STUDY SERIES
(Survey Methodology #2010-15)

Measuring Intent to Participate and Participation in the 2010 Census and Their Correlates and Trends: Comparisons of RDD Telephone and Non-probability Sample Internet Survey Data

Josh Pasek[1]
Jon A. Krosnick

[1] Stanford University

Statistical Research Division
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: December 28, 2010

Measuring Intent to Participate and Participation in the 2010 Census and Their Correlates and Trends: Comparisons of RDD Telephone and Non-probability Sample Internet Survey Data

Abstract:

This study explored whether probability sample telephone survey data and data from nonprobability sample Internet surveys yielded similar results regarding intent to complete the 2010 Census form and actual completion of the form, the correlates of these variables, and changes in these variables and their correlates over time. Using data collected between January and April, 2010, the telephone samples were more demographically representative of the nation's population than were the Internet samples after post-stratification. Furthermore, the distributions of opinions and behaviors were often significantly and substantially different across the two data streams, as were relations between the variables and changes over time in the variables. Thus, research conclusions would often be different depending on which data stream was used. Because the telephone data collection methodology rests on well-established theory of probability sampling and produced the most demographically representative samples, the substantive results yielded by these data may also be more accurate than the substantive results generated with the non-probability sample Internet data.

Jon Krosnick is University Fellow at Resources for the Future. Correspondence should be addressed to Josh Pasek or Jon Krosnick, Stanford University, 432 McClathy Hall, 450 Serra Mall, Stanford, CA 94305, josh@joshpasek.com or Krosnick@stanford.edu.

In 2009 and 2010, the U.S. Census Bureau commissioned the collection of two parallel data streams to monitor public reactions to the 2010 Census. In both a random digit dial (RDD) telephone survey conducted by the Gallup Organization and a series of non-probability Internet surveys administered by E-Rewards, respondents reported their intent to complete the Census form, whether they had completed it, a variety of purported predictors of intention and behavior, and demographic characteristics. These data were collected to track changes over time and to identify opinions that might enhance or reduce Census form completion.

This paper explores whether the results generated by the telephone and Internet data streams are equivalent. If the two data streams support identical conclusions about distributions of, changes over time in, and relations between the variables measured, then future Census Bureau efforts can choose to employ just one of these methods, perhaps the one that generates the most cases at the least cost per case. But if the data streams yielded different results, the Bureau must decide whether in the future, it makes sense to collect both or just one, and this decision can be facilitated by knowledge about which data stream was most accurate in describing the nation's population.

This paper outlines the results of our empirical analyses of these issues. Specifically, we report answers to five questions:

1. Did the two data streams differ in their degree of demographic representativeness of the nation's adult population?

2. Did the two data streams produce similar distributions of opinions and behaviors?

3. Were the relations between variables within the two data streams similar?

4. Were predictors of intent to complete the Census form and completion of the Census form similar across the two data streams?

5. Did measurements of opinions and behaviors in the two data streams yield similar patterns of change over time?

To answer each of these questions, the results of an analysis using the probability sample telephone survey data were compared with an identical analysis using the non-probability sample Internet survey data. We compared both data streams with known population benchmarks to determine which data collection method yielded the most representative samples.

We begin below by describing the methods of data collection that were employed to yield each data stream. Then, we describe the measures administered and the analytic strategies employed. Finally, we report our findings and describe their implications.

Methods

Data

Telephone Data Collection

The Gallup Organization (see Gallup, 2010) conducted interviews each day using a rolling cross-sectional design with Random Digit Dialing (RDD) of both landline and cellular telephone numbers. Gallup aimed to complete 1,000 interviews per day. Telephone numbers with area codes in the 50 states and the District of Columbia were each dialed a minimum of three times, or until someone answered. If a person was unavailable to be interviewed at the time, additional calls could be scheduled up to two months later. Each number was tried up to eight times total before it was dropped from the active sample. Calls were conducted between 4:00 and 9:30 PM in each time zone on weeknights, between 10:00AM and 3:00 PM in each time zone on Saturdays, and between noon and 9:30 PM in each time zone on Sundays. The AAPOR RR3 response rate for the telephone survey was 19.4 percent over the period examined in this analysis (American Association for Public Opinion Research, 2009).[1]

Quotas for sex were set within Census regions for each night's calling. Before the night's quota for female respondents within a region had been met, interviewers asked to speak with "the person, 18 years of age or older, living in [the] household, with the most recent birthday." After the quota for female respondents within a region had been met on a

[1] The value of e was fixed at .39 for this computation. AAPOR RR1 was 10.2 percent.

day, interviewers would ask to speak with “the man, 18 years of age or older, living in [the] household, with the most recent birthday” in that region.

Each day between December 3, 2009, and April 24, 2010, a group of individuals (ranging in number between 180 to 775 people) was randomly chosen to answer the questions used for the analyses reported here.[2] On most days, between 200 and 250 respondents were asked these questions; the median number of interviews completed per day that asked these questions was 216. Typically, 21 percent of the sampled telephone numbers were assigned to be asked the Census Bureau’s questions on any given day. This proportion was increased on certain dates to 31 percent or, on one occasion, to 61 percent of sampled telephone numbers.

Internet Data Collection

Panel recruitment. The Internet survey respondents were members of the E-Rewards panel. To recruit members of its panel, E-Rewards partnered with a variety of commercial companies, such as airlines, video stores, book sellers, and electronics retailers. Consumers who had relationships with the participating companies (e.g., members of the British Airways frequent flier program) were invited to join the panel and complete Internet surveys regularly. Only individuals who had a relationship with affiliated organizations could be invited to join E-Rewards, and only invited individuals were eligible to join. E-Rewards regularly examined the demographic profile of its panel members and sought out partnerships with companies that catered to demographic groups that were underrepresented. In exchange for completing surveys, E-Rewards members were rewarded with points that could be redeemed for prizes.

Sampling from the panel. An Internet survey was fielded each week between October 27, 2009, and December 8, 2009, and between January 18, 2010, and April 19, 2010. During each week, a series of stratified random samples of panel members who lived in the U.S. were invited to complete the Census questionnaire. The sampling was done to achieve two goals: (1) to obtain completed questionnaires from 900 people, (2) to ensure that an analyzable

[2] No telephone interviews were conducted on March 19, 2010, or April 4, 2010. Larger sampling proportions were specified on subsequent days to make up for the number of respondents who would have been interviewed on those days.

group of at least 100 individuals were each White, Black, Hispanic, and Asian-American. To achieve these goals, the sample of 900 people included an oversample of approximately 59 Asian-Americans in addition to a sample of individuals whose demographics resembled the nation's population in terms of sex, age, race, education, and region.[3]

During each week, a new sample of panel members was drawn in a way guided by the characteristics of people who had completed the week's survey so far (DraftFCB, 2010). On the first day of each week's survey (Tuesday), invitations were sent to individuals who were members of demographic groups that typically have relatively low response rates (e.g., low-income minorities). On each subsequent day, E-Rewards staff examined the demographic profile of the individuals who had competed the questionnaire and then drew the next day's sample so that it over-sampled individuals in demographic categories that were under-represented at that point. This was done by adjusting the sampling proportions within strata defined by sex, race, income, Census division, education, and age. Each panel member could participate in only one week's survey.

Invited individuals could complete the survey at any time between when they received their invitations and the end of the following Monday.

Invitations to complete each week's Internet survey were designed to elicit a minimum of 100 completed interviews in each of the nine Census divisions and a minimum of 100 completed interviews in each of four racial/ethnic categories: White, Black, Hispanic, and Asian.[4] Invitations were also designed to recruit a sample that was demographically representative with respect to education, household income, and the cross-tabulation of age by gender. Quotas were used to cap the number of individuals in each demographic category who completed the survey, such that no group would exceed its population proportion. If the minimum number of interviews in racial/ethnic categories or Census divisions had not been reached by the fifth day of each week (Saturday), E-Rewards staff relaxed quotas on

[3]Targets for these groups were specified using data from the 2008 American Community Survey.

[4]Asian respondents were over-represented relative to their population proportion. The oversampled individuals were invited to complete each week's survey in a sampling procedure separate from the sampling used to select all other potential respondents and was not subjected to any of the quotas imposed on the full population sample.

education, income, and age by gender, so that some categories could be over-represented by up to ten percent of their initial limits.

Comparing Across Data Streams

Differences between the data collection methods used by Gallup, in their telephone survey, and E-Rewards, in their Internet survey, mean that the two data streams were not directly comparable. The Gallup procedures were designed to yield a representative sample each day, whereas the E-Rewards Internet surveys were designed to yield a demographically representative sample each week. Therefore, data collected in the two data streams on a single day could not be compared to one another. We therefore compared the weekly Internet survey data to the telephone data collected during the same week.

Furthermore, different numbers of interviews were completed each day in the two data streams: the number of interviews conducted per day by telephone was more consistent than the number of respondents who completed the Internet survey each day. We therefore conducted our analyses so that the number of interviews completed on a given day within a week was functionally the same across the two data streams.

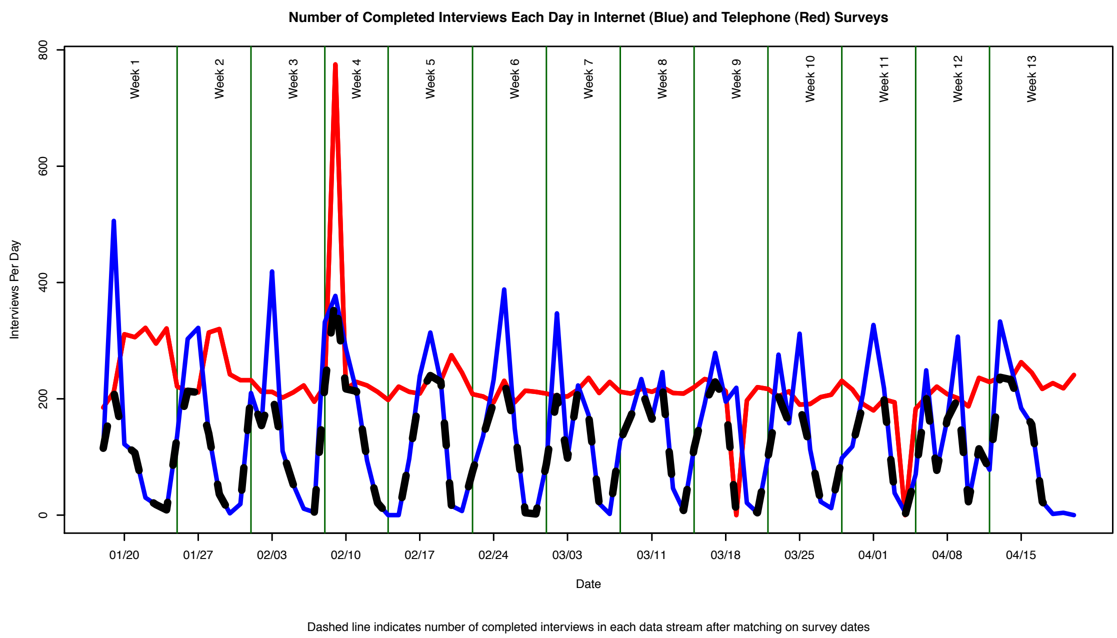
Date Overlap

The telephone and Internet data were not collected during identical time periods. Between October 27, 2009, and December 2, 2009, only Internet data were collected. Between December 9, 2009, and January 17, 2010, only telephone data were collected. Both streams of data were collected for a period of five days in December and during 13 weeks between January 18 and April 19, 2010. The analyses reported here focus on the latter period of overlap between the two data streams.

Matching Numbers of Interviews Each Day

The number of interviews completed per day in each data stream are plotted in Figure 1, where vertical lines divide the weeks from one another. Although the number of respondents interviewed each day by telephone was relatively consistent over time,

Figure 1.



the number of completed Internet questionnaires varied considerably across days. We implemented "downweighting" so that the effective number of completed interviews obtained on each day was the same across the telephone and Internet data streams (c.f. Heckman, Ichimura, & Todd, 1998).

To downweight, we identified the data stream that yielded fewer completed interviews on each day. We then divided the number of completed interviews in the smaller data stream that day by the number of completed interviews in the larger data stream that day. This ratio was then treated as a weight applied to the cases in the bigger data stream. The smaller data stream received a weight of 1.0 for that day. As a result, the effective sample size was the same in both data streams on each individual day, and the effective sample size varied from day to day. The dashed line in Figure 1 shows the effective sample size for each day in both data streams after downweighting.

This technique is not without drawbacks. First, by downweighting data collected in one data stream each day, we treat that data stream as if it involved responses from fewer people. A smaller effective sample size means larger standards errors will be computed for estimates, and that means our comparisons across data streams may be biased against finding significant differences between the data streams.

When any data were downweighted in the Internet survey stream during a given week, this could have compromised the demographic representativeness of that week's Internet data. However, as shown in Figure 1, the blue line, representing the number of completed Internet surveys, was above the red line, showing the number of completed telephone surveys, on a minority of days, so this sort of compromising happened rarely. To see whether this analytic approach affected our conclusions, we compared the results obtained when matching the two data streams in terms of number of completes per day with results obtained when not doing this matching.

Spanish Language Interviews

The telephone and Internet data streams also differed with regard to the languages in which questions were asked. In the telephone surveys, respondents who spoke Spanish could choose to be interviewed in Spanish, and some did. The Internet survey respondents all answered the questions in English and were not offered the opportunity to choose administration in Spanish. To address potential differences in results due to the use of different languages in the two modes, we conducted a series of analyses dropping telephone respondents who were interviewed in Spanish (N=632). The results were the same as those obtained when including these respondents.

Weighting

Base Weights

In telephone surveys, all American adults do not have equal probabilities of selection. Households that can be reached on more telephone lines are more likely to be reached than those with fewer working telephone lines. And within households, the probability that each individual will be selected is inversely proportional to the number of eligible adults living in the household. Therefore, base weights were computed for landline respondents by dividing the number of adults in each household by the number of landlines that could reach that household and rescaling those weights so they had a mean of one.

Because many individuals could be reached on both landline and cellular telephones, the base weights were designed to account for this as well. The 2008 National Health Interview Survey (NHIS) provided benchmarks of the proportions of the American population in various categories of cellular telephone and landline telephone use. Respondents in the NHIS could be divided into those who only used cellular telephones, those who could only be reached at home on landline telephones, those who had both but were primarily cellular telephone users, and those who had both who were not primarily cellular telephone users.[5] The telephone survey included questions designed to similarly demarcate respondents. Using responses to these questions, respondents reached via landline and cellular telephone were weighted to match the proportions observed in the NHIS.

Base weights for the telephone survey combined household adjustments for the landline telephone sampling with the weights used to combine the samples reached by landline and cellular telephone. To create these weights, every individual who reported having access to at least one non-business landline was assigned an initial weight equal to the number of individuals in the household divided by the number of telephone lines in the household. Post-stratification was implemented to force the number of individuals in each of the telephone use categories each week to match the NHIS benchmarks.

Base weights ranged from .70 to 4.66. These base weights were used for all analyses reported below.

House Weights

In the telephone data stream, weights were produced by the Gallup Organization for each day's data. To create these weights, Gallup started with the base weight computed above and raked on age by gender, education by age, race by gender, and ethnicity by gender.

[5] This later category combined individuals who reported using both about the same amount and those who reported that the majority of their use was on landline telephones.

Gallup limited these weights to range from a minimum of .25 to a maximum of 3.

ANES-Style Weights

The telephone and Internet data were also weighted to match the population in terms of demographics. Specifically, we implemented a raking procedure that followed the recommendations of a blue-ribbon panel of experts assembled by the American National Election Studies (DeBell and Krosnick, 2009). The *anesrake* R package was used to select variables for raking and to correct for demographic discrepancies between both data streams and population benchmarks (Pasek, 2010a).

Weights were produced separately for the two data streams for each week of the simultaneous survey period. The anesrake algorithm was told to use all variables for which the sample differed from the population by a total of more than five percentage points across all variable categories. Demographic variables indicating sex, race, age category, Census region, and education level were considered for weighting. Base weights were used as a starting weight vector in the telephone data. The weights generated by the downweighting procedure and the base weights were multiplied by one another to produce a starting weight vector. Final weights were capped at five following each iteration. Targets for the ANES-style weights came from the December 2009 Current Population Survey.

