same person was invited more than once; 1,137 respondents completed our questionnaire. Quotas were imposed for gender, race, and age, using estimates from the Current Population Survey. The vast majority of Firm 6's respondents received a \$4.50 credit on their Internet service bill in exchange for their participation. The remaining respondents received 300 frequent flyer miles from a major airline in exchange for their participation. Responses were collected between June 16, 2004, and July 1, 2004.

Non-probability sample Internet survey 7. The 2.2 million members of non-probability Internet survey firm 7's panel were recruited via online banner ads, text links in emails sent to purchased lists of potential panelists, and Internet links in newsletters emailed to potential panelists. A stratified random sample of 2,123 adults was selected in proportions matching the 2000 Census in terms of gender, race, age, income, and education, and 1,129 (53%) completed the survey. No quotas were used to restrict participation. Respondents were given \$1 for completing our survey. Email invitations were sent on July 26, 2004, and no responses were accepted after August 1, 2004.¹

¹ The firms varied in how they prevented respondents from completing many surveys per month, prevented respondents from completing the same survey more than once, and other such procedures. The different panels also had different attrition rates. These differences did not predict differences between firms in survey accuracy.

MEASURES

General survey procedures: Identical questions measuring primary demographic, secondary demographic, and non-demographic were asked in the same order in a survey that lasted about thirty minutes on average by telephone.²

Primary demographics. The primary demographics included sex, age, race/ethnicity, education, and region of residence.

Sex: “Are you male or female?” (*Categories:* Male, Female)

Age: “In what year were you born?” To calculate age, responses were subtracted from 2004 (the year in which the survey was conducted) (*Categories:* 18-27, 28-37, 38-47, 48-57, 58-67, 68+)

Ethnicity: “Are you Spanish, Hispanic, or Latino?” (*Categories:* Yes, no)

Race: “Which of the following races do you consider yourself to be?” (*Categories:* White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Other)

Education: “What is the highest level of school you have completed or the highest degree you have received?” (*Categories:* Less than 1st Grade, 1st Grade, 2nd Grade, 3rd Grade, 4th Grade, 5th Grade, 6th Grade, 7th Grade, 8th Grade, 9th Grade,

² Many other questions were asked on the questionnaire for which we could not obtain trusted benchmark values, though it might seem possible at first glance. For example, respondents reported whether they subscribed to various magazines, which would allow calculation of the percent of American adults who subscribe to each. The American Bureau of Circulation (ABC) reports the number of subscriptions that were sold for each magazine each year, but these numbers cannot be converted to percentages of people, because two people (e.g., spouses) might share a single subscription to a magazine. Furthermore, some subscriptions are sold to businesses (e.g., doctor’s offices) and to children rather than to adults. Therefore, we could not use the ABC figures to generate appropriate benchmarks to compare to our survey. Our survey also asked a series of questions about consumer behaviors (e.g., soft drink consumption, restaurant visits), taken from the Survey of the American Consumer conducted by MediaMark Research Intelligence (MRI). This survey involved interviewing a national area probability sample face-to-face, the questions asked in our survey were included on a paper questionnaire of the MRI survey that was left with respondents at the end of the face-to-face interview and that a relatively small portion of respondents completed and mailed back. Therefore, this could not be treated as a trustworthy benchmark given our definition.

10th Grade, 11th Grade, 12th Grade with no Diploma, High School Diploma or an equivalent, such as a GED, Some College But No Degree, Associate Degree from an Occupational/Vocational program, Associate Degree from an Academic Program, Bachelor's Degree, such as B.A., B.S., or A.B., Master's Degree, such as M.A., M.S., Masters in Engineering, Masters in Education, or Masters in Social Work, Professional School Degree, such as M.D., D.D.S., or D.V.M., Doctorate Degree, such as Ph.D., Ed.D.)

Region: “In what state do you live?” (*Categories:* Northeast, Midwest, South, West, using Census region classifications)

Secondary demographics. Seven questions asked respondents about secondary demographics: marital status, total number of people living in their household, work status, number of bedrooms in their home, number of vehicles owned, home ownership, and household income.³

Marital Status: “Are you now married, widowed, divorced, separated, or never married?” (*Categories:* Married, Widowed, Divorced, Separated, Never Married)

People in Household: Respondents were asked two questions to calculate the number of people in the household. First, “Including yourself, how many adults, 18 years old or older, live in your household? Do not include college students who are living away at college, persons stationed away from here in the armed forces, or persons away in institutions.” Next, they were asked “How many

³ One opt-in Internet firm used income as a strata to invite survey respondents. Because the other eight did not, we consider it to be a secondary demographic characteristic.

people age 17 or younger live in your household?” The two numbers were added together (*Categories*: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15)

Work Status: “Last week, did you do ANY work for either pay or profit?”
(*Categories*: Yes, no)

Number of Bedrooms: “How many bedrooms are in your house, apartment, or mobile home? That is, how many bedrooms would you list if your house, apartment, or mobile home were on the market for sale or rent?” (*Categories*: none, 1, 2, 3, 4, 5 or more)

Number of Vehicles: “How many automobiles, vans, and trucks of one-ton capacity or less are kept at home for use by members of your household?”
(*Categories*: none, 1, 2, 3, 4, 5, 6)

Owning a Home: “Are your living quarters... owned or being bought by a household member, rented for cash, occupied without payment of cash rent?”
(*Categories*: Owned, Rented, Some other arrangement)

Household Income: “Thinking about your total household income from all sources, including your job, how much was your total household income in 2003 before taxes?” If respondents refused to answer the question, they were asked “Was it more than \$35,000?” and, if so, they were asked if it was more than \$50,000, and so on, up to “\$250,000 or more.” If respondents said it was less than \$35,000, they were asked if it was more than \$10,000, and so on, up to \$35,000. (*Categories*: less than \$10, 000, \$10,000-\$14,999, \$15,000-\$24,999, \$25,000-\$34,999, \$35,000-\$49,999, \$50,000-\$59,999, \$60,000-\$74,999, \$75,000-

\$99,999, \$100,000-\$149,999, \$150,000-\$199,999, \$200,000-\$249,999,
\$250,000+)

Non-demographics. Six questions asked respondents about cigarette smoking, alcohol consumption, quality of health, and possession of a U.S. passport and a driver's license.

Smoking: "Do you smoke cigarettes every day, some days, or not at all?"

(*Categories:* Every day, Some days, Not at all)

Drinking in Lifetime: [Asked only of respondents who were 21 years old or older] "In your ENTIRE LIFE, have you had at least 12 drinks of any type of alcoholic beverage?" (*Categories:* Yes, no)

Drinking This Year: [Only asked if a respondents answered "yes" to the previous question] "In the PAST YEAR, on those days that you drank alcoholic beverages, on the average, how many drinks did you have?" Respondents who answered "no" to the previous question were coded "0" for this question.

(*Categories:* 0 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15)

Quality of Health: "Would you say your health in general is excellent, very good, good, fair or poor?" (*Categories:* Excellent, Very good, Good, Fair, Poor)

Passport: "Do you personally own a valid United States passport?"

(*Categories:* Yes, no)

Driver's License: "Do you personally have a current driver's license?"

(*Categories:* Yes, no)

Several questions determined the number of non-business land lines used for speaking:

(1) Altogether, how many different telephone numbers including cell phones, are there in your

home? (*Categories:* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20); (2) How many of these (INSERT ANSWER TO #1) telephone numbers are for cellular telephones?

(*Categories:* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20; question skipped if #1 = 0); (3) How many of the telephone numbers are for business use only? (*Categories:* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20; question skipped if #1 = 0 or #1 = #2); (4) How many of the telephone numbers in your home are never used for talking on the phone and instead are used only for a different reason, like a fax machine or computer modem? (*Categories:* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20; question skipped if #1 = 0); (5) How many of the (INSERT ANSWER TO #4) telephone numbers that are never used for talking on the phone are for business use only? (*Categories:* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20; question skipped if #1 = 0, if #4 = 1); (6) Is the telephone number that is never used for talking on the phone, used for business only? (*Categories:* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20; question skipped if #1 = 0 or if #4 > 1).

Each respondent's number of non-business landlines used for speaking was computed as follows: #1 - #2 - #3 - (#4 - #5 OR - #6). This formula assumes that all cellular phones are sometimes used for non-business talking.

BENCHMARKS

Primary and secondary demographics. The primary and secondary demographic benchmark values were obtained using data from the 2004 Current Population Survey (CPS)⁴ and the 2004 American Community Survey (ACS). The CPS is a face-to-face and telephone area probability sample survey of the U.S. civilian non-institutionalized population aged 16 and

⁴ Because Opt-in Internet Firm 1's data were collected in early 2005, we tested whether that sample improved in accuracy when compared to benchmarks from the 2005 CPS, and it did not.

older, with a response rate above 85%. The 2004 Annual Social and Economic (ASEC) supplement to the CPS asked an extensive set of labor force questions of respondents and was administered during February, March, and April, 2004 (total $N=213,241$). The ASEC supplement provided the primary demographic benchmarks and some secondary demographics: marital status, people in the household, home ownership, and income. The July, 2004, CPS was used to measure work status, because employment may change during the course of the year, and most of the main study's surveys were conducted close to July, 2004. Data were weighted using the appropriate person-level weight (MARSUPWT), which incorporates the CPS basic weight, the CPS special weighting factor, the CPS non-interview adjustment, the first-stage ratio adjustment, and the second-stage ratio adjustment procedure.⁵

The ACS is an area probability sample survey of more than 848,000 housing units conducted mostly face-to-face, with some interviews completed by telephone and mailed questionnaires; the response rate is more than 92%. The ACS was used to measure the number of bedrooms and the number of vehicles owned by the U.S. civilian non-institutionalized adult population.⁶ The data were weighted using the appropriate person-level weight (PWGTP).

Non-demographic benchmarks. The non-demographic benchmark values were obtained from three sources: the 2004 National Health Interview Survey (NHIS), the U.S. State Department, and the U.S. Federal Highway Administration. Data on cigarette smoking, alcohol consumption, and quality of health were obtained from the NHIS, which asked questions identical to those asked in the main study's surveys.⁷ The NHIS is an area probability face-to-face survey of the U.S. civilian, non-institutionalized household population, with a 72.5% response rate. The interviewed sample for 2004 31,326 adults aged 18 or over. The benchmark

⁵ A full technical report on weighting and sampling procedures can be found online at <http://www.census.gov/cps/>

⁶ ACS variable names are BDS and VEH

⁷ NHIS variable names were SMKSTAT2; ALCLIFE; ALCLF_STAT & PHSTAT

values were obtained using the appropriate weights (WTFA or WTFA_SA, depending on the NHIS questionnaire).⁸ Benchmarks for the alcohol consumption questions were measured using only the respondents who were 21 years old or older.

The total number of U.S. passports held by persons aged 16 and older in May of 2005 (47,329,846) was obtained via personal communication with an official in the U.S. State Department.⁹ This number was divided by the total U.S. population age 16 and older measured by the March 2005 ASEC supplement of the Current Population Survey (220,269,194) to produce 21.49% of U.S. adults having a passport, assuming that 16- and 17-year-olds were equally likely to have passports as adults of any other age. Some holders of U.S. passports might have lived outside the U.S. and would not have been represented in this study's survey samples. Adjusting this benchmark value up or down by 1% did not affect the results reported in the text.

The total number of U.S. driver's licenses in 2004 was downloaded from the U.S. Federal Highway Administration's official website (195,432,072).¹⁰ This number was divided by the U.S. adult population in 2004 (220,479,475), yielding an estimate of 89%. If some adults with driver's licenses lived outside the U.S., then this estimate would be higher than would be optimal to compare to the main study's survey results.

MISSING DATA

If a respondent was missing an answer to a primary demographic question, the weighting algorithm did the following: (1) subdivide the respondent into virtual clones; (2) allocate the total initial weight for the respondent among the virtual clones in proportion to the benchmark values for each of the response categories; (3) iterate weight adjustments on the clones, and (4) sum the weights for the virtual clones and assign the weight to that respondent. For example,

⁸ http://www.cdc.gov/nchs/about/major/nhis/nhis_2004_data_release.htm

⁹ This was the only benchmark on passport possession that was available from the U.S. Department of State.

¹⁰ <http://www.fhwa.dot.gov/policy/ohpi/hss/hsspubs.htm>

respondents with an initial weight of 1.0 who did not report their census region were divided into clones, one of which received a weight of .1863 and was assumed to be located in the “Northeast”, another of which received a weight of .2247 and was assumed to be located in the “Midwest,” etc. No missing values were imputed for any questions.

ANALYSIS PLAN

Bootstrapping standard errors. To test whether the differences between the firms in average absolute errors were statistically significant, we ran a bootstrapping procedure to estimate standard errors. These standard errors for the average absolute error for each survey were in turn used to conduct pair-wise t-tests. To calculate standard errors, the “bootstrap” command in Stata first randomly selected observations from the data set to produce a sample of equal size to the original sample. Next, a Stata program normalized the weights to a mean of 1 and calculated an average absolute error for the new sample. In the analysis with post-stratification, a unique set of post-stratification weights was calculated for each random sample. This was repeated 100 times. Finally, the procedure used the 100 average absolute errors to calculate a standard error.

Results

Table 1 shows t-statistics testing the significance of differences between the surveys in terms of average errors across all 19 benchmarks without post-stratification. The probability sample surveys were significantly more accurate than all of the non-probability sample Internet surveys, and the non-probability sample Internet surveys were rarely significantly different from one another in terms of accuracy. Non-probability sample Internet survey 7 was significantly less accurate than the other non-probability sample Internet surveys.

Table 2 shows comparable t-statistics when using only primary demographics without post-stratification. Again, the probability sample surveys were significantly more accurate than the non-probability sample Internet surveys. Non-probability sample Internet survey 7 was significantly less accurate than the other non-probability sample Internet surveys. And Non-probability sample Internet survey 1 was significantly more accurate than the other non-probability sample Internet surveys.

Table 3 shows t-tests assessing the statistical significance of differences between the surveys in terms of accuracy for the 13 secondary demographics and non-demographics with and without post-stratification. The probability samples were, on average, significantly more accurate than the non-probability sample surveys. In addition, the non-probability sample Internet surveys were rarely different from one another, although non-probability sample Internet survey 7 was often significantly less accurate than the others.

The right-most two columns of Table 4 shows the mean absolute error across the RDD surveys for each benchmark and the standard deviation of the errors. These average errors were relative small and minimally variable. The telephone survey used in the main study was the least accurate among these probability sample telephone surveys. These results suggest that the findings from the main study would likely have been replicated had a different probability sample telephone survey been used instead.

Table 5 shows similar results for probability sample Internet surveys: their errors were minimally variable and quite small. The probability sample Internet survey used in the main study was again less accurate than the others shown in Table 5, suggesting that the main study may have under-stated the superior accuracy of probability sample Internet surveys.

Table 6's numbers are comparable to those in Tables 5 and 6, this time for the non-probability Internet surveys and using the benchmarks that could be obtained for all the RDD surveys. The non-probability sample Internet surveys were quite a bit more variable in their errors and had relatively large average absolute errors (see the last two columns of Table 6, as compared to the last two columns of Tables 4 and 5).

Table 7 shows accuracy summary statistics when quality of health was omitted from the set of benchmarks. The results are comparable to those obtained when including quality of health.

Table 8 shows accuracy summary statistics generated using post-stratification weights from the three firms that provided them. These figures illustrate lower accuracy than when using the post-stratification weights we built for those three surveys.