Measures

Outcomes

Question wording and response options were not always identical in the telephone and Internet questionnaires. We coded responses to maximize comparability. In this section, we describe the question wordings and answer codings used to tap each construct in the surveys. Next to each construct name below is an indication of how similar the wording was across modes: an exact match, a close match, or a non-match.

R Plans to Complete the Census Form (Exact Match, except "Don't Know" offered on the Internet only)

Telephone. "How likely are you to participate in the 2010 Census? By participate, we mean fill out and mail in a Census form. Would you say you definitely will, probably will, might or might not, probably will not, or definitely will not participate?" Coding: Definitely will=1, Probably will=.75, Might or might not=.5, Probably will not=.25, Definitely will not=0.

Internet. "How likely are you to participate in the 2010 Census? By participate, we mean fill out and mail in a Census form. Would you say you..." Response choices were: "Definitely will," "Probably will," "Might or might not," "Probably will not," "Definitely will not," and "Don't Know." Two variables were created for this measure, one that treated all respondents who said "don't know" as having missing values when computing proportions, and one that treated "don't know" as a valid value when computing proportions. Coding: Definitely will=1, Probably will=.75, Might or might not=.5, Probably will not=.25, Definitely will not=0.

R Completed the Census Form (Close Match)

Telephone. Respondents were asked: "Did your household receive a census questionnaire delivered to you at your home?" Respondents who said that they had received a Census form at their homes were asked: "What have you done with your form?" Response options were "Not opened yet," "Opened but not started yet," "Started but not completed yet," "Completed but not mailed yet," "Mailed it back," "Threw it away," and "Or, Something else." Respondents who said that they had mailed it in were coded 1; all other responses were coded 0.

Internet. Respondents were asked: "Did your household receive a census questionnaire delivered to you at your home?" Respondents who said that they had received a Census form at their homes were asked: "What have you done with your form?" Response options were "Not opened yet," "Opened but not started yet," "Started but not completed yet," "Completed but not mailed yet," "Mailed it back," "Threw it away," "Other," and "Don't Know." Respondents who said that they had mailed it in were coded 1; all other responses were coded 0. Two variables were created for this measure, one that treated all respondents who said "don't know" as having missing values when computing proportions, and one that treated "don't know" as a valid value when computing proportions.

Predictors

The Census Could Help R (Exact Match, except for volunteered "benefit and harm" and presence of a "Don't Know" response option)

Telephone. "Do you believe that answering and sending back your Census form could personally benefit you in any way, personally harm you, or neither benefit nor harm?" Respondents who volunteered that the Census could benefit them or could both benefit and harm them were coded 1, and respondents who said the Census could harm them or could neither benefit nor harm them were coded 0.

Internet. "Do you believe that answering and sending back your census form could personally benefit you in any way, personally harm you, or neither benefit nor harm?" Offered response choices were: "Personally benefits me," "Personally harms me," "Neither benefits nor harms me," and "Don't Know." Respondents who said the Census could benefit them were coded 1, and respondents who gave any other answer were coded as 0. Two variables were created for this measure, one that treated all respondents who said "don't know" as having missing values when computing proportions, and one that treated "don't know" as a valid value when computing proportions.

The Census Could Harm R (Exact Match, except for volunteered "benefit and harm" and presence of a "Don't Know" response option)

Telephone. "Do you believe that answering and sending back your census form could personally benefit you in any way, personally harm you, or neither benefit nor harm?"

Respondents who said the Census could harm them or volunteered that it could both benefit and harm them were coded 1, and respondents who said the Census could benefit them or could neither benefit nor harm them were coded 0.

Internet. "Do you believe that answering and sending back your Census form could personally benefit you in any way, personally harm you, or neither benefit nor harm?" Internet response choices were: "Personally benefits me," "Personally harms me," "Neither benefits nor harms me," and "Don't Know." Respondents who said that the Census could harm them were coded 1, and respondents who said the Census could benefit them or could neither benefit nor harm them were coded 0. Two variables were created for this measure, one that treated all respondents who said "don't know" as having missing values when computing proportions, and one that treated "don't know" as a valid value when computing proportions.

The Census is Used to Locate People Who Are in the U.S. Illegally (Close Match)

Telephone. "Do you believe the Census is used to locate people living in the country illegally?" People who said that the Census was not used to locate illegal residents were coded 0, and people who said that the Census was used for that purpose were coded 1.

Internet. "People have different ideas about what the Census is used for. Below are some ideas that people have. As you read each one, please indicate - yes or no - whether you think that the Census is used for that purpose. Is the Census used... – To locate people living in the country illegally?" Response choices were: "Yes," "No," and "Don't Know." People who said that the Census was not used to locate illegal residents were coded 0, and people who said that the Census was used for that purpose were coded 1. Two variables were created for this measure, one that treated all respondents who said "don't know" as having missing values when computing proportions, and one that treated "don't know" responses as equivalent to "no" answers.

Can Trust the Confidentiality Promise (Close Match, Differences: the Internet question did not include a middle response category but did offer a "Don't Know" response option)

Telephone. "I am going to read some opinions about the Census. As I read each one, tell me if you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree. – The Census Bureau promise of confidentiality can be trusted." Coding: Strongly agree=1, Agree=.75, Neither agree nor disagree / Don't know (vol.)=.5, Disagree=.25, Strongly disagree=0.

Internet. "Below are some opinions that some people may have about the Census. As you read each one, please indicate if you strongly agree, agree, disagree, or strongly disagree. – The Census Bureau promise of confidentiality can be trusted." Response choices were: "Strongly Agree," "Agree," "Disagree," "Strongly Disagree," "No Opinion," and "Don't Know." Respondents who said "no opinion" or "don't know" were placed at the midpoint. Coding: Strongly agree=1, Agree=.75, No opinion / Don't know=.5, Disagree=.25, Strongly disagree=0.

R Doesn't Have Time to Fill Out the Form (Close Match, Differences: Internet question did not include a middle response category, but did include a "Don't Know" response option)

Telephone. "I am going to read some opinions about the Census. As I read each one, tell me if you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree. – It takes too long to fill out the Census information, I don't have time." Respondents who volunteered that they did not know were placed at the midpoint. Coding: Strongly agree=1, Agree=.75, Neither agree nor disagree / Don't know (vol.)=.5, Disagree=.25, Strongly disagree=0.

Internet. "Below are some opinions that some people may have about the Census. As you read each one, please indicate if you strongly agree, agree, disagree, or strongly disagree. – It takes too long to fill out the Census information, I don't have time." Response choices were: "Strongly Agree," "Agree," "Disagree," "Strongly Disagree," "No Opinion," and "Don't

Know." Respondents who said "no opinion" or "don't know" were placed at the midpoint. Coding: Strongly agree=1, Agree=.75, No opinion / Don't know=.5, Disagree=.25, Strongly disagree=0.

Importance of Counting Everyone (Close Match, Differences: Internet question did not include a middle response category, but did include a "Don't Know" response option)

Telephone. "I am going to read some opinions about the Census. As I read each one, tell me if you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree. – It is important for everyone to be counted in the Census." Respondents who volunteered that they did not know were placed at the midpoint. Coding: Strongly agree=1, Agree=.75, Neither agree nor disagree / Don't know (vol.)=.5, Disagree=.25, Strongly disagree=0.

Internet. "Below are some opinions that some people may have about the Census. As you read each one, please indicate if you strongly agree, agree, disagree, or strongly disagree. – It is important for everyone to be counted in the Census." Response choices were: "Strongly Agree," "Agree," "Disagree," "Strongly Disagree," "No Opinion," and "Don't Know." Respondents who said "no opinion" or "don't know" were placed at the midpoint. Coding: Strongly agree=1, Agree=.75, No opinion / Don't know=.5, Disagree=.25, Strongly disagree=0.

R's Participation Does Not Matter (Close Match, Differences: Internet question did not include a middle response category, but did include a "Don't Know" response option.)

Telephone. "I am going to read some opinions about the Census. As I read each one, tell me if you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree. – I just don't see that it matters much if I personally fill out the Census or not." Respondents who volunteered that they did not know were placed at the midpoint. Coding: Strongly agree=1, Agree=.75, Neither agree nor disagree / Don't know (vol.)=.5, Disagree=.25, Strongly disagree=0.

Internet. “Below are some opinions that some people may have about the Census. As you read each one, please indicate if you strongly agree, agree, disagree, or strongly disagree. – I just don’t see that it matters much if I personally fill out the Census or not.” Response choices were: “Strongly Agree,” “Agree,” “Disagree,” “Strongly Disagree,” “No Opinion,” and “Don’t Know.” Respondents who said “no opinion” or “don’t know” were placed at the midpoint. Coding: Strongly agree=1, Agree=.75, No opinion / Don’t know=.5, Disagree=.25, Strongly disagree=0.

Demographics

Female (Non-Match, Same Concept)

Telephone. Interviewers recorded whether each respondent was female (coded 1) or male (coded 0).

Internet. “What is your gender?” Response choices were “Male” and “Female.” Responses were coded 1 for female respondents and 0 for male respondents.

Hispanic (Non-Match, Same Concept)

Telephone. “Are you, yourself, of Hispanic origin or descent such as Mexican, Puerto Rican, Cuban, or other Spanish background?” Respondents who said yes were coded 1; all others were coded 0.

Internet. “Are you Hispanic or Latino?” Response choices were “Yes,” “No,” and “Prefer not to say.” Respondents who said yes were coded 1; all others were coded 0. Respondents who said that they would prefer not to say had their interviews terminated.

White Non-Hispanic (Non-Match, Same Concept)

Telephone. “What is your race? Are you White, African-American, Asian, or some other race?” Only one answer was recorded. Respondents who said White and did not identify themselves as Hispanic were coded 1, and all others were coded 0.

Internet. “What is your race? (Select all that apply.)” Response choices were “White,” “Black or African-American,” “Asian,” “Native Hawaiian or Pacific Islander,” “American Indian or Alaska Native,” “Other,” and “Prefer not to say.” Respondents who said that they would prefer not to say had their interviews terminated. Respondents were coded 1 if they said White and were not coded as Hispanic; all others were coded 0.

Black Non-Hispanic (Non-Match, Same Concept)

Telephone. “What is your race? Are you White, African-American, Asian, or some other race?” Respondents could make only a single selection. Responses were coded 1 for African-American respondents who were not Hispanic and 0 otherwise.

Internet. “What is your race? (Select all that apply.)” Response choices were “White,” “Black or African-American,” “Asian,” “Native Hawaiian or Pacific Islander,” “American Indian or Alaska Native,” “Other,” and “Prefer not to say.” Respondents who said they would prefer not to say had their interviews terminated. Respondents who said they were Black or African-American and were not Hispanic were coded 1, and all others were coded 0.

Age (Non-Match, Same Concept)

Telephone. “Please tell me your age.” Interviewers coded responses in four categories: “18-24,” “25-44,” “45-64,” and “65+.”

Internet. “What is your age?” Response choices were “Under 18 years old,” “18-24,” “25-34,” “35-44,” “45-54,” “55-64,” “65 or older,” and “Prefer not to say.” Respondents who were under 18 or who selected that they would prefer not to say had their interviews terminated. Responses were recoded into four categories to match the telephone survey categories: “18-24,” “25-44,” “45-64,” and “65 or older.”

Education (Close Match)

Telephone. “What is your highest completed level of education?” Responses were initially recorded in six categories: “Less than high school diploma,” “High school degree or diploma," "Technical/Vocational school," "Some college," "College graduate," and "Postgraduate work or degree." Responses were recoded into five categories: "Less than high school" ("Less than high school diploma"), "High school degree" ("High school degree or diploma"), "Some college" ("Technical/Vocational school" and "Some college"), "College graduate," and "Graduate" ("Postgraduate work or degree").

Internet. "What is the highest grade or year of regular school you completed?" Response choices were "Did not complete high school," "High school graduate," "Some college" "College graduate," "Post graduate education," and "Prefer not to say." Respondents who said they would prefer not to say had their interviews terminated. Responses were recoded into five categories: "Less than high school" ("Did not complete high school"), "High school degree" ("High school graduate"), "Some college," "College graduate," and "Graduate" ("Post graduate education").

Married (Non-Match, Same Concept)

Telephone. "What is your current marital status?" Answer choices were "Single/Never been married," "Married," "Separated," "Divorced," "Widowed," and "Domestic partnership/Living with partner (not legally married)." Respondents who said they were married or living with a domestic partner were coded as 1, and all others were coded 0.

Internet. "What is your marital status?" Response choices were "Married," "Widowed," "Divorced," "Separated," "Single, never married," and "Prefer not to say." Respondents who said they were married were coded as 1, and all others were coded 0. Respondents who reported that they preferred not to say were dropped from analyses with this variable.

English Speaking Household (Exact Match)

Telephone. "What language is spoken most often in this household?" Response options were "English," "Spanish," "An Asian or Pacific Islander Language," and "Some Other Language." Respondents who chose English were coded 1, and all others were coded 0.

Internet. "What language is spoken most often in this household?" Response options were "English," "Spanish," "An Asian or Pacific Islander Language (e.g. Chinese, Japanese, Tagalog, Vietnamese)," "Other," and "Prefer not to say." Respondents who chose English were coded 1, and all others were coded 0. Respondents who reported that they preferred not to say had their interviews terminated.

Spanish Speaking Household (Exact Match)

Telephone. "What language is spoken most often in this household?" Responses were coded as: "English," "Spanish," "An Asian or Pacific Islander Language," and "Some Other Language." Respondents who chose Spanish were coded 1, and all others were coded 0. Respondents who reported that they preferred not to say had their interviews terminated.

Internet. "What language is spoken most often in this household?" Response choices were "English," "Spanish," "An Asian or Pacific Islander Language (e.g. Chinese, Japanese, Tagalog, Vietnamese)," "Other" and "Prefer not to say." Respondents who chose Spanish were coded 1, and all others were coded 0. Respondents who reported that they preferred not to say had their interviews terminated.

R Owns Residence (Extremely Close Match - Note Change in Order of Response Options in Question)

Telephone. "Do you own or rent your home?" Responses choices were "Rent," "Own," or "Other." People who said they owned their residences were coded 1, and all others were coded 0.

Internet. "Do you rent or own your home?" Response choices were "Rent," "Own," "Other" and "Prefer not to say." People who said they owned their residences were coded 1, and all others were coded 0. Respondents who said that they preferred not to say were dropped from analyses with this variable.

R Rents Residence (Extremely Close Match - Note Change in Order of Response Options in Question)

Telephone. "Do you own or rent your home?" Responses were coded as "Rent," "Own," or "Other." People who said they rented their residences were coded 1, and all others were coded 0.

Internet. "Do you rent or own your home?" Response choices were "Rent," "Own," "Other," and "Prefer not to say." People who said they rented their residences were coded 1, and all others were coded 0. Respondents who reported that they preferred not to say were dropped from analyses with this variable.

Number of Persons in the Household (Non-Match, Same Concept)

Telephone. "Including yourself, how many adults 18 years of age or older live in this household?" and "How many children under the age of 18 are living in your household?" These two answers were summed, and all values greater than six were recoded as six.

Internet. "Including yourself, how many people live in your household?" A drop-down box offered answer choices from 1 to 9, "10 or more," and "Prefer not to say." Responses of larger than six were recoded to six. Respondents who reported that they preferred not to say were dropped from analyses with this variable.

Any Children in the Household (Non-Match, Same Concept)

Telephone. "How many children, under the age of 18, are living in your household?" People who answered one or more were coded 1, and all others were coded 0.

Internet. "Do you have children who are under 18 living at home with you?" Response choices were "Yes," "No," and "Prefer not to say." People who said yes were coded 1, and people who said no were coded 0. Respondents who reported that they preferred not to say were dropped from analyses with this variable.

Analysis Strategy

The telephone and Internet data were compared to one another in five ways. First, demographic variables were compared to population benchmarks to gauge the demographic representativeness of the two data streams. Representativeness was assessed separately for demographic variables used in quotas or weighting and for variables that were not used for quotas and weighting.

Second, distributions of opinions and behaviors were compared. For each variable, we computed the average difference between the proportion of respondents in each response category across all 13 weeks. We made these comparisons various ways: with and without weighting, and with and without matching the number of interviews per day across the data streams.