Table 9 shows results when not capping the post-stratification weights we constructed. These figures show that capping caused accuracy to improve for five of the non-probability sample surveys and did not affect the accuracy of the probability sample surveys.

Table 10 reports tests of the statistical difference of the change in accuracy caused by using the post-stratification weights that we build. The figures show that post-stratification weighting significantly increased the accuracy of the two probability samples. Weighting only significantly or marginally increased the accuracy of three of the seven non-probability samples, it marginally *decreased* the accuracy of one of the non-probability samples, and had no detectable effect on the remaining three samples.

Table 11 shows the targets obtained from the CPS that were used to create the survey weights.

Table 1. T-tests Comparing Average Absolute Errors (Without Post-Stratification) Across Samples on All Nineteen Demographics, Using Bootstrapped Standard Errors

Survey	Probability Sample Surveys		Non-Probability Sample Internet Surveys						
	Telephone	Internet	1	2	3	4	5	6	7
<i>Average Error</i>	3.58%	3.49%	4.90%	5.53%	6.17%	5.28%	5.51%	5.59%	10.16%
Telephone	-	-0.09	1.32 **	1.95 ***	2.60 ***	1.71 ***	1.93 ***	2.02 ***	6.58 ***
Prob-Internet	0.09	-	1.41 ***	2.04 ***	2.69 ***	1.79 ***	2.02 ***	2.11 ***	6.67 ***
Non-Prob Internet 1	-1.32 **	-1.41 ***	-	0.63	1.27 **	0.38	0.61 +	0.69 +	5.26 ***
Non-Prob Internet 2	-1.95 ***	-2.04 ***	-0.63	-	0.65	-0.24	-0.02	0.07	4.63 ***
Non-Prob Internet 3	-2.60 ***	-2.69 ***	-1.27 **	-0.65	-	-0.89 *	-0.67	-0.58	3.98 ***
Non-Prob Internet 4	-1.71 ***	-1.79 ***	-0.38	0.24	0.89 *	-	0.22	0.31	4.87 ***
Non-Prob Internet 5	-1.93 ***	-2.02 ***	-0.61 +	0.02	0.67	-0.22	-	0.09 +	4.65 ***
Non-Prob Internet 6	-2.02 ***	-2.11 ***	-0.69 +	-0.07	0.58	-0.31	-0.09 +	-	4.56 ***
Non-Prob Internet 7	-6.58 ***	-6.67 ***	-5.26 ***	-4.63 ***	-3.98 ***	-4.87 ***	-4.65 ***	-4.56 ***	-

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 2. T-tests Comparing Average Absolute Errors (Without Post-Stratification) Across Samples on Primary Demographics Using Bootstrapped Standard Errors

Survey	Probability Sample Surveys		Non-Probability Sample Internet Surveys						
	Telephone	Internet	1	2	3	4	5	6	7
<i>Average Error</i>	3.43%	2.47%	4.14%	4.96%	6.44%	6.35%	7.01%	6.05%	12.82%
Telephone	-	-0.96	0.71	1.53 *	3.01 ***	2.92 ***	3.57 ***	2.62 ***	9.39 ***
Prob-Internet	0.96	-	1.67 **	2.49 ***	3.98 ***	3.88 ***	4.54 ***	3.58 ***	10.35 ***
Non-Prob Internet 1	-0.71	-1.67 **	-	0.82	2.30 ***	2.21 ***	2.87 ***	1.91 **	8.68 ***
Non-Prob Internet 2	-1.53 *	-2.49 ***	-0.82	-	1.49 *	1.39 *	2.05 **	1.09	7.86 ***
Non-Prob Internet 3	-3.01 ***	-3.98 ***	-2.30 ***	-1.49 *	-	-0.09	0.56	-0.39	6.38 ***
Non-Prob Internet 4	-2.92 ***	-3.88 ***	-2.21 ***	-1.39 *	0.09	-	0.66	-0.30	6.47 ***
Non-Prob Internet 5	-3.57 ***	-4.54 ***	-2.87 ***	-2.05 **	-0.56	-0.66	-	-0.96 +	5.81 ***
Non-Prob Internet 6	-2.62 ***	-3.58 ***	-1.91 **	-1.09	0.39	0.30	0.96 +	-	6.77 ***
Non-Prob Internet 7	-9.39 ***	-10.35 ***	-8.68 ***	-7.86 ***	-6.38 ***	-6.47 ***	-5.81 ***	-6.77 ***	-

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 3. T-tests Comparing Average Absolute Errors on Secondary Demographic and Non-demographic Benchmarks Using Bootstrapped Standard Errors

Survey	Probability Sample Surveys				Non-Probability Sample Internet Surveys												
	Telephone		Internet		1	2	3	4	5	6	7						
Without Post-stratification																	
<i>Average Error</i>	3.64%		3.96%		5.25%	5.79%	5.85%	4.79%	4.81%	5.38%	8.93%						
Telephone	-	0.32	1.61	**	2.15	***	2.20	***	1.15	*	1.17	*	1.74	**	5.28	***	
Prob-Internet	-0.32	-	1.29	**	1.83	**	1.89	***	0.83	0.86	1.42	*	4.97	***			
Non-Prob Internet 1	-1.61	**	-1.29	**	-	0.54	0.60	-0.46	-0.44	0.13	3.68	***					
Non-Prob Internet 2	-2.15	***	-1.83	**	-0.54	-	0.06	-1.00	+	-0.98	+	-0.41	3.14	***			
Non-Prob Internet 3	-2.20	***	-1.89	***	-0.60	-0.06	-	-1.06	*	-1.03	+	-0.47	3.08	***			
Non-Prob Internet 4	-1.15	*	-0.83	0.46	1.00	+	1.06	*	-	0.02	0.59	4.14	***				
Non-Prob Internet 5	-1.17	*	-0.86	0.44	0.98	+	1.03	+	-0.02	-	0.57	+	4.11	***			
Non-Prob Internet 6	-1.74	**	-1.42	*	-0.13	0.41	0.47	-0.59	-0.57	+	-	3.55	***				
Non-Prob Internet 7	-5.28	***	-4.97	***	-3.68	***	-3.14	***	-3.08	***	-4.14	***	-4.11	***	-3.55	***	-
With Post-stratification																	
<i>Average Error</i>	2.93%		3.37%		4.57%	5.23%	4.55%	5.52%	5.19%	5.02%	6.89%						
Telephone	-	0.44	1.64	**	2.31	***	1.63	*	2.59	***	2.26	***	2.09	***	3.97	***	
Prob-Internet	-0.44	-	1.20	*	1.86	***	1.18	+	2.15	***	1.82	**	1.65	**	3.52	***	
Non-Prob Internet 1	-1.64	**	-1.20	*	-	0.67	-0.01	0.95	0.62	0.45	2.33	*					
Non-Prob Internet 2	-2.31	***	-1.86	***	-0.67	-	-0.68	0.28	-0.05	-0.22	1.66	+					
Non-Prob Internet 3	-1.63	*	-1.18	+	0.01	0.68	-	0.96	0.64	0.46	2.34	*					
Non-Prob Internet 4	-2.59	***	-2.15	***	-0.95	-0.28	-0.96	-	-0.33	-0.50	1.38						
Non-Prob Internet 5	-2.26	***	-1.82	**	-0.62	0.05	-0.64	0.33	-	-0.17	+	1.71	+				
Non-Prob Internet 6	-2.09	***	-1.65	**	-0.45	0.22	-0.46	0.50	0.17	+	-	1.88	*				
Non-Prob Internet 7	-3.97	***	-3.52	***	-2.33	*	-1.66	+	-2.34	*	-1.38	-1.71	+	-1.88	*	-	

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 4. Absolute Error (without Post-Stratification) for Primary Demographics and Some Secondary Demographics the Main Study’s Telephone Survey and Six Other Telephone Surveys Fielded Between June – August of 2004

Demographic Category	Benchmark value	Main Study Telephone Survey	Other RDD Surveys						Summary	
			2	3	4	5	6	7	<i>M</i> of errors	<i>SD</i> of errors
Female	51.68%									
%		55.36%	50.76%	52.17%	51.6%	50.76%	52.17%	50.5%		
Absolute error		3.68	0.92	0.49	0.08	0.92	0.49	1.18	1.11	1.19
Aged 30-44	29.7									
%		34.72	27.18	26.41	27.74	27.18	26.41	29.36		
Absolute error		5.02	2.52	3.29	1.96	2.52	3.29	0.34	2.71	1.43
White	82.02									
%		78.31	85.64	81.36	79.79	85.64	81.36	85.02		
Absolute error		3.71	3.62	0.66	2.23	3.62	0.66	3.00	2.50	1.36
Non-Hispanic	87.62									
%		94.63	94.22	93.97	93.25	94.22	93.97	93.48		
Absolute error		7.01	6.60	6.35	5.63	6.60	6.35	5.86	6.34	0.47
High school degree only	31.75									
%		27.41	25.24	25.5	25.1	25.24	25.5	24.78		
Absolute error		4.34	6.51	6.25	6.65	6.51	6.25	6.97	6.21	0.86
South	35.92									
%		34.22	33.78	37.22	35.27	33.78	37.22	33.85		
Absolute error		1.70	2.14	1.30	0.65	2.14	1.30	2.07	1.61	0.56
Married	56.5									
%		61.93	63.83	65.31	65.77	63.83	65.31	65.93		
Absolute error		5.43	7.33	8.81	9.27	7.33	8.81	9.43	8.06	1.44
2 adults living in the house	56.28									
%		56.22	59.34	58.28	57.72	59.34	58.28	58.22		
Absolute error		0.06	3.06	2.00	1.44	3.06	2.00	1.94	1.94	1.02
Household income \$50K-\$75K	19.79									
%		16.28	18.42	18.38	19.07	18.42	18.38	19.44		
Absolute error		3.51	1.37	1.41	0.72	1.37	1.41	0.35	1.45	1.00
Overall <i>M</i> of errors		3.83	3.79	3.40	3.18	3.79	3.40	3.46	3.55	0.25
Overall <i>SD</i> of errors		2.04	2.42	3.01	3.21	2.42	3.01	3.22	2.76	0.46

Note: Different modal categories were used in this analysis vs. the main analysis reported in the text for age, household size, and income, because the RDD survey questions did not allow grouping responses into the same categories. For example, for income, RDD survey respondents chose one category from a list that did not include with the modal category chosen for the other analyses (50-60K). Similarly, the RDD surveys did not ask for the number of children in the household, so the number of adults was used as a benchmark. For these questions, the modal category from the RDD surveys was chosen, new benchmarks were estimated, and the estimates from the Internet surveys were re-calculated.

Table 5. Absolute Error (Without Post-Stratification) for Primary Demographics and Some Secondary Demographics for the Probability Sample Internet Survey and Six Other Probability Sample Internet Surveys Conducted Between June – August of 2004

Demographic Category	Benchmark value	Main Study Probability Sample Internet Survey	Other Probability Sample Internet Surveys						Summary		
			2	3	4	5	6	7	<i>M</i> of errors	<i>SD</i> of errors	
Female	51.68%										
%		50.57%	52.49%	50.17%	51.11%	50.00%	52.57%	49.68%			
Absolute error		1.11	0.81	1.51	0.57	1.68	0.89	2.00	1.22	0.52	
Aged 30-44	29.7										
%		34.67	26.39	30.83	29.78	29.37	30.42	27.59			
Absolute error		4.97	3.31	1.13	0.08	0.33	0.72	2.11	1.81	1.79	
White	82.02										
%		76.09	81.63	82.59	79.22	82.33	78.92	81.57			
Absolute error		5.93	0.39	0.57	2.80	0.31	3.10	0.45	1.94	2.13	
Non-Hispanic	87.62										
%		90.86	89.36	91.82	90.22	92.3	89.25	90.95			
Absolute error		3.24	1.74	4.20	2.60	4.68	1.63	3.33	3.06	1.16	
High school degree only	31.75										
%		31.41	35.08	34.46	32.22	26.5	32.99	32.11			
Absolute error		0.34	3.33	2.71	0.47	5.25	1.24	0.36	1.96	1.88	
South	35.92										
%		38.82	33.98	35.58	34.22	34.86	34.11	36.31			
Absolute error		2.9	1.94	0.34	1.70	1.06	1.81	0.39	1.45	0.92	
Married	56.5										
%		59.82	57.46	60.66	58.11	57.7	61	59.27			
Absolute error		3.32	0.96	4.16	1.61	1.20	4.50	2.77	2.65	1.43	
2 adults living in the house	56.28										
%		59.08	57.73	57.23	57.56	57.05	56.5	57.44			
Absolute error		2.80	1.45	0.95	1.28	0.77	0.22	1.16	1.23	0.80	
Household income \$50K-\$75K	19.79										
%		22.20	21.27	23.1	21.56	21.93	20.23	19.29			
Absolute error		2.41	1.48	3.31	1.77	2.14	0.44	0.50	1.72	1.03	
Overall <i>M</i> of errors		3.00	1.71	2.10	1.43	1.94	1.62	1.45	1.89	0.55	
Overall <i>SD</i> of errors		1.72	1.03	1.52	0.93	1.82	1.39	1.14	1.36	0.34	

Table 6. Absolute Error for Primary Demographics and Some Secondary Demographics for Non-Probability Sample Internet Surveys without Post-Stratification Using Categories Used to Assess Variation in Accuracy Across the 7 RDD Surveys and the 7 Probability Sample Internet Surveys

Demographic Category	Benchmark value	Non-Probability Sample Internet Surveys							Summary	
		1	2	3	4	5	6	7	<i>M</i> of errors	<i>SD</i> of errors
Female	51.68%									
%		53.80%	49.82%	48.73%	50.73%	52.26%	49.61%	55.45%		
Absolute error		2.12	1.86	2.95	0.95	0.58	2.07	3.77	2.04	1.09
Aged 30-44	29.7									
%		32.83	35.89	38.59	32.27	32.32	35.84	38.02		
Absolute error		3.13	6.19	8.89	2.57	2.62	6.14	8.32	5.41	2.67
White	82.02									
%		79.58	85.83	89.62	88.49	66.89	73.88	41.54		
Absolute error		2.44	3.81	7.60	6.47	15.13	8.14	40.48	12.01	13.19
Non-Hispanic	87.62									
%		88.64	95.09	96.65	96.64	94.78	93.44	90.19		
Absolute error		1.02	7.47	9.03	9.02	7.16	5.82	2.57	6.01	3.12
High school degree only	31.75									
%		13.76	18.52	16.20	16.50	19.03	20.45	16.60		
Absolute error		17.99	13.23	15.55	15.25	12.72	11.30	15.15	14.45	2.21
South	35.92									
%		36.13	38.01	39.03	29.80	40.99	44.85	26.25		
Absolute error		0.21	2.09	3.11	6.12	5.07	8.93	9.67	5.03	3.50
Married	56.5									
%		58.77	59.93	61.49	56.82	58.15	53.71	45.54		
Absolute error		2.27	3.43	4.99	0.32	1.65	2.79	10.96	3.77	3.48
2 adults living in the house	56.28									
%		58.28	60.56	63.37	56.27	58.51	55.96	53.86		
Absolute error		2.00	4.28	7.09	0.01	2.23	0.32	2.42	2.62	2.43
Household income \$50K-\$75K	19.79									
%		20.93	22.74	23.20	21.58	24.49	21.70	21.55		
Absolute error		1.14	2.95	3.41	1.79	4.70	1.91	1.76	2.52	1.23
Overall <i>M</i> of errors		3.59	5.04	6.96	4.72	5.76	5.27	10.56	5.99	2.26
Overall <i>SD</i> of errors		5.47	3.57	4.02	5.04	5.07	3.73	12.12	5.58	2.98