Third, we estimated the parameters of regression equations predicting intention to complete the Census form and actual form completion using all demographic and substantive variables. This allowed us to determine whether predictors in the two data streams differed significantly from one another.

Fourth, we compared the trends of the variables in the two data streams over time. We examined changes in the proportion of respondents choosing each response option. Correlations over weeks between the two data streams in terms of these proportions were computed to describe the closeness of correspondence between the trends documented. χ^2 tests were conducted and parameters of regression equations were estimated to determine whether the trends over time for the two data streams were statistically significantly different from one another.

Finally, relations between variables in the two data streams were examined. All pairs of variables in the telephone data stream were correlated with one another, and these correlations were compared to equivalent correlations in the Internet data.

Results

Demographic Representativeness

To assess demographic representativeness, data from the telephone and Internet surveys were compared to population benchmarks obtained from two datasets: the December 2009 Current Population Survey (CPS) and the 2008 American Community Survey (ACS). Benchmark values for sex, age, race, Hispanic identification, education, and marital status were identified using the CPS.[6] Benchmarks for the number of people living in each respondent's household, the proportion of Americans in households with children, the proportion of Americans that owned or rented their homes, and the primary language spoken in each American's household were computed using the ACS.

Two measures of representativeness were produced for each variable in each data stream. First, we computed the absolute difference between the proportion of respondents in the modal response category of each variable in the benchmark survey and the proportion in that category in each data stream at each wave. Then, we averaged all absolute errors for each variable across survey waves to determine the overall average error for the variable. Second, for each variable, we determined what percent of the absolute errors for each data stream were statistically different form zero ($p<.05$).

Summary statistics assessing representativeness were computed for two sets of variables: (1) the demographic data that were used to set quotas during data collection or were used for computing weights in either data stream (Census region (South), race (White), education level (High School Degree), sex (Female), marital status (Married), age (25 to 44), and number of individuals in the household (two persons))[7], and (2) the demographic data that were not used for quotas or weighting (the primary language spoken in the household (English), whether respondents owned their homes (Own), and whether any children lived in

[6]The individual weights, not the household weights, were used.

[7]Region, race, education, sex, and age were used to set targets during the Internet data collection. region and sex were used for quotas in the telephone survey. Number of adults in the household was used to produce base weights for the telephone survey. Region, race, education, sex, marital status, and age were used for post-stratification. House weights for the telephone data stream were produced using region, age, education, race, and sex. Modal categories in parentheses.

the household (Yes)). In each data stream, absolute errors for all variables were averaged across weeks to produce a summary statistic, and the proportion of significant differences across all weeks and all variables was computed for each data stream as well.

Representativeness for Variables Used for Quotas or Weighting

Using only base weights and not matching on interview dates, the telephone samples were less representative than the Internet samples. The telephone data's average absolute error was 6.8 percentage points (row 1, column 1, of Table 1), significantly larger than the Internet data stream's average error of 5.6 percentage points (Table 1, row 1, column 2; $p<.001$ difference).[8] More of the telephone data's absolute differences were significantly different from zero (89.01%) than was true for the Internet data's errors (82.42%), though the difference between data streams was not statistically significant (Table 1, row 1, columns 3 and 4). These errors, which ranged from zero to more than 18, are displayed in the first row of Figure 2; 62.6 percent of variable-weeks had errors greater than five percentage points in the telephone data stream, and 41.8 percent variable-weeks had errors of greater than five percentage points in the Internet data stream, a significant difference (p=.005).

When using house weights without matching on interview dates, the telephone samples were more representative than the Internet samples. The average absolute error in the telephone data was 4.8 percentage points (Table 1, row 3, column 1), whereas the Internet data – unadjusted because no weights were provided by E-Rewards – differed from benchmarks by an average of 5.6 percentage points (Table 1, row 3, column 2), a significant difference (p=.006). Differences between the proportions of respondents in modal demographic categories and population benchmarks were statistically significant for 69.2 percent of comparisons in the telephone data and 82.4 percent of comparisons in the Internet data (Table 1, row 3, columns 3 and 4), a significant difference (p=.04).

When using ANES-style weights and matching on dates,[9] the telephone samples were more representative than the Internet samples in terms of the variables used for quotas and

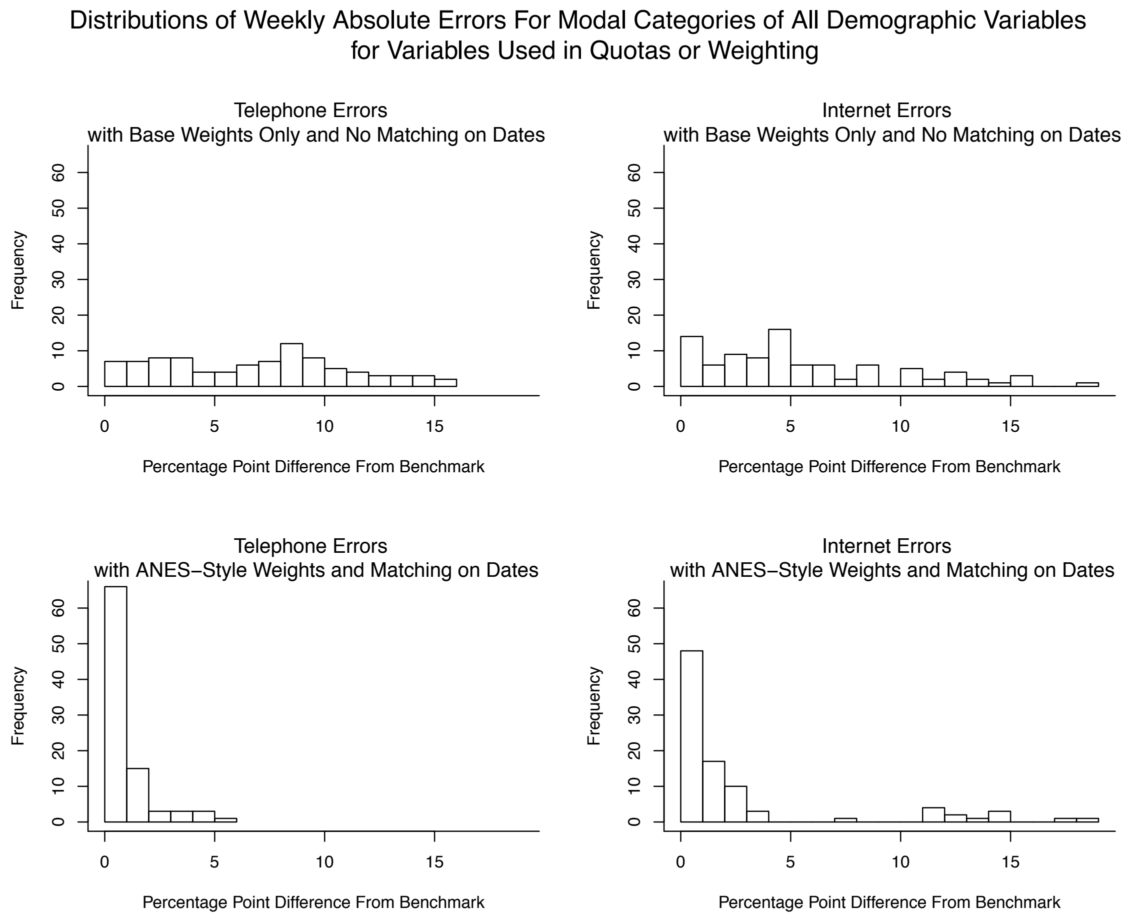
[8]The p values comparing accuracy in the two surveys were calculated using non-parametric bootstraps.
[9]Results matching on dates did not substantively differ from results without matching on dates.

Table 1:: Deviations from Benchmarks for Demographic Variables

Computation Method	*Demographic Variables Used in Quotas or Weighting*			
	Average Percentage Point Deviation		Percent of Variables with Significant Deviations	
	Telephone	Internet	Telephone	Internet
With Base Weights Only				
Not Matched on Survey Dates	6.84	5.58	89.01	82.42
Matched on Survey Dates	6.90	5.77	89.01	81.32
With House Weights for the Telephone Data				
Not Matched on Survey Dates	4.82	5.58	69.23	82.42
Matched on Survey Dates	4.83	5.77	64.84	80.22
With ANES-Style Weights				
Not Matched on Survey Dates	.75	2.57	14.29	39.56
Matched on Survey Dates	.67	2.62	12.09	28.57
	Demographic Variables Not Used in Quotas or Weighting			
With Base Weights Only				
Not Matched on Survey Dates	15.41	12.71	94.87	89.74
Matched on Survey Dates	15.05	12.65	87.18	87.18
With House Weights for the Telephone Data				
Not Matched on Survey Dates	12.38	12.71	92.31	89.74
Matched on Survey Dates	12.08	12.65	87.18	84.62
With ANES-Style Weights				
Not Matched on Survey Dates	12.01	13.83	100.00	92.31
Matched on Survey Dates	11.42	13.75	94.87	89.74

Notes: The figures in columns 1 and 2 averages of the absolute deviations between the percent of respondents in the modal category of each demographic variable and a benchmark survey's estimate of the proportion of people in that category, first averaged across demographics in each week's survey, and then averaged across weeks. Columns 3 and 4 report the average proportion of weeks when the survey's estimate of a proportion was significantly different from the benchmark (p<.05, two-tailed). Demographic variables used in quotas or weighting and their modal response categories were: region (South), race (White), education (High School Degree), sex (Female), marital status (Married), age (25 to 44), and number of individuals in the household (two persons). Demographic variables not used in quotas or weighting include: Primary language of the household (English), whether respondents rented or owned their homes (Own), and whether there were children in the household (Yes).

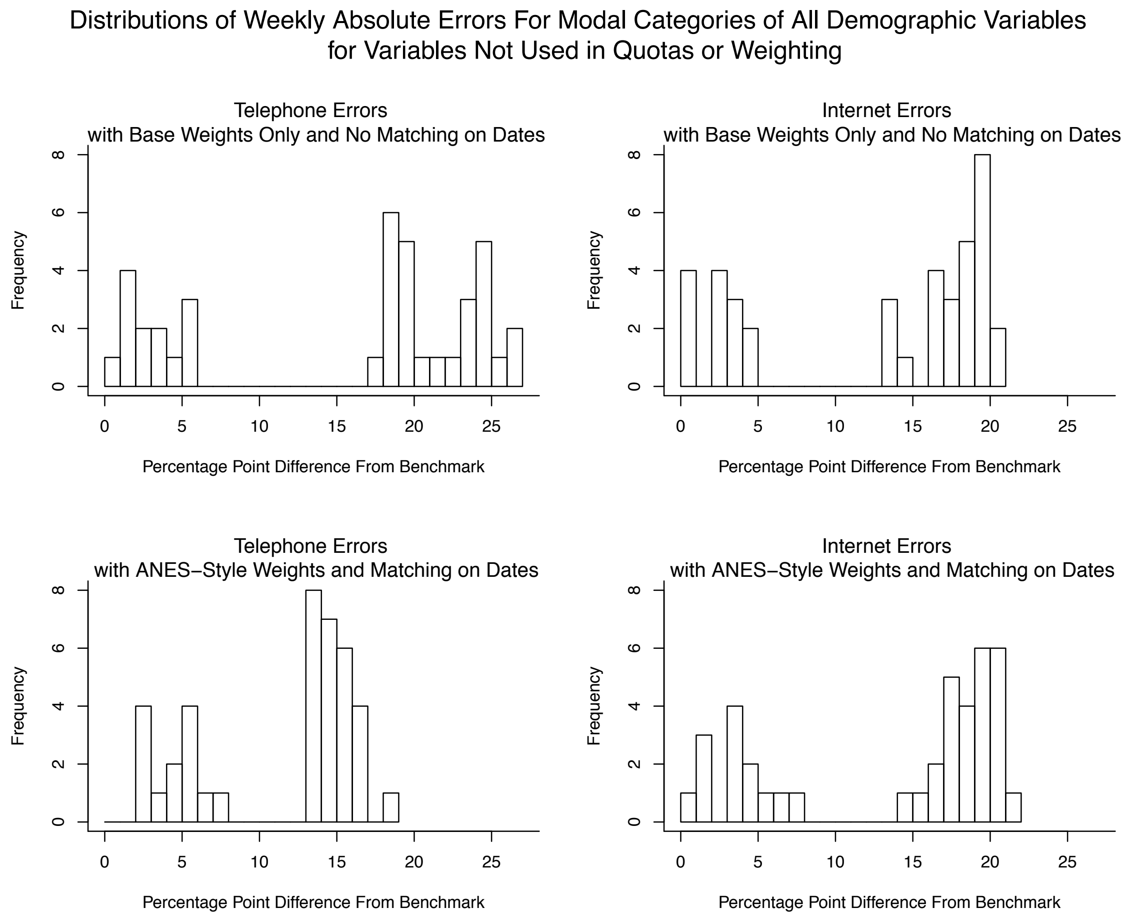
Figure 2.



weighting. The telephone data's average absolute error was just .67 percentage points (Table 1, row 6, column 1), compared to 2.6 percentage points for the Internet data (see Table 1, row 6, column 2), a significant difference (p<.001). Only 12.1 percent of the absolute errors were significantly different from zero in the telephone data (Table 1, row 6, column 3), whereas 28.6 percent of the absolute errors were significantly different from zero in the Internet data (Table 1, row 6, column 4), a significant difference (p=.006). Absolute errors, which ranged from zero to 18.0 percentage points, are displayed in the second row of Figure 2;[10]

[10]These variables do not perfectly equal benchmarks because of the caps imposed on the weights and because the number of persons per household was used for the telephone base weight but was not used in

Figure 3.



1.1 percent of variable-weeks had errors greater than five percentage points in the telephone data stream, and 14.3 percent of variable-weeks had errors of greater than five percentage points in the Internet data stream, a significant difference (p<.001).

Representativeness for Variables Not Used for Quotas or Weighting

With only base weights and without matching on dates, the telephone samples were less representative than the Internet samples for the variables not used in quotas or weighting.

post-stratification. The direction and significance of the difference between the data streams holds when the number of persons in the household is dropped as a measure.

The average absolute error in the telephone data was 15.4 percentage points (Table 1, row 7, column 1), whereas the average absolute error in the Internet data was 12.7 percentage points (Table 1, row 7, column 2), a significant difference (p<.001). 94.9 percent of the absolute errors were significant in the telephone data stream, whereas 89.7 percent were significant in the Internet data (Table 1, row 7, columns 3 and 4), which was not significantly different. The first row of Figure 3 displays these differences, which were in some cases as large as 26 percentage points.

The telephone and Internet data samples were equally representative when we compared the samples using the weights supplied with the Gallup data and no weighting for the Internet data. For the variables not used in quotas or weighting, absolute errors in the telephone data averaged 12.4 percentage points, compared to 12.7 percentage points in the Internet data (Table 1, row 9, columns 1 and 2), not significantly different. 92.3 percent of absolute errors were statistically significant in the telephone data, and 89.7 percent were significant in the Internet data (Table 1, row 9, columns 3 and 4), which was again not significantly different.[11]

Using the ANES-style weights and matching on dates, the telephone samples were more representative than the Internet samples. Whereas the average absolute error in the telephone data was 11.4 percentage points, the average absolute error in the Internet data was significantly larger (p<.001): 13.8 percentage points (Table 1, row 12, columns 1 and 2). Absolute errors were statistically significant for 94.9 percent of the variable-weeks in the telephone data and for 89.7 percent of the variable-weeks in the Internet data (Table 1, row 12, columns 3 and 4), which was not significantly different.

The second row of Figure 3 displays the absolute errors for each data stream with ANES-style weights. In the telephone data, the largest absolute errors were smaller when using the ANES-style weights (18.6 percentage points) than when using base weights only (26.9 percentage points), whereas in the Internet data, the largest absolute errors tended to

[11] As with all weighted data, standard errors for these analyses are inexact. Exact standard errors depend on a series of assumptions that a researcher could make and could not be calculated because of complexities in the sample designs, the challenge inherent in comparing data across modes, and the inexact relationship between post-stratification weighting and standard error adjustments (Gelman, 2007).

be larger using the ANES-style weights (21.1 percentage points) than when using the base weights only (20.5 percentage points).