Table 7. Summary Statistics When Excluding Quality of Health From the Benchmark Calculations

Evaluative Criterion	Probability Sample Surveys		Non-probability Sample Internet Surveys						
	Telephone	Internet	1	2	3	4	5	6	7
Average absolute error									
Primary demographics									
Without Post-stratification	3.43% ^b	2.47%	4.14% ^b	4.96% ^{ab}	6.44% ^{ab}	6.35% ^{ab}	7.01% ^{ab}	6.05% ^{ab}	12.82% ^{ab}
All benchmarks									
Without Post-stratification	3.68%	3.35%	4.79% ^{ab}	5.36% ^{ab}	6.05% ^{ab}	5.19% ^{ab}	5.69% ^{ab}	5.65% ^{ab}	10.11% ^{ab}
Secondary and non-demographics									
Without Post-stratification	3.80%	3.79%	5.12% ^{ab}	5.56% ^{ab}	5.58% ^{ab}	4.61% ^a	5.03% ^a	5.46% ^{ab}	8.76% ^{ab}
With post-stratification	3.04%	3.07%	4.50% ^{ab}	5.07% ^{ab}	4.22% ^a	5.33% ^{ab}	5.46% ^{ab}	5.29% ^{ab}	6.80% ^{ab}
Rank: Average absolute error									
Primary demographics									
Without Post-stratification	2	1	3	4	7	6	8	5	9
All benchmarks									
Without Post-stratification	2	1	3	5	8	4	7	6	9
Secondary and non-demographics									
Without Post-stratification	2	1	5	7	8	3	4	6	9
With post-stratification	1	2	4	5	3	7	8	6	9
Largest absolute error									
All benchmarks									
Without Post-stratification	11.71%	9.59%	17.99%	13.23%	15.55%	15.25%	15.13%	15.97%	40.48%
Secondary and non-demographics									
Without Post-stratification	11.71%	9.59%	14.68%	12.12%	13.03%	14.80%	13.70%	15.97%	20.04%
With post-stratification	8.94%	8.42%	14.75%	12.05%	12.09%	15.21%	13.14%	9.74%	17.47%
% Significant differences from benchmark									
All benchmarks									
Without Post-stratification	58%	63%	63%	68%	79%	58%	68%	63%	89%
Secondary and non-demographics									
Without Post-stratification	54%	69%	77%	77%	77%	54%	69%	62%	85%
With post-stratification	38%	38%	77%	69%	69%	77%	69%	77%	77%

Table 8. Summary Statistics for Secondary Demographics and Non-Demographics When Using Weights Provided by the Survey Firms for the Probability Sample Surveys and Non-Probability Sample Internet Survey 1

Evaluative Criteria	Probability Sample Surveys		Non-Probability Sample Internet
	Telephone	Internet	Survey 1
Average absolute error	3.42%	4.11%	4.99%
Largest absolute error	6.70%	9.04%	11.90%
% Significant differences from benchmark	58%	75%	58%

Table 9. Summary Statistics for Secondary Demographics and Non-Demographics Using Un-Capped Post-Stratification Weights

Evaluative Criteria	Probability Sample Surveys		Non-Probability Sample Internet Surveys						
	Telephone	Internet	1	2	3	4	5	6	7
Average Error	2.98%	3.37%	4.15%	5.17%	5.27%	5.97%	5.74%	5.17%	6.17%
Rank: Average Error	1	2	3	5	6	8	7	4	9
Largest Error	9.00%	8.41%	14.97%	11.90%	12.82%	15.51%	12.67%	10.22%	17.22%
% Significant Differences	25%	42%	67%	67%	58%	92%	75%	75%	58%

Table 10. Bootstrapped Standard Errors and T-tests of the Difference Between the Average Absolute Error with and without Post-stratification for Each Survey

Firm	Δ in average absolute error due to post- stratification	Bootstrapped standard error of Δ	t	p
Telephone	0.72	0.37	1.96	0.05
Prob-Internet	0.59	0.24	2.42	0.02
Non-Prob Internet 1	0.68	0.38	1.79	0.07
Non-Prob Internet 2	0.56	0.39	1.44	0.15
Non-Prob Internet 3	1.50	0.45	3.34	0.00
Non-Prob Internet 4	-0.73	0.43	-1.67	0.09
Non-Prob Internet 5	-0.37	0.60	-0.62	0.54
Non-Prob Internet 6	0.37	0.42	0.88	0.38
Non-Prob Internet 7	2.03	0.78	2.62	0.01

Table 11. Weighting Targets for Post-Stratification Weights (Source: 2004 CPS ASEC Supplement)

Demographic	Weighting Target %		
	Male	Female	Total
Sex and Age			
18-24	6.61%	6.35%	12.96%
25-34	9.13	9.13	18.26
35-44	10.03	10.27	20.3
45-54	9.35	9.77	19.12
55-64	6.31	6.9	13.21
65+	6.89	9.25	16.14
Sex and Education			
Less than high school	7.97	7.82	15.79
High school graduate	15.11	16.64	31.75
Some college	11.04	12.53	23.57
College graduate	9.65	10.8	20.45
Graduate degree	4.54	3.89	8.43
Race			
White only			82.02
Black only			11.47
Other race			6.51
Ethnicity			
Hispanic / Latino			87.62
Not Hispanic / Latino			12.38
Region			
Northeast			18.63
Midwest			22.47
South			35.92
West			22.98

Appendix 1.*Stata Program to Compare Survey Accuracy, Compute the Post-Stratification Weights, and Bootstrap the Average Absolute Error Both Before Post-Stratification and With Post-Stratification Weights*

```

*****
* A program compare the accuracy of a survey to other surveys and benchmark values
*****
*****
clear
set more off

set mem 500m

use "data.dta"

recode source (2=8) (8=1) (1=2) (3=9) (5=4) (9=3) (4=5)

recode source (3 =4) (4=3)

*****
*** Creating the variables
*****

gen age = abs(yrborn-104)

gen regions = .

*** generating northeast values
foreach num of numlist 7 20 22 30 40 46 31 33 39 {
    replace regions = 1 if state==`num'
}

foreach num of numlist 14 15 23 36 50 16 17 24 26 28 35 42 {
    replace regions = 2 if state==`num'
}

foreach num of numlist 8 9 10 11 21 34 41 47 49 1 18 25 43 4 19 37 44 {
    replace regions = 3 if state==`num'
}

foreach num of numlist 3 6 13 27 29 32 45 51 2 5 12 38 48 {
    replace regions = 4 if state==`num'
}

gen racecats = .
replace racecats = 1 if cpsrace==1
replace racecats=2 if cpsrace==2
replace racecats=3 if cpsrace>2

gen hispanic = 1 if cpshsp ==1
replace hispanic = 2 if cpshsp ~.=1
replace hispanic = . if cpshsp==.

gen educats=.
replace educats = 1 if cpseduc<14 & cpseduc~=.
replace educats=2 if cpseduc==14 & cpseduc~=.
replace educats=3 if cpseduc==15 & cpseduc~.= | cpseduc==16 & cpseduc~=.
replace educats=4 if cpseduc==18 | cpseduc ==17 & cpseduc~=.
replace educats=5 if cpseduc>18 & cpseduc~=.
label variable educats "no hs, hs, some college, college, graduate"

** note: vocational degree is with some college, associate's degree is with college degree

gen agecats = .
replace agecats = 1 if age >17 & age~=.
replace agecats = 2 if age >27 & age~=.
replace agecats = 3 if age >37 & age~=.
replace agecats = 4 if age >47 & age~=.
replace agecats = 5 if age >57 & age~=.
replace agecats = 6 if age >67 & age~=.

gen marcats = .
replace marcats = 3 if cpsmart>1 & cpsmart<5
replace marcats=2 if cpsmart==5
replace marcats =1 if cpsmart==1
label variable marcats "married, never married, other"

gen incomecats = .
replace incomecats = 1 if hhinc>0 & hhinc~=.
replace incomecats = 2 if hhinc>10000 & hhinc~.= | hhinc_p ==1

```

```

replace incomecat = 3 if hhinc>15000 & hhinc<=. | hhinc_q ==1
replace incomecat = 4 if hhinc>25000 & hhinc<=. | hhinc_r ==1
replace incomecat = 5 if hhinc>35000 & hhinc<=. | hhinc_a ==1
replace incomecat = 6 if hhinc>50000 & hhinc<=. | hhinc_b ==1
replace incomecat = 7 if hhinc>60000 & hhinc<=. | hhinc_c ==1
replace incomecat = 8 if hhinc>75000 & hhinc<=. | hhinc_d ==1
replace incomecat = 9 if hhinc>100000 & hhinc<=. | hhinc_e ==1
replace incomecat = 10 if hhinc>150000 & hhinc<=. | hhinc_f ==1
replace incomecat = 11 if hhinc>200000 & hhinc<=. | hhinc_g ==1
replace incomecat = 12 if hhinc>250000 & hhinc<=. | hhinc_h ==1

gen totalnum = numkd+numad

*** Creating a drinks variable

gen numberofdrinks = 1 if nhis_ac2==1 & nhis_ac1==1
replace numberofdrinks = 2 if nhis_ac2==2 & nhis_ac1==1
replace numberofdrinks=3 if nhis_ac2>2 & nhis_ac1==1
replace numberofdrinks=1 if nhis_ac2==0
*replace numberofdrinks=0 if nhis_ac1==0

order cpssex agecats racecats hispanic educats regions marcats totalnum cpspuwk cpswj census38 census43 cpslivq incomecat brftu nhis_ac1 numberofdrinks brfex
nhishs mripp mridl

*** tabbing to create a variable for each value
foreach var of varlist cpssex - mridl{
    quietly tab `var', gen(`var')
}

order cpssex2 agecats1 racecats1 hispanic2 educats4 regions3 marcats1 totalnum3 cpspuwk1 census384 census433 cpslivq1 incomecat5 brftu3 nhis_ac11
numberofdrinks1 nhishs2 nhishs2 mripp2 mridl1

****
* Creating categories for weighting
****

gen agecat =
replace agecat= 1 if age<25
replace agecat=2 if age>24 & age<35
replace agecat=3 if age>34 & age<45
replace agecat=4 if age>44 & age<55
replace agecat=5 if age>54 & age<65
replace agecat=6 if age>64

gen agegender = agecat if cpssex==1
replace agegender = agecat+6 if cpssex==2

*** generating edugender variable
gen edugender=educats if cpssex==1
replace edugender=educats+5 if cpssex==2

*** generating the racecats variable
* already done

*** regions already done

****
*** Creating phone lines variables
****

gen constant=1

*** Generating phone number probabilities

gen phone = q48 - q49 - q50 - (q51-q51a)
replace phone = q48 - q49 - q50 - (q51-q51b)
replace phone = 1 if phone<1 | phone ==.
replace phone = 3 if phone>3

**** Generating probability of selection weights

gen pwhouse = 1/phone

replace pwhouse =1 if source ~=1 & source~=2 | pwhouse==.

gen numadweight = numad
replace numadweight = 3 if numad>3
replace numadweight=1 if numadweight==.
gen pwperson = (1/phone)*numadweight
replace pwperson = 1 if source~=1 & source~=2 | pwperson==.
replace pwperson = 1 if source==2

****
*** We'll want to re-adjust the mean of the proportional sampling weights for each sample
****

capture program drop propweights
program propweights, rclass
    version 10.1
    quietly {

        replace phone = q48 - q49 - q50 - (q51-q51a)

```

```

replace phone = q48 - q49 - q50 - (q51-q51b)
replace phone = 1 if phone<1 | phone ==.
replace phone = 3 if phone>3

**** Generating probability of selection weights

replace pwhouse = 1/phone

replace pwhouse =1 if source ~=1 & source~=2 | pwhouse==.

replace numadweight = numad
replace numadweight = 3 if numad>3
replace numadweight=1 if numadweight==.
replace pwperson = (1/phone)*numadweight
replace pwperson = 1 if source~=1 & source~=2 | pwperson==.

foreach var of varlist pwhouse pwperson {
    replace `var' = (`var')/(sum(`var')/sum(constant)) if source==2
    replace `var' = (`var')/(sum(`var')/sum(constant)) if source==1
}

replace pwperson = 1 if source ==2

end

*****
*** Setting up the "Marginals" program
*****

**** Creating the benchmark questions and their "true values"
global num = 0

replace educats4 = 1 if cpseduc==14
replace educats4 = 0 if cpseduc~=14
replace educats4 = . if cpseduc==.

gen white = 1 if cpsrace==1
replace white = 0 if cpsrace~=1
replace white = . if cpsrace==.

global num = 0

foreach var of varlist cpssex2 agecats3 racecats1 hispanic2 educats4 regions3 marcats1 totalnum3 cpspuwk1 cnsus384 cnsus433 cpslivq1
incomecat5 brftu3 nhis_ac11 numberofdrinks1 nhishs2 mripp2 mridl1 {
    global num = $num+1
    clonevar demovar_ $num = `var'
}

**** copy the percentage X 100 for each of the benchmark questions below

global p =0

**** Looping over the "true values"

foreach num of numlist 5168 2083 8202 8762 3175 3592 5650 3384 6080 4338 4146 7250 1511 7825 7745 3767 3193 7850 8900 {
    global num = `num'
    global truevalue = ($num)/100
    global p = $p+1
    global truevalue_ $p = $truevalue
}

*****
* Setting up the post file to save results
*****

quietly capture postclose resultsbuffer
postfile resultsbuffer m_demovar_1 p_demovar_1 m_demovar_2 p_demovar_2 m_demovar_3 p_demovar_3 m_demovar_4 p_demovar_4 m_demovar_5 p_demovar_5
m_demovar_6 p_demovar_6 m_demovar_7 p_demovar_7 m_demovar_8 p_demovar_8 m_demovar_9 p_demovar_9 using "/benchmarkoutput.dta", replace

*****
* Looping over the 9 firms and saving values
*****

**** finding out how many sources there are
su source
global sourcenum = r(max)

**** Running the weighting programs
propweights
gen prop = pwperson
gen post = .
forvalues x = 1/$sourcenum {
    replace pwperson = prop
    replace pwperson = 0 if source==`x'
    *Specify the original weighting variable (which might be all ones)
    global weight_var = "pwperson"
}

```

*Build flag variables for each category we are weighting for

```
foreach xn of numlist 1/31 {
  capture drop flag_`xn'
  gen flag_`xn' = 0
}
```

*Set the flags to 1 if person belongs in category

```
replace flag_1 = 1 if agegender==1
replace flag_2 = 1 if agegender==2
replace flag_3 = 1 if agegender==3
replace flag_4 = 1 if agegender==4
replace flag_5 = 1 if agegender==5
replace flag_6 = 1 if agegender==6
replace flag_7 = 1 if agegender==7
replace flag_8 = 1 if agegender==8
replace flag_9 = 1 if agegender==9
replace flag_10 = 1 if agegender==10
replace flag_11 = 1 if agegender==11
replace flag_12 = 1 if agegender==12
```

```
replace flag_13 = 1 if edugender==1
replace flag_14 = 1 if edugender==2
replace flag_15 = 1 if edugender==3
replace flag_16 = 1 if edugender==4
replace flag_17 = 1 if edugender==5
replace flag_18 = 1 if edugender==6
replace flag_19 = 1 if edugender==7
replace flag_20 = 1 if edugender==8
replace flag_21 = 1 if edugender==9
replace flag_22 = 1 if edugender==10
```

```
replace flag_23 = 1 if racecats==1
replace flag_24 = 1 if racecats==2
replace flag_25 = 1 if racecats==3
```

```
replace flag_26 = 1 if hispanic==1
replace flag_27 = 1 if hispanic==2
```

```
replace flag_28 = 1 if regions==1
replace flag_29 = 1 if regions==2
replace flag_30 = 1 if regions==3
replace flag_31 = 1 if regions==4
```