With ANES-style weights, the telephone samples were more representative than using any other weighting scheme with either data stream. The absolute error of 11.4 percentage points in the telephone data (Table 1, row 12, column 1) was significantly smaller (p<.001) than the absolute error of 12.7 percentage points for the Internet data with base weights only (Table 1, row 7, column 2). When comparing the telephone data with ANES-style weights to the Internet data with base weights only, absolute errors were significant for 94.9 percent of the variable-weeks in the telephone data and 89.7 percent of the variable-weeks in the Internet data (Table 1, row 12, columns 3 and row 7, column 4), not significantly different. The largest absolute error in the telephone data (18.7 percentage points; see Figure 3 row 2) was also smaller than the largest absolute error in the Internet data (20.5 percentage points; see Figure 3, row 1).

In conclusion, with the most defensible analytic approach (post-stratification and matching on dates), the telephone samples were more representative than the Internet samples.

Methods Used in Additional Analyses

For the sake of parsimony, we present a limited set of results for additional analyses in the remainder of this document. Only analyses conducted using ANES-style weights and matching on interview dates are discussed below because that approach yielded the most demographic representativeness. Where doing so did not add considerably to the length of this document, tables present results generated using all six weighting schemes.

Distributions of Opinions and Behaviors

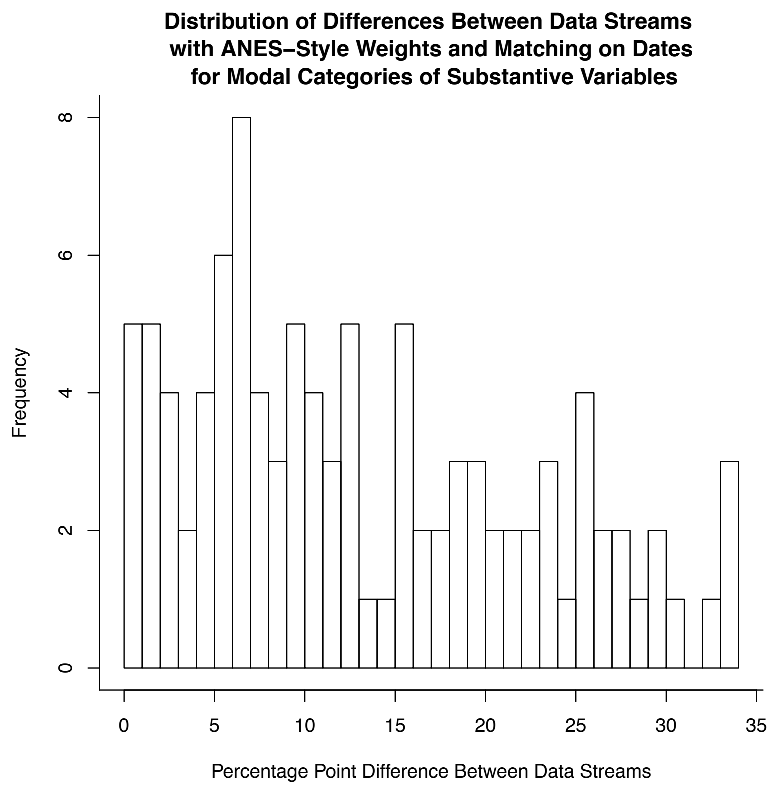
Distributions of opinions and behaviors often differed significantly between the telephone and Internet data streams. Using the ANES-style weights and matching on dates, the proportion of people giving modal responses differed between the two data streams by an

Table 2:: Differences Between the Telephone and Internet Surveys' Estimates for Opinions and Behaviors

Computation Method	Mean Discrepancy	Percent Significantly Different
With Base Weights Only		
Not Matched on Survey Dates	11.53	74.26
Matched on Survey Dates	11.73	70.30
With House Weights for the Telephone Data		
Not Matched on Survey Dates	12.25	77.23
Matched on Survey Dates	12.33	73.27
With ANES-Style Weights		
Not Matched on Survey Dates	13.48	84.16
Matched on Survey Dates	13.30	81.19

Notes: Column 1 displays the average absolute percentage point deviation between the proportion of respondents selecting the modal response category in one survey and the proportion selecting that category in the other survey. Column 2 displays the average percent of weeks (averaged across measures) when the proportions estimated for each variable by the two data streams were significantly different ($p<.05$, two-tailed). The variables included and their modal response categories were: R Plans To Participate – Definitely Will, R Received Census Form, R Completed Census Form, Count Importance – Agree, R's Participation Does Not Matter – Disagree, Trust Confidentiality – Agree, Don't Have Time to Fill Out – Disagree, The Census Can Help R, The Census is Used to Locate People Who Are in the U.S. Illegally. All variables measured for 13 weeks except R Received Census Form and R Completed Census Form.

Figure 4.



average of 13.3 percentage points (p<.001; Table 2, row 6, column 1).[12] These differences were statistically significant for 81.2 percent of the comparisons (p<.001; Table 2, row 1, column 2). The sizes of the differences are shown in Figure 4. For only a handful of the opinion and behavior measures did the data streams differ by less than a percentage point, while the largest differences were over 30 percentage points. The two data streams, therefore, often yielded very different portraits of the distributions of opinions and behaviors in the population.

Agree-Disagree Items

The proportions of respondents in the modal category of the agree-disagree rating scales differed significantly between the telephone and Internet data for all of the questions

[12]Modal response categories were identified using the largest total categories across both data streams.

Table 3:: Differences Between Data Streams in Responses to the Agree-Disagree Question

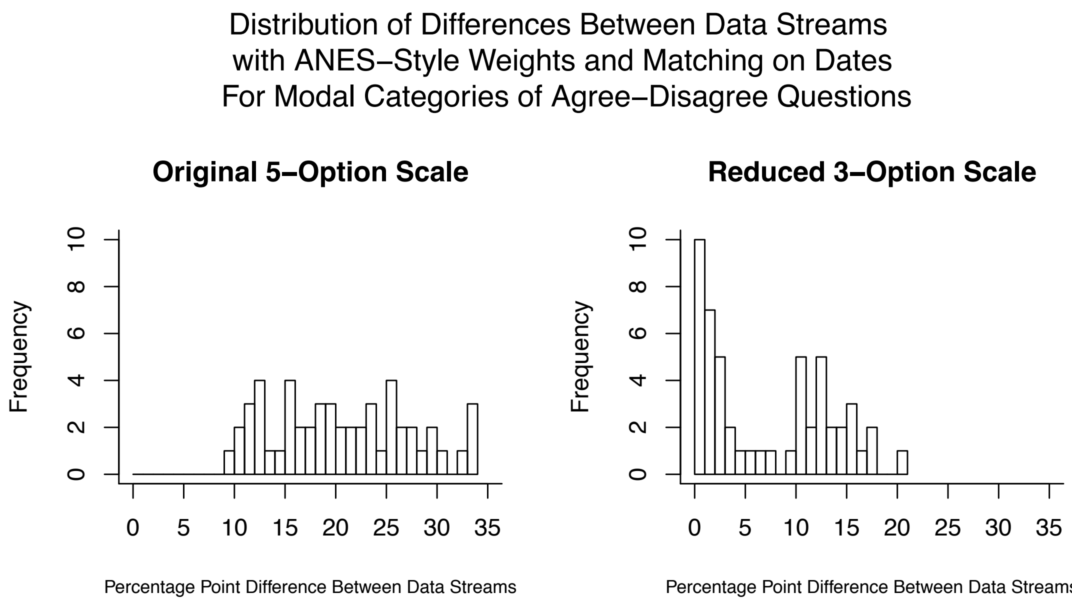
Computation Method	5-Point Scale Mean Discrepancy		3-Point Scale Percent Significantly Different	
With Base Weights Only				
Not Matched on Survey Dates	19.28	100.00	7.67	71.15
Matched on Survey Dates	19.33	100.00	7.76	67.31
With House Weights for the Telephone Data				
Not Matched on Survey Dates	19.76	100.00	7.46	63.46
Matched on Survey Dates	19.90	100.00	7.63	63.46
With ANES-Style Weights				
Not Matched on Survey Dates	20.79	100.00	7.56	55.77
Matched on Survey Dates	20.64	100.00	7.30	53.85

Notes: The 3-Point Scale combines "agree" and "strongly agree" responses and combines "disagree" and "strongly disagree" responses. Columns 1 and 3 display the average absolute percentage point differences between the proportion of people selecting the modal category in the telephone data and the proportion of people selecting that category in the Internet data. Columns 2 and 4 show the average proportions of weeks when the proportions for each variable in the two data streams were significantly different (p<.05, two-tailed). Variables and modal categories used were: Count Importance – Agree, R's Participation Does Not Matter – Disagree, Trust Confidentiality – Agree, Don't Have Time to Fill Out – Disagree. Modal categories used to generate columns 3 and 4 include: Count Importance – Agree or Strongly Agree, R's Participation Does Not Matter – Disagree or Strongly Disagree, Trust Confidentiality – Agree or Strongly Agree, Don't Have Time to Fill Out – Disagree or Strongly Disagree.

(see Table 3, column 2). The mean discrepancies were very large, that is between 19 and 21 percentage points (See Table 3, column 1). The largest difference between the two data streams was 33.5 percentage points.

These discrepancies occurred partly because the telephone survey respondents were less likely to report strong opinions (strongly agree or strongly disagree) than were the Internet survey respondents. Strong opinions were 21.2 percent of the responses offered by the telephone respondents, whereas the same figure was 34.5 percent for the Internet respondents, which was a significant difference (p<.001).

Figure 5.



After eliminating these differences in opinion strength by recoding the agree-disagree responses into three categories (agree, disagree, neither), the differences between the two data streams in terms of the percent of respondents in the modal category was significant for fewer of the questions (ranging from 54 percent to 71 percent, see Table 3, column 4). The mean discrepancies between the data streams in terms of the percent of respondents in the modal category was smaller, ranging from 7 to 8 percentage points. The largest difference between the two data streams was 20.3 percentage points (see Figure 5). Nonetheless, the distributions of responses in the two data streams were often notably different.[13]

[13]Descriptions of results of analyses using the agree-disagree items that follow sometimes include descriptions of results generated using both the 5-point version and the 3-point version. Sometimes, we only describe results with the 3-point version in order to minimize the presentation. The 3-point version always yielded results that were more similar across data streams, so in this sense, it is the more conservative approach.

Table 4:: How Different Coding of "don't know" Responses Affects the Apparent Magnitude of Differences Between Results Obtained with the Telephone and Internet Data Streams

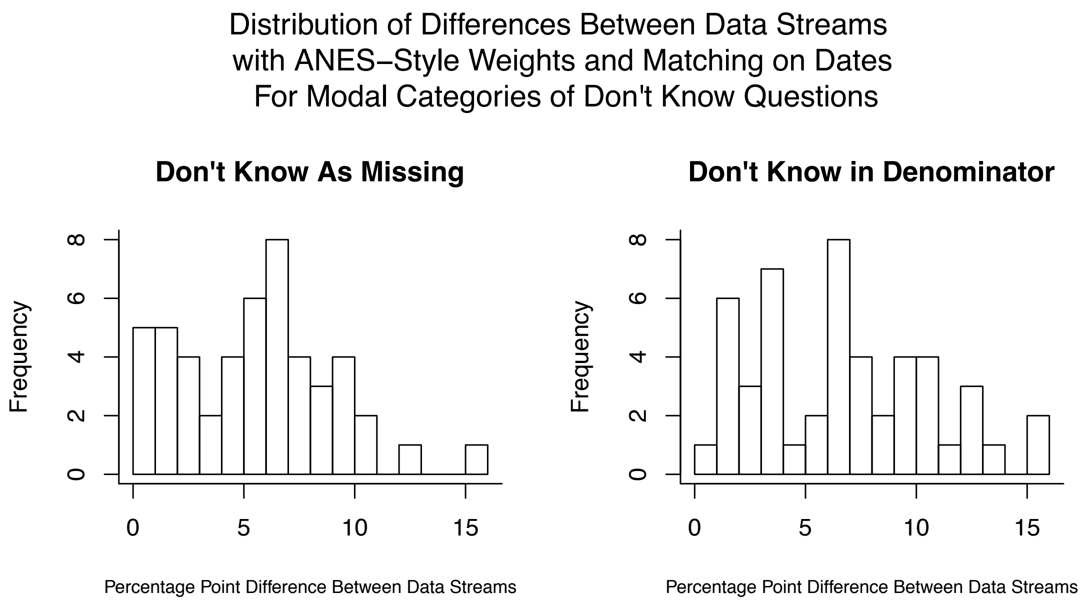
Computation Method	"don't know" Responses Dropped		"don't know" Responses in the Denominator	
	Mean Discrepancy		Percent Significantly Different	
With Base Weights Only				
Not Matched on Survey Dates	1.95	42.11	2.00	32.65
Matched on Survey Dates	2.09	33.33	2.11	28.57
With House Weights for the Telephone Data				
Not Matched on Survey Dates	2.22	45.61	1.96	34.69
Matched on Survey Dates	2.21	26.32	2.19	22.45
With ANES-Style Weights				
Not Matched on Survey Dates	3.24	49.12	2.66	46.94
Matched on Survey Dates	3.01	36.84	2.60	28.57

Notes: Columns 1 and 3 display the average absolute percentage point differences between the proportion of people selecting the modal category in the telephone data and the proportion of people selecting that category in the Internet data. Columns 2 and 4 show the average proportions of weeks when the proportions for each variable in the two data streams were significantly different ($p<.05$, two-tailed). Variables and modal categories used were: R Definitely Will Complete the Census form, R Received Census Form, R Completed Census Form, The Census Can Help R, The Census is Used to Locate People Who Are in the U.S. Illegally. All variables measured for 13 weeks except R Received Census Form and R Completed Census Form.

Methods of Handling "Don't Know" Responses

The telephone survey respondents were never offered an explicit "don't know" option, whereas the Internet survey respondents were offered such an option with many of the opinion and behavior questions. We therefore expected that the telephone respondents would be considerably more likely to answer these questions substantively than were the Internet respondents. This difference in presentation formats could be responsible for the differences across data streams in the pattern of substantive answers to the opinion and behavior questions.

Figure 6.



To explore this possibility, we compared the substantive results obtained using two different strategies for handling the "don't know" responses. In one approach, "don't know" responses were considered missing data and were dropped from the analysis. In a second approach, "don't know" responses were treated as substantive and were included in the denominator when calculating the proportion of the sample who reported each opinion or behavior.

As expected, when a question was asked identically in the telephone and Internet data streams except for the presence of a "don't know" option in the Internet version, the telephone respondents were more likely than Internet respondents to provide a substantive answer. An average of 3.9 percent of telephone respondents declined to answer these questions substantively, whereas this figure was 9.5 percent among Internet respondents, which was a significantly larger number (p<.001).

Apparent differences between the telephone and Internet data streams were the same regardless of how "don't know" responses were handled. When offering or omitting the "don't know" option was the only distinction between the data streams in terms of how a question was asked, the average difference between the proportion of respondents in the modal response categories in the telephone and Internet data was 3.0 percentage points when "don't know" responses were treated as missing and was significantly different from zero, $p<.001$; see Table 4, row 6, column 1) and 2.6 percentage points when those responses were instead included in the denominator, which was again significantly different from zero, $p<.001$; see Table 4, row 6, column 3). These two numbers were not significantly different from one another.

Whereas the data streams yielded a significantly different result for 36.8 percent of the variable-weeks when "don't know" responses were dropped ($p<.001$; Table 4, row 6, column 2), that figure was 28.6 percent when "don't know" responses were included in the denominators ($p=.001$; Table 4, row 6, column 4), not significantly different. The largest difference between the data streams was 15.6 percentage points when "don't know" responses were dropped and an equivalent 15.6 percentage points when "don't know" responses were included in the denominator (see Figure 6).

Predictors of Census Form Completion Intention and Completion

Next, we compared the data streams in terms of the degrees to which the substantive and demographic variables predicted the respondents' intention to complete the Census form and completion of the Census form.