*Specify targets for each category

*Weighting targets, from the 2004 CPS ASEC, targets go in this order: 12 age X gender categories; 8 education X gender categories; 4 race / ethnicity categories; and 4 regional categories

```
global target_1= .0661
global target_2= .0913
global target_3= .1003
global target_4= .0935
global target_5= .0631
global target_6= .0689
global target_7= .0635
global target_8= .0913
global target_9= .1027
global target_10= .0977
global target_11= .069
global target_12= .0925
```

```
global target_13= .0797
global target_14= .1511
global target_15= .1104
global target_16= .0965
global target_17= .0454
global target_18= .0782
global target_19= .1664
global target_20= .1253
global target_21= .1080
global target_22= .0389
```

```
global target_23= .8202
global target_24= .1147
global target_25= .0651
```

```
global target_26= .1238
global target_27= .8762
```

```
global target_28= .1863
global target_29= .2247
global target_30= .3592
global target_31= .2298
```

*Set the flags equal to the true value if a person is missing data from that category


```

*** Calculating the marginals and running a t-test.
*** Program will also calculate the error and put it in a row below the marginal value
*****

* cap weights here

*replace post = 5 if post>5 & post=-.

* change to firm weights here

*replace post = prop3 if source==4
*replace post = totalwt if source == 2
*replace post = weights if source== 1

*replace post = weight4 if source==4
*replace post = weight2 if source == 2
*replace post = weight1 if source== 1

*** One row is the un-weighted analysis, the next row is the error, followed by weighted marginals and error
gen weight=.

global totalerror
forvalues i = 1/$p {
    forvalues s = 1/$sourcenum {
        replace weight = prop
        summarize demovar_`i' if source==`s' [aweight=weight]
        global m_demovar_`s' = r(mean)
        global sd_demovar_`s' = r(sd)
        global n_demovar_`s' = r(N)
        global m_error = abs($m_demovar_`s' - ($truevalue_`i')/100)
        ttesti $n_demovar_`s' $m_error $sd_demovar_`s' 0
        global p_demovar_`s' = r(p)
    }
    post resultsbuffer ($m_demovar_1) ($p_demovar_1) ($m_demovar_2) ($p_demovar_2) ($m_demovar_3) ($p_demovar_3) ($m_demovar_4) ($p_demovar_4)
    ($m_demovar_5) ($p_demovar_5) ($m_demovar_6) ($p_demovar_6) ($m_demovar_7) ($p_demovar_7) ($m_demovar_8) ($p_demovar_8) ($m_demovar_9) ($p_demovar_9)
    forvalues s = 1/$sourcenum {
        global m_demovar_`s' = abs($m_demovar_`s' - ($truevalue_`i')/100)
        global p_demovar_`s' = 0
    }
    post resultsbuffer ($m_demovar_1) ($p_demovar_1) ($m_demovar_2) ($p_demovar_2) ($m_demovar_3) ($p_demovar_3) ($m_demovar_4) ($p_demovar_4)
    ($m_demovar_5) ($p_demovar_5) ($m_demovar_6) ($p_demovar_6) ($m_demovar_7) ($p_demovar_7) ($m_demovar_8) ($p_demovar_8) ($m_demovar_9) ($p_demovar_9)
    *** Now for the weighted analyses
    forvalues s = 1/$sourcenum {
        replace weight = post
        summarize demovar_`i' if source==`s' [aweight=weight]
        global m_demovar_`s' = r(mean)
        global sd_demovar_`s' = r(sd)
        global n_demovar_`s' = r(N)
        global m_error = abs($m_demovar_`s' - ($truevalue_`i')/100)
        ttesti $n_demovar_`s' $m_error $sd_demovar_`s' 0
        global p_demovar_`s' = r(p)
    }
    post resultsbuffer ($m_demovar_1) ($p_demovar_1) ($m_demovar_2) ($p_demovar_2) ($m_demovar_3) ($p_demovar_3) ($m_demovar_4) ($p_demovar_4)
    ($m_demovar_5) ($p_demovar_5) ($m_demovar_6) ($p_demovar_6) ($m_demovar_7) ($p_demovar_7) ($m_demovar_8) ($p_demovar_8) ($m_demovar_9) ($p_demovar_9)
    forvalues s = 1/$sourcenum {
        global m_demovar_`s' = abs($m_demovar_`s' - ($truevalue_`i')/100)
        global p_demovar_`s' = 0
    }
    post resultsbuffer ($m_demovar_1) ($p_demovar_1) ($m_demovar_2) ($p_demovar_2) ($m_demovar_3) ($p_demovar_3) ($m_demovar_4) ($p_demovar_4)
    ($m_demovar_5) ($p_demovar_5) ($m_demovar_6) ($p_demovar_6) ($m_demovar_7) ($p_demovar_7) ($m_demovar_8) ($p_demovar_8) ($m_demovar_9) ($p_demovar_9)
    }
}

*** Doing the average error for the primary benchmarks, unweighted

forvalues s = 1/9 {
    global epsilon = 0
    forvalues i = 1/$p {
        quietly summarize demovar_`i' [aweight=wperson] if source == `s'
        global mean = r(mean)
        global error_`i' = abs($mean - $truevalue_`i')/100
        global epsilon = abs($mean - $truevalue_`i')/100 + $epsilon
    }
    global epsilon = $epsilon*100/$p
    global averageerror_`s' = $epsilon
}

post resultsbuffer ($averageerror_1) ($p_demovar_1) ($averageerror_2) ($p_demovar_2) ($averageerror_3) ($p_demovar_3) ($averageerror_4) ($p_demovar_4)
($averageerror_5) ($p_demovar_5) ($averageerror_6) ($p_demovar_6) ($averageerror_7) ($p_demovar_7) ($averageerror_8) ($p_demovar_8) ($averageerror_9) ($p_demovar_9)

*** Doing the average for the non-primary benchmarks

drop demovar_1 - demovar_16

global num=0
foreach var of varlist marcats1 totalnum3 epspuwk1 cnsus384 cnsus433 cpslivq1 incomecat5 brftu3 nhis_acl1 numberofdrinks1 nhishs2 mripp2
mridl1 {
    global num = $num+1
}

```

```

clonevar demovar_ $num = `var'
}

*** copy the percentage X 100 for each of the benchmark questions below

global p = 0

*** Looping over the "true values"

foreach num of numlist 5650 3384 6080 4338 4146 7250 1511 7825 7745 3767 3193 7850 8900 {
    global num = `num'
    global truevalue = ($num)
    global p = $p+1
    global truevalue_ $p = ($truevalue)/100
}

forvalues s = 1/9 {
    global epsilon = 0
    forvalues i = 1/$p {
        quietly summarize demovar_ `i' [aweight=pwperson] if source == `s'
        global mean = r(mean)
        global error_ `i' = abs($mean)-${truevalue_`i'}/100
        global epsilon = abs($mean)-${truevalue_`i'}/100 + $sepsilon
    }
    global epsilon = $sepsilon*100/$p
    global averageerror_ `s' = $sepsilon
}

post resultsbuffer (Saverageerror_1) ($p_demovar_1) (Saverageerror_2) ($p_demovar_2) (Saverageerror_3) ($p_demovar_3) (Saverageerror_4) ($p_demovar_4)
(Saverageerror_5) ($p_demovar_5) (Saverageerror_6) ($p_demovar_6) (Saverageerror_7) ($p_demovar_7) (Saverageerror_8) ($p_demovar_8) (Saverageerror_9) ($p_demovar_9)

*** weighted analyses - non-primary benchmarks only

forvalues s = 1/9 {
    global epsilon = 0
    forvalues i = 1/$p {
        quietly summarize demovar_ `i' [aweight=post] if source == `s'
        global mean = r(mean)
        global error_ `i' = abs($mean)-${truevalue_`i'}/100
        global epsilon = abs($mean)-${truevalue_`i'}/100 + $sepsilon
    }
    global epsilon = $sepsilon*100/$p
    global averageerror_ `s' = $sepsilon
}

post resultsbuffer (Saverageerror_1) ($p_demovar_1) (Saverageerror_2) ($p_demovar_2) (Saverageerror_3) ($p_demovar_3) (Saverageerror_4) ($p_demovar_4)
(Saverageerror_5) ($p_demovar_5) (Saverageerror_6) ($p_demovar_6) (Saverageerror_7) ($p_demovar_7) (Saverageerror_8) ($p_demovar_8) (Saverageerror_9) ($p_demovar_9)

postclose resultsbuffer

save "Mode 04 analyzed.dta", replace

*****
*** Bootstrapping average errors program for un-weighted data
*** change the demovar list above in order to change it from non-primary demographics
*****

capture program drop survey_eval

program survey_eval, rclass
    version 10.1
    quietly {
        global epsilon = 0
        propweights
        replace weight = pwperson
        forvalues i = 1/$p {
            summarize demovar_ `i' [aweight=weight] if source==9
            global mean_ `i' = r(mean)
            global error_ `i' = abs($mean_ `i')-${truevalue_`i'}/100
            global epsilon = abs($mean_ `i')-${truevalue_`i'}/100 + $sepsilon
        }
        global epsilon = $sepsilon*100/$p
        return scalar epsilon = $sepsilon
    }
end

bootstrap mean=r(epsilon) , reps(100): survey_eval

*****
* Bootstrapping the weighted data for non-primary demographics
*****

drop demovar_1 - demovar_16

gen prop = pwperson
gen post = .

```

```
capture program drop survey_eval
```

```
program survey_eval, rclass
  version 10.1
  quietly {
    global epsilon = 0
    propweights
```

```
*** CHANGE SOURCE
```

```
replace pwperson = 0 if source~=1
*Specify the original weighting variable (which might be all ones)
global weight_var = "pwperson"
```

```
*Build flag variables for each category we are weighting for
```

```
                                foreach xn of numlist 1/31 {
  capture drop flag_`xn'
  gen flag_`xn' = 0
}
```

```
*Set the flags to 1 if person belongs in category
```

```
replace flag_1 = 1 if agegender==1
replace flag_2 = 1 if agegender==2
replace flag_3 = 1 if agegender==3
replace flag_4 = 1 if agegender==4
replace flag_5 = 1 if agegender==5
replace flag_6 = 1 if agegender==6
replace flag_7 = 1 if agegender==7
replace flag_8 = 1 if agegender==8
replace flag_9 = 1 if agegender==9
replace flag_10 = 1 if agegender==10
replace flag_11 = 1 if agegender==11
replace flag_12 = 1 if agegender==12
```

```
replace flag_13 = 1 if edugender==1
replace flag_14 = 1 if edugender==2
replace flag_15 = 1 if edugender==3
replace flag_16 = 1 if edugender==4
replace flag_17 = 1 if edugender==5
replace flag_18 = 1 if edugender==6
replace flag_19 = 1 if edugender==7
replace flag_20 = 1 if edugender==8
replace flag_21 = 1 if edugender==9
replace flag_22 = 1 if edugender==10
```

```
replace flag_23 = 1 if racecats==1
replace flag_24 = 1 if racecats==2
replace flag_25 = 1 if racecats==3
```

```
replace flag_26 = 1 if hispanic==1
replace flag_27 = 1 if hispanic==2
```

```
replace flag_28 = 1 if regions==1
replace flag_29 = 1 if regions==2
replace flag_30 = 1 if regions==3
replace flag_31 = 1 if regions==4
```

```
*Specify targets for each category
```

```
*Weighting targets, from the 2004 CPS ASEC, targets go in this order: 12 age X gender categories; 8 education X
gender categories; 4 race / ethnicity categories; and 4 regional categories
```

```
global target_1 = .0661
global target_2 = .0913
global target_3 = .1003
global target_4 = .0935
global target_5 = .0631
global target_6 = .0689
global target_7 = .0635
global target_8 = .0913
global target_9 = .1027
global target_10 = .0977
global target_11 = .069
global target_12 = .0925
```

```
global target_13 = .0797
global target_14 = .1511
global target_15 = .1104
global target_16 = .0965
global target_17 = .0454
global target_18 = .0782
global target_19 = .1664
global target_20 = .1253
global target_21 = .1080
global target_22 = .0389
```

```
global target_23 = .8202
global target_24 = .1147
```

```

global target_25=.0651

global target_26=.1238
global target_27=.8762

global target_28=.1863
global target_29=.2247
global target_30=.3592
global target_31=.2298

*Set the flags equal to the true value if a person is missing data from that category

forvalues i = 1/12 {
    replace flag_`i' = ${target_`i'} if agegender == .
}

forvalues i = 13/22 {
    replace flag_`i' = ${target_`i'} if edugender == .
}

forvalues i = 23/25 {
    replace flag_`i' = ${target_`i'} if racecats == .
}

forvalues i = 26/27 {
    replace flag_`i' = ${target_`i'} if hispanic == .
}

forvalues i = 28/31 {
    replace flag_`i' = ${target_`i'} if regions == .
}

*Gen a new weight variable
capture drop new_wgt

*** CHANGE SOURCE

gen new_wgt = ${weight_var} if source == 1

*Calculate total target

*** CHANGE SOURCE

summarize ${weight_var} if source == 1
global totwgt = r(mean) * r(N)

*Specify attenuation of adjustment
global attenuate = 0.25

***Specify varlist
global flag_cps_list = "flag_cps_1-flag_cps_31"
foreach xn of numlist 1/31 {
    capture drop flag_cps_`xn'
    gen flag_cps_`xn' = .
}

*****
*Iterate
*****
* You can change the number of iterations below if you'd like
*****

foreach rep of numlist 1/50 {
    *display_newline(2) "Iteration = `rep' " _newline(2)
    foreach xn of numlist 1/31 {

*** CHANGE SOURCE

        quietly summarize flag_`xn' [iweight = new_wgt] if source == 1
        global mean = r(mean)
        global adj = 1 + Sattenuate * ( ${target_`xn'}/$mean - 1)
        quietly replace new_wgt = $adj * new_wgt if flag_`xn' == 1

*** CHANGE SOURCE

        quietly summarize new_wgt if source == 1
        global new_totwgt = r(mean) * r(N)
        global adj_totwgt = Stotwgt / $new_totwgt
        quietly replace new_wgt = ${adj_totwgt} * new_wgt
        quietly replace flag_cps_`xn' = flag_`xn' - ${target_`xn'}

    }

}

order new_wgt
quietly drop flag_1-flag_31 flag_cps_1-flag_cps_31
replace weight = new_wgt
forvalues i = 1/$p {

*** CHANGE SOURCE

    summarize demovar_`i' [aweight=weight] if source==1
    global mean_`i' = r(mean)
    global error_`i' = abs(${mean_`i'} - ${truevalue_`i'})/100
    global epsilon = abs(${mean_`i'} - ${truevalue_`i'})/100 + $epsilon

```

```
    }  
    global epsilon = Sepsilon*100/$p  
    return scalar epsilon = Sepsilon  
  }  
end  
  
bootstrap mean=r(epsilon) , reps(100) strata(source): survey_eval
```