Predictors of Intent to Complete the Census Form

Parameters of ordinal logit regression equations were estimated to predict respondents' intention to complete the Census form. For all regressions, data from the telephone and Internet surveys were combined, and a dummy variable predictor indicated whether each respondent came from the telephone (coded 0) or the Internet (coded 1) data stream. By having this dummy variable interact with each predictor, we could assess whether the relation between the predictor and the respondents' intent to complete the Census form was different between the telephone and Internet data streams.

A series of regressions examined each predictor separately (see Table 5, columns 1, 2, and 3), and all predictors were then combined into a single large regression equation as well (see Table 5, columns 3, 4, and 5). Single-predictor regressions in Table 5 show the coefficients for each level of each predictor (Telephone; Table 5, column 1), the coefficients for each level of each predictor when the coding of the data source variable is reversed (Internet; Table 5, column 2), and the coefficients for the interactions between the predictors and the Internet dummy variable (Difference; Table 5, column 3).

Substantive predictors in the single-predictor regressions. In the telephone data, with only one exception, all of the substantive variables significantly predicted the respondents' intent to complete the Census form in the expected direction in the single-predictor regressions. Respondents who thought that the Census could help them were more likely to say they will complete the form than did respondents who did not think so, and respondents who thought that the Census could harm them were less likely to say that they would complete the form than were respondents who did not think so (bs=-.97 and -1.08 respectively, ps<.001; Table 5, column 1, rows 1-2). Respondents who thought that the Census would be used to locate illegal immigrants were considerably less likely to intend to complete the form than respondents who did not think so (b=-.58, p<.001; Table 5, column 1, row 3).

Telephone respondents were more likely to intend to complete the form if they said the following: that the confidentiality promise could be trusted, that they had time to fill it out, that it was important to count everyone, and that that their participation did matter (bs=.63, 1.12, 1.90, and 1.48, ps<.001 respectively; Table 5, column 1, rows 4, 7, 8, and 11). Telephone respondents were less likely to intend to complete the form if they thought that the confidentiality promise could not be trusted, r did not have time to fill it out, and did not think that counting everyone was important (bs=-.44, -.47, and -.68, ps<.001 respectively; Table 5, column 1, rows 5, 6, and 9). The only exception to the expected results was that respondents who said that their participation did not matter were not significantly less likely to intend to complete the form than were respondents who neither agreed nor disagreed

Table 5:: Ordinal Logistic Regressions Predicting Intent to Complete the Census Form with ANES-Style Weights and Matching on Survey Dates

	Single Predictor Regressions			Multiple Predictor Regression		
Variable	Telephone	Internet	Difference	Telephone	Internet	Difference
The Census Can Help R	.97***	1.44***	.48***	.61***	1.19***	.57***
The Census Can Harm R	-1.08***	-1.56***	-.49**	-.61***	-.98***	-.37
The Census is Used to Locate People Who Are in the U.S. Illegally	-.58***	-.72***	-.14	-.19**	-.32***	-.13
Can Trust Confidentiality Promise - Agree or Strongly Agree	.63***	1.18***	.55***	.21*	.26*	.05
Can Trust Confidentiality Promise - Disagree or Strongly Disagree	-.44***	-.17**	.27**	-.31**	-.25*	.07
Don't Have Time to Fill Out - Agree or Strongly Agree	-.47***	.03	.50***	-.31***	-.05	.26
Don't Have Time to Fill Out - Disagree or Strongly Disagree	1.12***	1.70***	.58***	.70***	.61***	-.09
Importance of Counting Everyone - Agree or Strongly Agree	1.90***	2.10***	.20	.92***	.71***	-.21
Importance of Counting Everyone - Disagree or Strongly Disagree	-.68***	.87***	1.56***	-.62***	.14	.76**
R's Participation Does Not Matter - Agree or Strongly Agree	-.10	-.15*	-.05	-.38**	-.36**	.02
R's Participation Does Not Matter - Disagree or Strongly Disagree	1.48***	2.00***	.52***	.67***	.95***	.28
Female	.26***	.10*	-.15*	.10	.06	-.04
White	.54***	.46***	-.07	.28**	.18	-.10
Black	.48***	.43***	-.05	.39**	.34	-.04
Hispanic	.49***	.37***	-.12	.41**	.56**	.15
Age 18-24	-1.42***	-1.88***	-.46***	-1.76***	-1.81***	-.05
Age 25-44	-.19**	-1.35***	-1.16***	-.76***	-1.29***	-.54**
Age 45-64	.21**	-.58***	-.79***	-.21*	-.62***	-.42*
Region - Northeast	.06	-.04	-.10	-.06	.27*	.33*
Region - North Central	.11	.09	-.02	.10	.11	.01
Region - South	.08	-.09	-.17*	.12	-.09	-.21
Education - Less Than High School	-1.39***	-1.01***	.39**	-1.16***	-.47*	.69**
Education - High School Degree	-1.09***	-.59***	.50***	-.97***	-.38*	.59**
Education - Some College	-.86***	-.56***	.31*	-.67***	-.37*	.31
Education - College Graduate	-.45***	-.46***	-.01	-.44***	-.26	.18
Married	.55***	.57***	.02	.18**	.08	-.10
English Speaking Household	.52**	.70***	.18	-.08	.00	.08
Spanish Speaking Household	.66***	.62**	-.05	.97***	-.38	-1.34***

Table 5:: Ordinal Logistic Regressions Predicting Intent to Complete the Census Form with ANES-Style Weights and Matching on Survey Dates (cont.)

	Single Predictor Regressions			Multiple Predictor Regression		
Variable	Telephone	Internet	Difference	Telephone	Internet	Difference
R Owns Residence	.72***	.94***	.23	.36*	.22	-.14
R Rents Residence	.23	.33***	.10	.11	-.01	-.12
One Person in Household	.40***	.36**	-.04	.17	–	–
Two Persons in Household	.62***	.37***	-.25	.13	-.15	-.28
Three Persons in Household	.36***	-.02	-.38**	.02	-.27	-.29
Four Persons in Household	.27**	-.09	-.35*	.03	-.18	-.21
Five Persons in Household	.19	.04	-.15	-.07	.03	.10
Any Children in Household	-.16***	-.19***	-.03	.02	.16	.14
Week 1	.07	.11	.04	-.38*	-.29	.10
Week 2	.27	.61***	.34	-.07	-.09	-.01
Week 3	.26	.39*	.12	-.08	-.22	-.13
Week 4	.25	.32	.07	-.08	-.21	-.13
Week 5	.04	.59***	.55*	-.39*	-.02	.38
Week 6	.31	.48**	.17	.02	.06	.04
Week 7	.38*	.52**	.14	-.11	.09	.21
Week 8	.67***	.81***	.14	.38	.43	.06
Week 9	.89***	1.03***	.15	.73***	.87*	.15
Week 10	.64***	.29	-.34	.43	.15	-.27
Week 11	.21	.19	-.01	.20	.30	.11
Week 12	-.03	-.08	-.05	-.11	-.24	-.13
Source	–	–	–	–	–	.72
N					11124	
McFadden's Pseudo R^2					.47	
AIC					17915.47	

Note: Columns 1, 2, 4, and 5 display regression coefficients, and columns 3 and 6 display differences between the data streams as estimated by an equation including an interaction term. All variables representing responses to each survey question were included in each single predictor regression. The omitted reference categories for categorical variables were: Male, Other Race, Age 65 and older, Region - West, Education - Graduate, Not Married, Other Household Language, Respondent Has Some Other Living Arrangement, Six or More Persons in Household, and No Children in Household. * $p<.05$ | ** $p<.01$ | *** $p<.001$.

with the statement about completion mattering (b=-.10, p=.25; Table 5, column 1, row 10).

Nine of the 10 substantive variables that had predicted intent to complete the Census form in the telephone data showed similar associations that were significant in the Internet data. As in the telephone data, Internet respondents were more likely to report intent to complete the forms if they thought the Census could help them, and respondents were less likely to report that they would complete the form if they thought it could harm them (bs=1.44 and -1.56, ps<.001; Table 5, column 2, rows 1-2). Belief that the Census would be used to locate illegal residents was associated with a diminished likelihood of intent to complete the form (b=-.72, p<.001; Table 5, column 2, row 3).

Internet respondents were also more likely to report intent to complete their forms if they trusted the confidentiality promise, said they had time to fill it out, said that counting everyone was important, or said that their responses did matter (bs=1.18, 1.70, 2.10, and 2.00, ps<.001; Table 5, column 2, rows 4, 7, 8, and 11). Respondents who did not trust the confidentiality promise or thought that their participation was unimportant were less likely to intend to complete the form (bs=-.17 and -.15, ps=.009 and .04; Table 5, column 2, rows 5 and 10). Internet respondents who said that they did not have time to fill out the Census form were no more or less likely to say they would complete the form than were people who offered no opinion about having time to complete it (b=.03, p=.68; Table 5, column 2, row 6). Surprisingly, respondents who said that it was not important to count everyone were more likely to report that they would complete the form than did respondents with no opinion on the matter (b=.87, p<.001; Table 5, column 2, row 9).

Whereas the existence and directions of relations in the single-predictor regressions were similar between the telephone and Internet data, the strength of associations were significantly different between the two data streams for eight of the eleven predictors. For five of these variables, relations in the telephone data were significantly weaker than those in the Internet data (Census Can Help R, Census Can Harm R, Can Trust Confidentiality Promise – Agree or Strongly Agree, Don't Have Time to Fill Out – Disagree or Strongly Disagree, and R's Participation Does Not Matter – Disagree or Strongly Disagree; Table

5, column 3, rows 1, 2, 4, 7, and 11). For three variables, associations were significantly stronger in the telephone data than in the Internet data (Can Trust Confidentiality Promise – Disagree or Strongly Disagree, Don't Have Time to Fill Out – Agree or Strongly Agree, and Importance of Counting Everyone – Disagree or Strongly Disagree; Table 5, column 3, rows 5, 6, and 9). Only three predictors of intent to complete the Census form manifested similar strengths of relations in the two data streams. Thus, for the most part, the relations of predictors with intentions were not identical between the data streams.

Substantive predictors in the multiple-predictor regressions. When including all predictors at once in an equation, all eleven substantive variables manifested significant associations with intention in the expected directions in the telephone data. People who thought that the Census could help them were more likely to intend to complete the form than were individuals who thought that the Census would neither help nor harm them; both sets of individuals, in turn, reported being more likely to complete the form than did individuals who thought that the Census would harm them (*b*s=.61 and -.61, *p*s<.001; Table 5, column 4, rows 1-2). Respondents reported that they were less likely to complete the Census form if they thought that it was used to find illegal residents (b=-.19, p=.001; Table 5, column 4, row 3). Respondents were also more likely to say they would fill out the Census form the more they agreed that they could trust the confidentiality of the Census, the more they disagreed that they lacked the time to fill out the form, the more important they thought it was to count everyone, and the less they agreed with the notion that their participation was irrelevant (Table 5, column 4, rows 4-11).

Nine of the eleven variables that significantly predicted intent to complete the Census form in the telephone data in the combined regression also predicted significantly and in the expected direction in the Internet data. The more respondents believed that the Census could help them, agreed that they could trust the confidentiality promise, disagreed with the notion that they did not have time to fill out the Census form, thought that counting everyone was important, and rejected the sentiment that their participation did not matter, the more likely they were to say they would complete the form (Table 5, column 5, rows 1, 4, 7, 8, and 11). Respondents were less likely to report that they would complete the form to the extent that they thought that the Census could harm them, was used to identify illegal immigrants, or had a confidentiality promise that could not be trusted (Table 5, column 5, rows 2, 3, 5). Respondents who thought that their participation did not matter were also less likely to say they would complete their forms than did respondents who did not hold those sentiments (b=-.36, p=.009; Table 5, column 5, row 10). All else being equal, Internet respondents who reported that they did not have time to fill out the Census form or that counting everyone was unimportant were not significantly more or less likely to report that they would complete the Census form (bs=-.05, and .14, ps = .70 and .41 respectively; Table 5, column 5, rows 6 and 9).

When all predictors were included in the regression, only two of the coefficients for substantive variables differed significantly between the telephone and Internet data streams. Perception that the Census could help respondents accounted for less variance in intentions among the telephone respondents than among the Internet respondents (b=.57, p<.001; Table 5, column 6, row 1). Compared to the telephone respondents, the Internet respondents were less dissuaded from sending in their forms by notions that counting everyone was unimportant (b=.76, p=.001; Table 5, column 6, row 9). Thus, of eleven substantive variables, eight manifested the same direction and strength of relations with intentions in the telephone data as they had in the Internet data.

Demographic predictors in single-predictor regressions. In the telephone data, nine of the ten demographic variables significantly predicted respondents' intent to complete the Census form in the single-predictor regressions. Sex, race, age, education, marital status, language spoken in the household, home ownership, number of persons in the household, and the presence of children in the household were all related to self-reported likelihood of response (Table 5, column 1, rows 12-36). Only region was not related to reported intentions (Table 5, column 1, rows 19-21).

The same nine demographic variables that predicted telephone respondents' intentions were significant predictors in the Internet data. In the single-predictor models, intent to complete the Census form was related to sex, race, age, education, marital status, primary household language, home ownership, number of individuals living in the household, and the presence of children in the household (Table 5, column 2, rows 12-36).

The specific response categories that predicted respondents' intent to complete the Census form differed slightly between the two data streams, however. In contrast with the telephone respondents, the Internet respondents who reported that they rented their homes were significantly more likely to report their intention to complete the form than were individuals who neither owned nor rented (b=.33, p<.001; Table 5, column 2, row 30). In the telephone data, respondents living in households with three or four individuals were more likely to report intending to complete the form than were respondents in six person households; individuals in three and four person households were no more likely to report that they would do so in the Internet data (Table 5, column 2, rows 33-34).

Whereas the two data streams showed similar relations between demographic information and intent to complete the Census form in the single-predictor regressions, the respective strengths of the relations differed. Among the nine demographic variables that significantly predicted intention in the each data stream, five coefficients were significantly different between the two data streams: sex, age, region, education, and number of persons in the household (Table 5, column 3, rows 12-36). Sometimes, a coefficient was significantly stronger in the telephone data than in the Internet data. Other times, the association was stronger in the Internet data than in the telephone data. Differences between data streams were sufficiently common that a researcher's choice of a data stream to use would affect the conclusions he or she would reach.

Demographic predictors in the multiple-predictor regressions. In the combined regression, only six of ten demographic variables were related to telephone respondents' reported likelihood of completing the form (race, age, education, marital status, language spoken in the household, and home ownership; Table 5, column 4, rows 12-36). Sex, region, and the number of individuals in the household were not significantly related to respondents' intent to complete the Census form.

Four demographic variables were found to be significantly related to respondents' self-reported intent to complete the form in the Internet data stream; three of these variables mirrored predictors in the telephone data: Race, age, region, and education (Table 5, column 5, rows 12-36). Sex, marital status, language spoken in the household, home ownership, and the number of persons living in the household did not significantly predict respondents' projected likelihood of completing the Census form (Table 5, column 5, rows 12-36).

In the multiple-predictor regressions, four of the seven variables that predicted intent to complete the form differed between the data streams. Census region, education, and primary language in the household produced significantly different coefficients in the two data streams (Table 5, column 6, rows 12-36). Of 13 response categories for demographic variables that significantly predicted intent to complete the form in the telephone data, only three response categories in the Internet data had statistically significant coefficients, in the same direction, and did not significantly differ from the telephone results (Hispanic, Age 18-24, and Education – Some College). Again, the two data streams supported different conclusions about demographic correlates.

Predictors of Census Form Completion

The parameters of binomial logistic regressions were estimated predicting respondents' reports of whether they had completed the Census form (which we refer to as "completion behavior"). We again estimated the parameters of two sets of regressions to determine whether the data streams led to similar inferences (see Table 6). For all regressions, data from the telephone and Internet surveys were combined, and a source dummy variable was used to indicate whether any given respondent came from the telephone (coded 0) or the Internet (coded 1) data stream. A series of regressions examined each predictor separately (Table 6, columns 1, 2, and 3), and all predictors were then included in a single large equation (Table 6, columns 3, 4, and 5). We examined the effect of each predictor in the telephone data (Table 6, column 1) and in the Internet data (Table 6, column 2). The differences between these coefficients were tested via interactions between the predictors and

Table 6:: Logistic Regressions Predicting Census Form Completion with ANES-Style Weights and Matching on Survey Dates

	Single Predictor Regressions			Multiple Predictor Regression		
Variable	Telephone	Internet	Difference	Telephone	Internet	Difference
The Census Can Help R	.13	.28**	.15	.01	.22	.21
The Census Can Harm R	-.42	-.64*	-.22	-.22	-.41	-.19
The Census is Used to Locate People Who Are in the U.S. Illegally	-.04	-.30*	-.25	.13	.12	-.01
Can Trust Confidentiality Promise - Agree or Strongly Agree	.30	.29**	.00	.05	.05	.01
Can Trust Confidentiality Promise - Disagree or Strongly Disagree	.06	-.09	-.15	-.16	.05	.22
Don't Have Time to Fill Out - Agree or Strongly Agree	.34	.16	-.17	-.03	.56	.60
Don't Have Time to Fill Out - Disagree or Strongly Disagree	1.17***	1.12***	-.05	.88**	.68	-.19
Importance of Counting Everyone - Agree or Strongly Agree	.89***	.57***	-.32	.90*	-.37	-1.27*
Importance of Counting Everyone - Disagree or Strongly Disagree	-.12	.13	.25	-.11	-.36	-.25
R's Participation Does Not Matter - Agree or Strongly Agree	.18	-.15	-.33	-.03	-.35	-.33
R's Participation Does Not Matter - Disagree or Strongly Disagree	.63**	.71***	.08	.14	.22	.07
Female	.04	-.05	-.08	.01	.04	.03
White	.40*	.37*	-.04	.12	.49	.37
Black	.18	.12	-.05	.14	.31	.17
Hispanic	-.12	.02	.14	.21	.25	.04
Age 18-24	-1.76***	-1.19***	.57*	-1.74***	-.96*	.79
Age 25-44	-1.36***	-.66***	.70**	-1.51***	-.68*	.83*
Age 45-64	-.63***	-.42**	.20	-.77***	-.53	.24
Region - Northeast	-.05	-.01	.04	-.15	-.05	.10
Region - North Central	.18	.16	-.03	.16	.43	.27
Region - South	-.13	-.08	.05	-.38*	.08	.46
Education - Less Than High School	-.09	-.59**	-.50	-.04	-.33	-.29
Education - High School Degree	.09	.15	.06	.02	-.09	-.10
Education - Some College	-.03	-.14	-.11	.09	-.13	-.22
Education - College Graduate	.08	-.04	-.12	.11	-.22	-.33
Married	.21**	.39***	.18	.31*	.58**	.27

Table 6:: Logistic Regressions Predicting Census Form Completion with ANES-Style Weights and Matching on Survey Dates (cont.)

	Single Predictor Regressions			Multiple Predictor Regression		
Variable	Telephone	Internet	Difference	Telephone	Internet	Difference
English Speaking Household	.11	.14	.03	.13	-.62	-.76
Spanish Speaking Household	-.58	-.83*	-.25	-.13	-1.18	-1.04
R Owns Residence	-.47	.14	.61	-.55	-1.13*	-.57
R Rents Residence	-.96*	-.29	.68	-.81	-1.24**	-.43
One Person in Household	.82***	.56*	-.26	.82*	–	–
Two Persons in Household	.92***	.74**	-.19	.80**	.65	-.14
Three Persons in Household	.61**	.56*	-.04	.62*	.52	-.10
Four Persons in Household	.11	.31	.20	.36	.39	.03
Five Persons in Household	.15	.17	.02	.36	.44	.08
Any Children in Household	-.58***	-.36**	.22	.19	-.12	-.31
Week 9	1.28**	1.10*	-.18	1.14*	.77	-.37
Week 10	2.39***	2.55***	.16	2.49***	2.56***	.07
Week 11	2.88***	3.31***	.43	3.02***	3.48***	.46
Week 12	3.12***	3.76***	.64*	3.29***	4.03***	.74
Week 13	3.69***	–	–	3.97***	–	–
Intercept	–	–	–	-2.81*	-.98	1.83
N					5062	
McFadden's Pseudo R^2					.46	
AIC					4343.68	

Note: Columns 1, 2, 4, and 5 display regression coefficients, and columns 3 and 6 display differences between the data streams as estimated by an equation including an interaction term. All variables representing responses to each survey question were included in each single predictor regression. The omitted reference categories for categorical variables were: Male, Other Race, Age 65 and older, Region - West, Education - Graduate, Not Married, Other Household Language, Respondent Has Some Other Living Arrangement, Six or More Persons in Household, and No Children in Household. * $p<.05$ | ** $p<.01$ | *** $p<.001$.

the Internet dummy variable (Table 6, column 3).

Substantive predictors in the single-predictor regressions. Only three substantive variables predicted Census form completion in the single-predictor regressions in the telephone data, each in the expected direction. There was no significant difference in form completion depending on whether respondents thought the Census could help or harm them (*b*s=.13 and -.42, *p*s=.19 and .10 respectively; Table 6, column 1, rows 1-2), or whether they believed that the Census was used to locate illegal residents (*b*=-.04, *p*=.68; Table 6, column 1, row 3). Compared to individuals holding neutral viewpoints, telephone respondents were more likely to complete the form if they disagreed that it took too much time to fill out, if they agreed that it was important to count everyone, and if they disagreed with the notion that their participation did not matter (*b*s=.1.17, .89, and .63 respectively, *p*s<.002; Table 6, column 1, rows 7, 8, and 11). Form completion was not significantly related to respondents' trust in the confidentiality of the Census (*b*s=.30 and .06, *p*s=.06 and .74; Table 6, column 1, rows 4-5). Similarly, relative to individuals with neutral opinions, respondents were no more or less likely to complete the form if they reported that they did not have time to fill it out, if they disagreed that it was important to count everyone, or if they agreed with the notion that their participation did not matter (*b*s=.34, -.12, and .18, *p*s=.13, .71 and .39 respectively; Table 6, column 1, rows 6, 9, and 10).

Seven of the 11 substantive variables significantly predicted form completion in the Internet data, including all three that had predicted it in the telephone data. Respondents were more likely to have completed the form if they thought the Census could help them than if they did not think so, and were even less likely to say they had completed the form if they thought it could harm them (*b*s=.28 and -.64, *p*s=.008 and .04 respectively; Table 6, column 2, rows 1-2). Respondents were also less likely to report that they completed the form if they thought that it was used to locate illegal residents (*b*=-.30, *p*=.03; Table 6, column 2, row 3).

Compared to respondents who either said that they didn't know or had no opinion, individuals were more likely to have completed the form if they trusted the confidentiality promise, thought they had time to fill out the form, regarded counting everyone as important, and perceived their own participation as relevant (*b*s=.29, 1.12, .57, and .71 respectively, *p*s<.01; Table 6, column 2, rows 4, 7, 8, and 11). There were no significant differences in completion rates between respondents with no opinion and those who thought that the confidentiality of the Census could not be trusted, that they did not have time to fill it out, that counting everyone was unimportant, or that their participation did not matter (*b*s=-.09, .16, .13, and -.15, *p*s=.53, .46, .54, and .40 respectively; Table 6, column 2, rows 5, 6, 9, and 10).

There were no significant differences between the telephone and Internet data streams in terms of which substantive predictors of form completion appeared to be consequential in single-predictor regressions. Although fewer predictors were significant in the telephone data than in the Internet data, the differences between the two data streams never reached statistical significance (Table 6, column 3, rows 1-11). Therefore, the two data streams yielded equivalent portraits of the predictors of form completion when each predictor was examined individually.

Substantive predictors in the multiple predictor regression. In the regressions with all predictors at once, only two of the substantive predictors were related to form completion. Both of these relations were in the expected direction. As in the single predictor regressions, respondents who disagreed that they lacked time to fill out the Census form or who agreed that it was important to count everyone were more likely to complete their forms than individuals who held other views (*b*s=.88 and .90, *p*s=.001 and .02; Table 6, column 4, rows 7 and 8).

In contrast, none of the substantive variables in the Internet data stream significantly predicted form completion in the equation with all predictors (Table 6, column 5, rows 1-11).

For one of the two substantive variables that predicted form completion in the telephone data, its association with form completion in the Internet data was significantly different. Respondents who agreed that counting everyone was important were more likely than individuals who did not hold that viewpoint to report that they had completed their forms in the telephone data, but this relation did not appear in the Internet data (*b*s=.90 and -.37, *p*s=.02 and .33; Table 6, columns 4-5, row 8). This difference was significant (*b*=-1.27, *p*=.02; Table 6, column 6, row 8). The only other substantive predictor to reach statistical significance in the telephone regression was not a significant predictor in the Internet data (Don't Have Time to Fill Out – Disagree or Strongly Disagree). There were no variables for which the telephone and Internet data streams indicated similar significant relations.

Demographic predictors in single predictor regressions. In the telephone data, six of the ten demographic variables significantly predicted form completion in the single predictor regressions: race, age, marital status, home ownership, number of persons in the household, and the presence of children in the household (Table 6, column 1, rows 12-36). Sex, region, education, and primary household language were unrelated to form completion (Table 6, column 1, rows 12-36).

Seven demographics predicted form completion in single predictor regressions in the Internet data, including five demographics that had been significant predictors with the telephone data. Among the predictors with significant relations in the telephone data, race, age, marital status, number of persons in the household, and the presence of children in the household were significantly related to form completion in the Internet data (Table 6, column 2, rows 12-36). Whereas renting a home had been inversely related to Census form completion in the telephone data (*b*=-.96, *p*=.04; Table 6, column 1, row 30), there was no significant relation in the Internet data (*b*=-.27, *p*=.24; Table 6, column 2, row 30). Education and primary household language were also related to Census form completion among Internet respondents (Table 6, column 2, rows 22-25 and 27-28).

Demographic predictors of form completion were similar in the two data streams in single predictor regressions. Only age predicted significantly differently across the data streams. Whereas younger respondents tended to be less likely to return their forms according to both data streams, these differences were significantly more pronounced in the telephone data than in the Internet data (Table 6, column 3, rows 16-18). Relations with form completion were not significantly different across the data streams for any other predictors

(Table 6, column 3, rows 12-36). Coefficients for seven demographic response categories were statistically significant, in the same direction, and not significantly different from one another in the two data streams (White, Age 45-64, Married, One person in household, two persons in household, three persons in household, and any children in household; Table 6, column 3, rows 12-36). Coefficients for three additional variables were not significantly different from one another, but only were statistically significant in one of the two data streams (Education, primary language in the household, and home ownership; Table 6, column 3, rows 12-36). Thus, in these comparisons, most predictors were related to form completion equivalently.

Demographic predictors in the multiple predictor regression. Four demographics were related to form completion in the multiple predictor regression with the telephone data. Age, region, marital status, and the number of persons in the household significantly predicted whether respondents completed their forms (Table 6, column 4, rows 12-36). Sex, race, education, language spoken in the household, home ownership, and presence of children in the household were not related to form completion (Table 6, column 4, rows 12-36).

Only three demographics predicted form completion in the Internet data, two of which had predicted similarly in the telephone data. Age, marital status, and home ownership were related to Census form completion among Internet respondents (Table 6, column 5, rows 12-36). Sex, race, Census region, education, primary language in the household, number of persons in the household, and the presence of children in the household were unrelated to form completion (Table 6, column 5, rows 12-36).

The relations between form completion and our predictors was different for only one demographic variable in the multiple predictor regression. Whereas individuals between the ages of 25 and 44 were less likely to report form completion than were older individuals in both data streams, membership in this age group was related more strongly to form completion in the telephone data than in the Internet data (b=.83, p=.02; Table 6, column 6, row 17). No other predictors manifested significant differences between the data streams (Table 6, column 6, row 12-36).

Changes Over Time

Variables That Manifested Change Over Time

Before assessing whether the telephone and Internet data streams documented similar changes over time in opinions and behaviors, we examined whether significant variation over time occurred for each variable within each data stream. In each data stream, χ^2 statistics were calculated to identify instances in which the proportion of respondents providing an answer varied significantly across the 13 weeks examined here. Significance markers in columns 1 and 2 of Tables 7 and 8 are based on the p values from these analyses. To provide a sense of the magnitude of these changes, we calculated the average absolute difference between the proportion of the sample selecting each response in any given week and the proportion of all samples that selected that response across all weeks. These average absolute differences are shown in columns 1 and 2 of Tables 7 and 8.[14]

Changes over time in the telephone data stream. For most variables in the telephone data, the proportion of respondents selecting each response category changed significantly over the 13 weeks (see column 1 in Table 7). The rate of selection of 17 of the 26 response categories varied significantly from week to week. Significant changes over time were observed for all response categories for four variables (R plans to complete the Census form, R received Census form, R completed Census form, and Don't have time to fill out), for some response categories in another three variables (The Census can help/harm R, Can trust confidentiality promise, and R's participation does not matter), and for no response categories in only two variables (The Census is used to locate people who are in the U.S. illegally, Importance of counting everyone). The extent of over-time change in each variable in the telephone data can be seen in the black lines in Figure 7 for response categories that manifested significant variation over time in at least one data stream and in Figure 8 for response categories that did not manifest significant over-time variation in either data stream.[15]

[14] Larger absolute differences across response options do not perfectly translate into more significant results, because χ^2 tests use the mean squared error instead of the mean absolute error.

[15] Neither, No Opinion, and Don't Know responses are shown in Appendix A.

Changes over time in the Internet data stream. Significant variation over time was also apparent for many measures in the Internet data (see column 2 in Table 7). 14 of the 17 response categories with significant over-time changes in the telephone data also varied significantly between weekly groups of Internet respondents. An additional four variables that did not significantly vary over time in the telephone data varied significantly over time in the Internet data. In the Internet data, significant over-time change appeared for all response options of the same four questions as in the telephone data (R plans to complete the Census form, R received Census form, R completed Census form, and Don't have time to fill out) and for some response options for an additional four variables (The Census can help/harm R, Can trust confidentiality promise, Importance of counting everyone, and R's participation does not matter). No over-time change was identified for only a single variable (The Census is used to locate people who are in the U.S. illegally). Change over time in each variable in the telephone data is shown by the gray lines in Figure 7 for response categories that varied significantly over time in at least one data stream and in Figure 8 for response categories that did not manifest significant variation in either data stream.

Comparing Changes Over Time

Two metrics were used to assess the similarity of the changes over time portrayed by the two data streams. χ^2 statistics were computed to identify significant differences between change over time in the two data streams.[16] Asterisks in column 3 of Table 7 indicate the results of significance tests assessing whether relations between weeks and selection of a response option differed between the two data streams.

To illustrate the magnitude of these differences, the numbers in column 3 of Table 7 are the average absolute difference in percentage points between variations in the two data

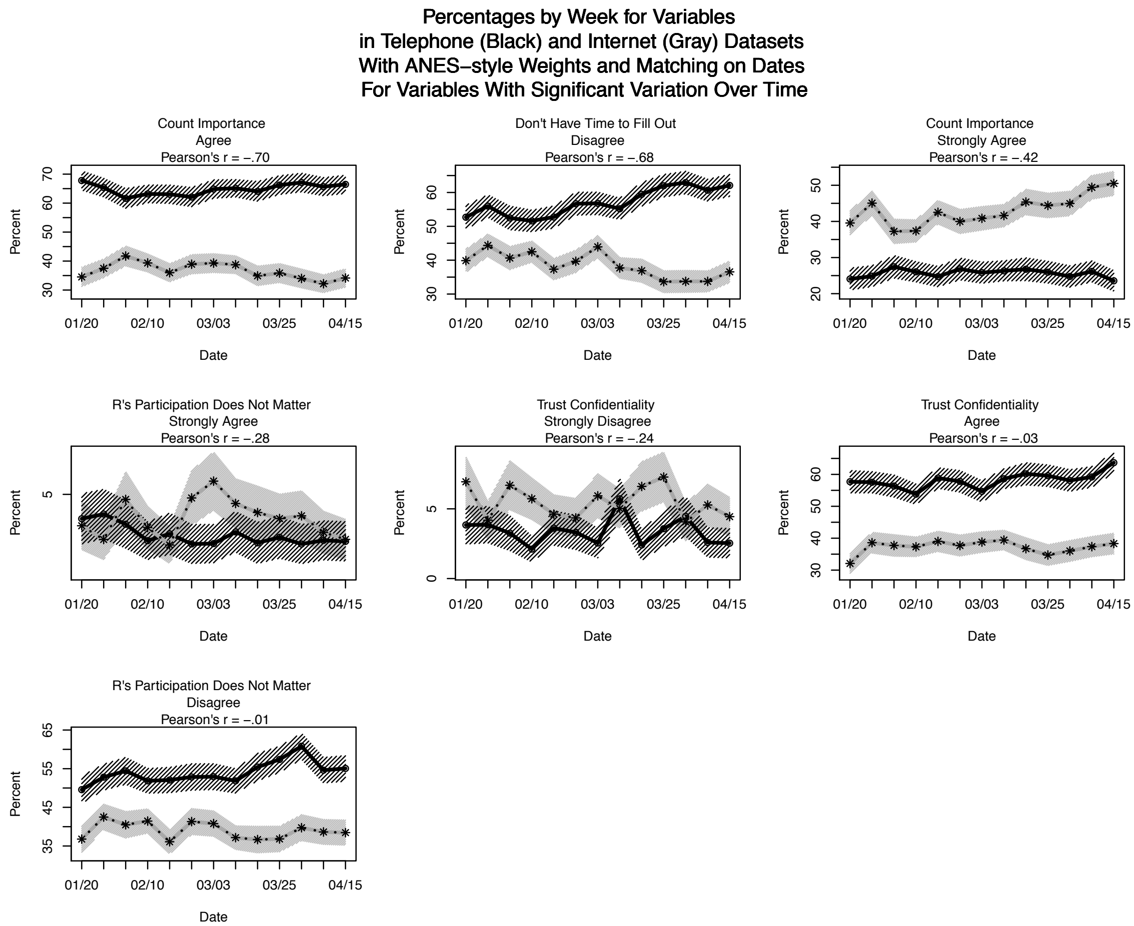
[16]This χ^2 statistic was computed in a three-step process. First, the χ^2 statistics in both the telephone and Internet data streams were added together to identify the total differences between observed and expected values within the two data streams. Second, we computed the χ^2 statistic comparing observed and expected values for each week with both data streams combined. The χ^2 statistic of interest was equal to the difference between the sum of the two separate data streams and the χ^2 for the combined data. Each of these statistics had degrees of freedom equal to the one less than the number of weeks during which the variable was measured.

Table 7:: Tests of the Significance of Change Over Time in the Telephone and Internet Data Streams and Tests of the Significance of the Difference Between The Data Streams in the Patterns of Over-time Change With ANES-Style Weights and Matching on Survey Dates

Variable	Variation Across Weeks – Telephone Data	Variation Across Weeks – Internet Data	Difference Between the Telephone and Internet Data
R Plans To Participate – Definitely Not	1.28***	.62*	1.29**
R Plans To Participate – Probably Not	1.21*	1.36***	1.56**
R Plans To Participate – Might or Might Not	1.36**	2.00***	2.16*
R Plans To Participate – Probably Will	2.06***	2.30***	1.54
R Plans To Participate – Definitely Will	4.66***	5.45***	3.05**
R Received Census Form	10.53***	2.46***	5.95
R Completed Census Form	22.39***	15.82***	9.13***
The Census Can Help R	2.57***	4.62***	2.69
The Census Can Harm R	.53	.45	.60
The Census is Used to Locate People Who Are in the U.S. Illegally	1.51	1.56	2.01
Trust Confidentiality – Strongly Disagree	.74**	.98*	1.45***
Trust Confidentiality – Disagree	1.10	.83	1.50
Trust Confidentiality – Agree	1.72*	1.43	2.70*
Trust Confidentiality – Strongly Agree	.70	2.36***	2.30*
Don't Have Time to Fill Out – Strongly Disagree	3.71***	11.03***	7.32***
Don't Have Time to Fill Out – Disagree	3.39***	3.07***	6.15***
Don't Have Time to Fill Out – Agree	3.50***	2.45***	1.73**
Don't Have Time to Fill Out – Strongly Agree	.55**	.66*	.77
Count Importance – Strongly Disagree	.19	.42	.33
Count Importance – Disagree	.77	.60	1.04*
Count Importance – Agree	1.55	2.36**	3.54***
Count Importance – Strongly Agree	.96	3.35***	3.70***
R's Participation Does Not Matter – Strongly Disagree	1.71**	3.32***	2.91**
R's Participation Does Not Matter – Disagree	2.13**	1.90	2.93**
R's Participation Does Not Matter – Agree	2.11***	.98	1.45
R's Participation Does Not Matter – Strongly Agree	.44	.82**	1.00*

Notes: Columns 1 and 2 display the average absolute percentage point differences between the global mean and each week's mean, averaged across weeks. Column 3 displays the average weekly difference between means in the two data streams after controlling for the global average difference. χ^2 tests of each outcome variable by weeks are used to test statistical significance in the two data streams. * $p<.05$ | ** $p<.01$ | *** $p<.001$.

Figure 7a

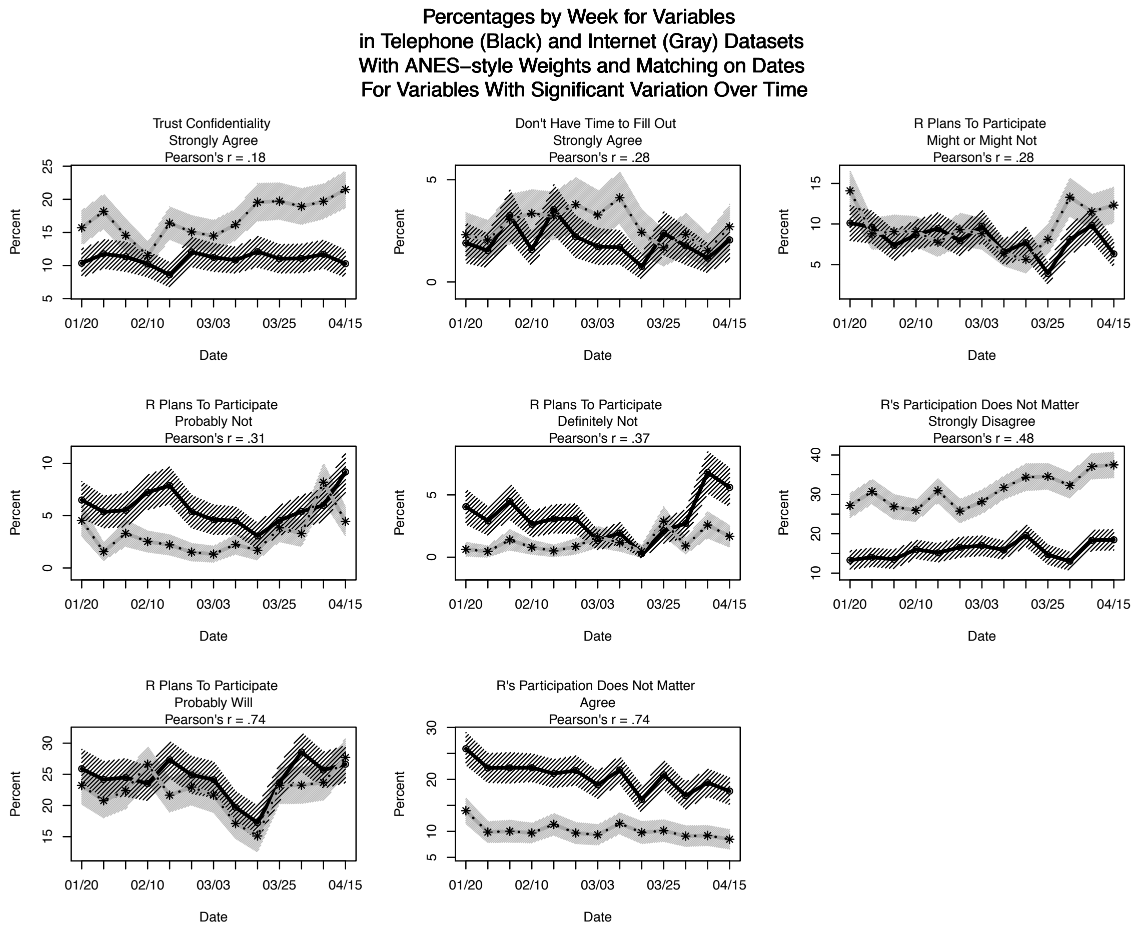


streams. These were produced in a three-step process. For step one, in each data stream, we subtracted the proportion of all samples combined who selected each response option from the proportion of each week's sample who selected that response option. Every response category therefore had 13 discrepancies in each data stream; these discrepancies indicated how much each week differed from the full time period's proportion. In step two, for each response category in each week, we found the absolute difference between the discrepancies in the two data streams. In the final step, these absolute differences were averaged across weeks to produce a single coefficient. To compute a second metric, we relaxed the assumption that variation in the two data streams would have the same magnitude. Pearson product moment correlations shown above each variable in Figure 7 indicate the extent to which the patterns of change in the two data streams were similar.

Differences in variation over time. For 16 of the 21 response categories where changes over time were significant in at least one data stream, changes were significantly different between the two data streams (Table 7, column 3). Differences between the two data streams were also observed for one additional variable where neither data stream yielded significant change over time (Count Importance – Disagree; Table 7, column 3, row 9). With ten of the response categories for which significant variation over time was observed, the differences between the data streams were as large or larger than the changes within either data stream. For example, this was the case for reports of whether the respondent would definitely not complete the Census form. Whereas the proportion choosing this option varied from its average by an average of 1.3 percentage points in the telephone data (p<.001; Table 7, column 1, row 1) and .6 percentage points in the Internet data (p=.03; Table 7, column 2, row 1) in any given week, the difference between these changes also averaged 1.3 percentage points (p=.004; Table 7, column 3, row 1). There were only four response options for which changes were observed in both the telephone and Internet data streams and the difference between the data streams in patterns of over-time change was not significant.

Patterns of change over time. Patterns of change over time in the telephone and Internet data streams were sometimes very different. Figure 7a shows instances where the two data streams told very different stories about change over time. The most striking instances of discrepancy involve negative correlations between the two over-time lines. For example, as shown in the top row of Figure 7a, in weeks when more people disagreed with the statement that they did not have time to fill out the Census form in the telephone data, fewer individuals chose that same response option in the Internet data. The correlations between these two lines is a striking -.68. An even stronger negative correlation appears in the upper left of Figure 7a – where over-time trends in judgments about the importance

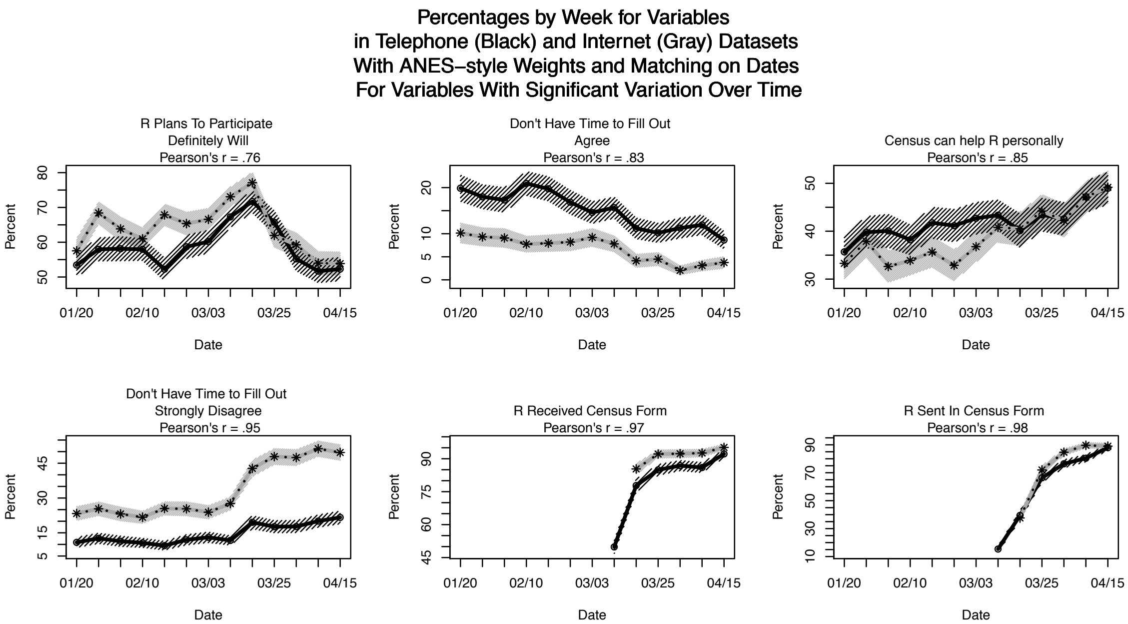
Figure 7b



of completing the form were correlated -.70. In the remainder of figure 7a are instances of weaker negative correlations or correlations of essentially zero, meaning that over-time trends in one line did not meaningfully covary with over-time trends in the other line.

Figure 7b shows instances in which there were stronger positive associations between the over-time trends from the two data streams, ranging from .18 in the upper left corner to .74 in the bottom right corner. Visual inspection of these figures shows that the more robust positive associations here are mostly due to similarly of a general monotonic trend in the lines rather than due to close correspondence of short-term fluctuations. For example, in

Figure 7c

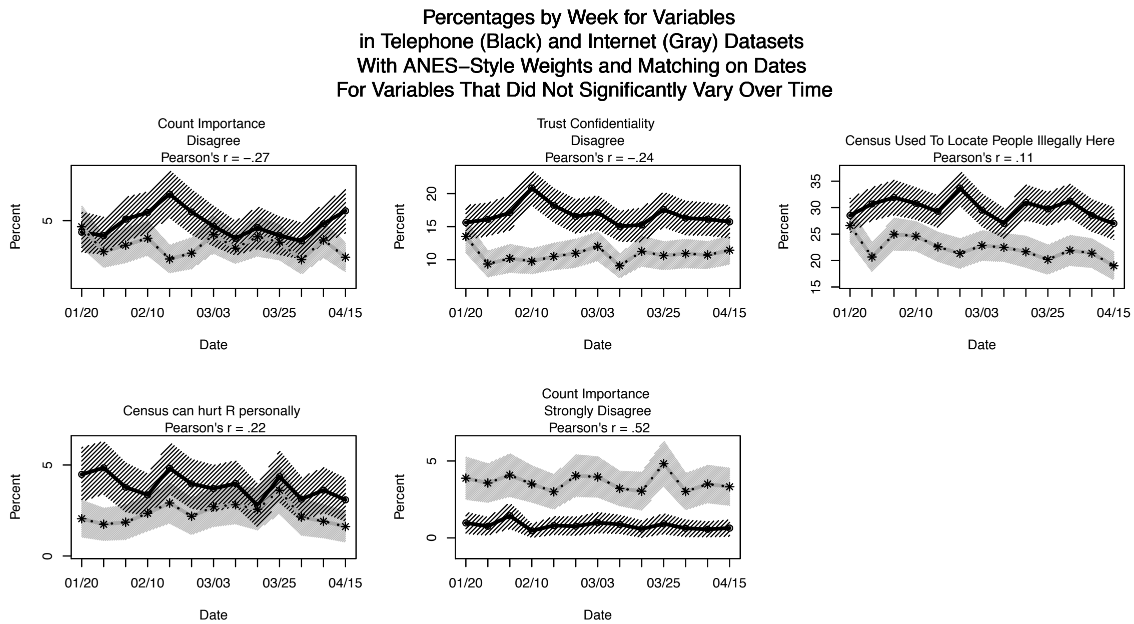


the lower right corner, we see that both data streams documented a downward trajectory in the proportion of people who said that their own Census participation does not matter, but the downward trend is much more pronounced in the telephone data than in the Internet data, and the ups and downs in the telephone data are minimal in the Internet data. Thus, the correlation of .74 between these lines might seem to overstate their correspondence.

Figure 7c shows stronger associations, and these appeared for relatively few of the measures. The correlations across data streams between the over-time trends range from .76 in the upper left corner to .98 in the lower right corner. Again, as is apparent in the graph showing the proportion of respondents who disagreed with the statement that they did not have time to complete the form (shown in the middle of Figure 7c), the over-time trend is much more dramatic in the Internet data than in the telephone data, despite the .95 correlation between the two lines.

Thus, though some over-time trends appeared to be fairly similar across the data

Figure 8.



streams, most were not.

Figure 8 shows relations between over-time trends for variables that did not manifest statistically significant over-time trends. One might imagine that in these instances, there is no reason to expect the over-time trends from the two data streams to resemble one another, because there was no reliable over-time change in either line. Three of the correlations between the data streams were negative, and the other two were positive. But in the instance of the strongest correlation (r=.52), visual inspection of the over-time trends suggests little meaningful correspondence, so this moderate correlation might be viewed as misleading in its magnitude.

It is interesting to note the results in Figure 7c shown for receiving the form and completing the form. The Internet respondents were more likely than the telephone respondents to report having received the form in all weekly surveys in which the question was asked. And in some weeks, the Internet respondents were more likely than the telephone respondents

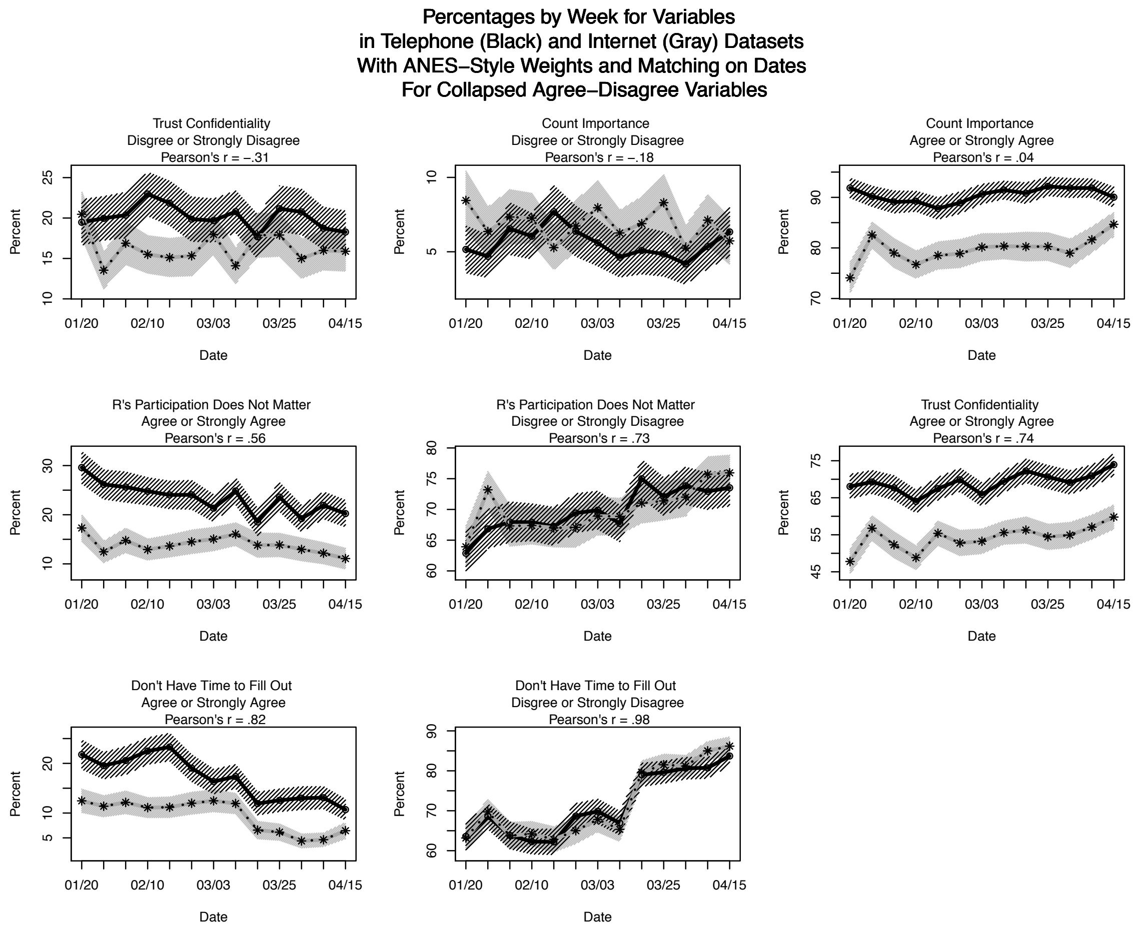
Table 8:: Changes Over Time in Telephone and Internet Data Streams and Interaction Between Changes With ANES-Style Weights and Matching on Survey Dates For Collapsed Agree-Disagree Questions

Variable	Variation Across Weeks – Telephone Data	Variation Across Weeks – Internet Data	Difference Between the Telephone and Internet Data
Trust Confidentiality – Agree or Strongly Agree	1.92**	2.51***	1.67
Trust Confidentiality – Disagree or Strongly Disagree	1.10	1.5*	2.24*
Don't Have Time to Fill Out – Agree or Strongly Agree	3.80***	2.93***	2.04**
Don't Have Time to Fill Out – Disagree or Strongly Disagree	7.10***	8.27***	1.53
Count Importance – Agree or Strongly Agree	1.15*	1.85***	1.90**
Count Importance – Disagree or Strongly Disagree	.79	0.84	1.10
R's Participation Does Not Matter – Agree or Strongly Agree	2.41***	1.27*	1.88
R's Participation Does Not Matter – Disagree or Strongly Disagree	2.85***	2.98***	1.97

Notes: Columns 1 and 2 display the average absolute percentage point differences between the global mean and each week's mean, averaged across weeks. Column 3 displays the average weekly difference between means in the two data streams after controlling for the global average difference. χ^2 tests of each outcome variable by weeks are used to test statistical significance in the two data streams. * $p<.05$ | ** $p<.01$ | *** $p<.001$.

to say that they had returned the form. In the remaining weeks, the form completion rates in the telephone and Internet data were about equal. These results should be interpreted with regards to past studies showing the Internet surveys are minimally distorted by social desirability pressures (e.g., Holbrook & Krosnick, 2010). If claiming to have received and returned the form are socially desirable answers, then we might imagine these answers would have been more prevalent in the telephone interviews than in the Internet interviews. The absence of this pattern suggests that the telephone reports were not in fact distorted by social desirability pressures.

Figure 9.



Changes Over Time for Collapsed Agree-Disagree Responses

We also compared the data streams in terms of the similarity of change over time using the 3-point agree-disagree scales. Variation across weeks of the telephone data, variation across weeks of the Internet data, and the differences between those patterns of variation are presented in Table 8. These numbers tell much the same story as we saw in Table 7. Figure 9 shows the patterns of over-time change and correlations between these patterns across data streams, which again tell a similar story.

Significant changes over time were present for six of the eight response categories in the telephone data (Trust Confidentiality – Agree or Strongly Agree, Don't Have Time to Fill Out – Agree or Strongly Agree, Don't Have Time to Fill Out – Disagree or Strongly Disagree, Count Importance – Agree or Strongly Agree, R's Participation Does Not Matter – Agree or Strongly Agree, R's Participation Does Not Matter – Disagree or Strongly Disagree; Table 8, column 1). In the Internet data, changes over time were statistically significant for all six of these categories as well as the proportion of respondents who disagreed or strongly disagreed that they trusted the confidentiality of the Census (Table 8, column 2). Only the proportion of individuals who disagreed that it was important to count everyone did not vary in either data stream (Table 8, row 2). For most of these collapsed measures, therefore, it was possible to compare changes over time.

Differences between data streams in terms of variation over time. For three of the seven measures that manifested significant change over time in at least one data stream, the patterns of over-time change differed significantly between the data streams (Table 8, column 3). In two of these three instances, differences between the two data streams were larger than the changes within either data stream (Trust Confidentiality – Disagree or Strongly Disagree, Count Importance – Agree or Strongly Agree; Table 8, rows 3 and 5). Significant variations were observed in both data streams for four response categories when the within-stream changes were not significant.

Patterns of change over time. Using the 3-point coding of the agree-disagree scales, changes over time in the two data streams were often inconsistent. Figure 9 tells much the same story as we saw in Figures 7a, 7, 7c, and 8: Some negative correlations appeared. One near zero correlation appeared. And some moderate to strong correlations appeared. Thus, sometimes, trends matched between the data streams, and sometimes they did not.

Relations Between Measures

To assess whether the two data streams led to similar conclusions about relations between variables, we identified pairs of variables that could be correlated in both data streams. Of 1,152 pairs of variables present in both datasets, 1,061 pairs were suitable (because the two variables were not responses to the same question and neither variable was from a filter question that determined whether another question would be asked). We produced weighted correlations comparing these dyads in each of the two data streams. The correlations varied from -.48 to .47 in the telephone data and from -.44 to .52 among Internet respondents. We then plotted the correlations in the two data streams for each pair of variables in Figure 10. We generated a summary statistic to illustrate the correspondence between these relations by regressing the vector of correlations from the Internet data on the vector of correlations from the telephone data.

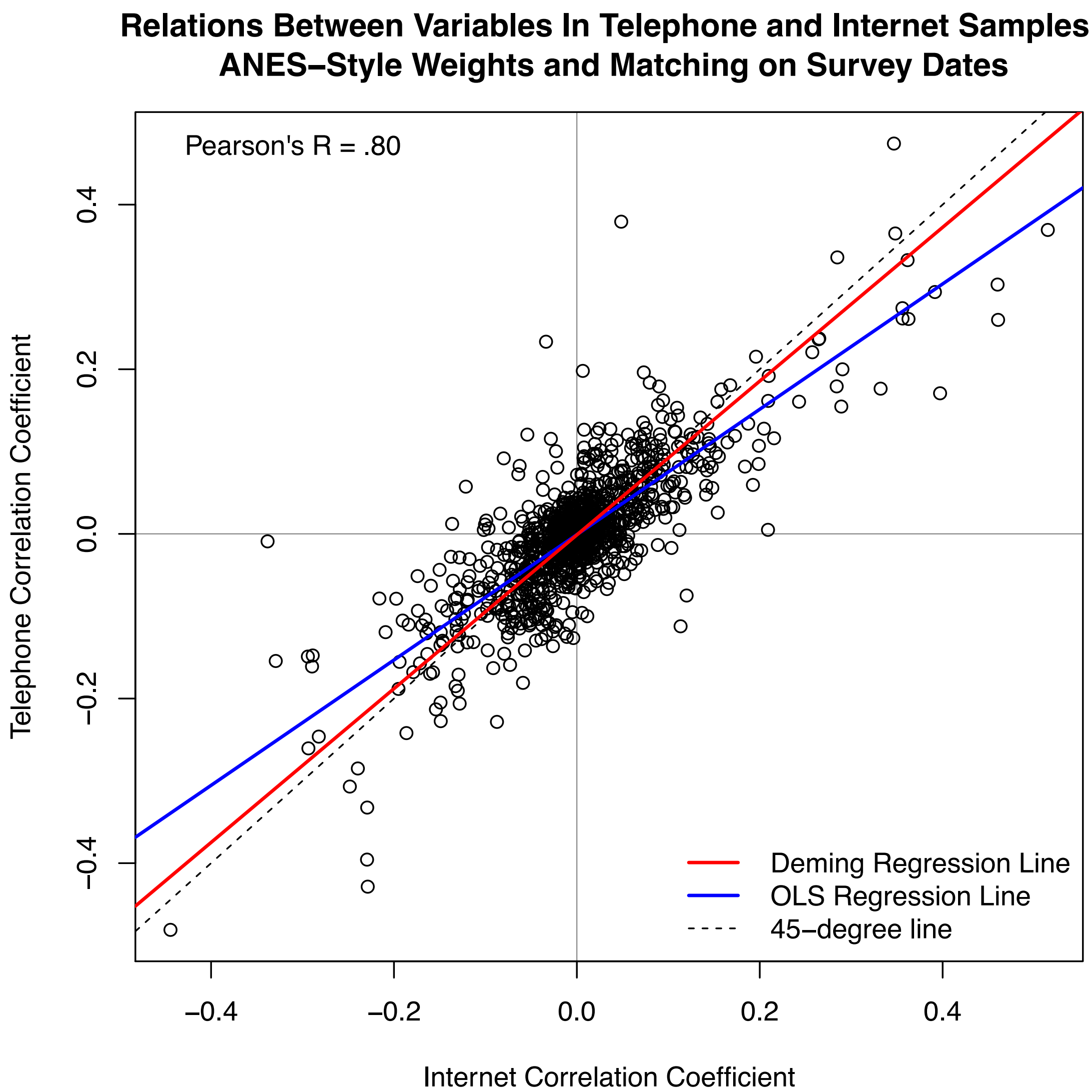
Relations between variables in the Internet data were stronger than those in the telephone data. On average, the correlations in the telephone data were .9 times as large as identical comparisons in the Internet data, a difference that was significantly different from 1:1 at the p=.02 level. After taking this difference in average strength into account, the relative strengths of the relations between variables were fairly consistent across data streams. Correlations in the telephone data stream each explained about 65 percent of the variation in correlations in the Internet data stream. The stronger but similar relations in the Internet data stream can be seen in the best-fitting regression lines in Figure 10.

Discussion

This investigation revealed systematic and often sizable differences between probability sample telephone data and non-probability sample Internet data in terms of demographic representativeness of the samples, the proportion of respondents reporting various opinions and behaviors, the predictors of intent to complete the Census form and actual completion of the form, changes over time in responses, and relations between variables. After post-stratification, the telephone samples were more representative than were the Internet samples in terms of demographics not used for quotas or weighting. These results therefore suggest that all of the probability telephone data's substantive results may have been more accurate than the Internet data's in describing the country as a whole.

For each type of inference examined, results from the two data streams frequently

Figure 10.


Relations Between Variables In Telephone and Internet Samples
ANES–Style Weights and Matching on Survey Dates
Pearson's R = .80
Deming Regression Line
OLS Regression Line
45–degree line
Telephone Correlation Coefficient
Internet Correlation Coefficient
0.4
0.2
0.0
–0.2
–0.4
–0.4
–0.2
0.0
0.2
0.4

differed. The proportion of the samples reporting various opinions and behaviors often differed by five, ten, or even 15 percentage points. The predictors of intent to complete the Census form differed significantly for all but three of the substantive variables examined in single-predictor regressions. Changes over time in the two data streams diverged more than would be predicted by chance in almost two-thirds of the cases examined. And correlations between variables in the two data streams were systematically different from one another.

The frequency and unpredictability of sizable differences between the two data streams indicate that they are not interchangeable. To conclude that the two data streams could be used equivalently for any given type of inference, we would need to see reliable, generalizable conclusions using that type of inference from the two data streams. The results of these analyses instead indicate that the choice between probability telephone and non-probability Internet data collection is consequential.

If the Census Bureau wishes to make a choice about whether to conduct probability sample telephone surveys or non-probability sample Internet surveys, the results presented here suggest that the former may yield more accurate results. Whereas the non-probability Internet samples were more representative without post-stratification, the method that yielded the most representative results was the application of ANES-style weights to the telephone data. This approach best approximated known population demographics and provided substantive results in line with what had been expected from most analyses.

Sample Representativeness

The Internet samples were more representative before post-stratification. This is likely to have occurred because E-Rewards constantly monitored the demographic profiles of respondents and adjusted sampling procedures to reach targets. The Internet data stream's considerably larger discrepancies from the benchmarks not used for quotas or weighting suggests that the representativeness of the Internet samples when no weights were applied might have occurred only for variables used in the constant adjustment procedure, which would not necessarily assure representativeness on other variables, demographic or substantive.

The closest matches to the benchmarks not used for quotas or weighting came from the post-stratified telephone survey. Remarkably, post-stratification weights actually reduced accuracy of the Internet surveys. This counterintuitive result implies that individuals within various demographic groups in the Internet surveys are substantially different from the individuals within those same groups in the general population.

Differences in Distributions of Opinions and Behaviors

Consistent differences were apparent between the distributions of opinions and behaviors in the two data streams. These findings too fall in line with prior comparisons between probability and non-probability samples (e.g. Malhotra and Krosnick, 2007). The proportions of the population selecting various response categories in the two surveys differed across all weighting and matching schemes in the current analyses. Understandings of the proportion of the population selecting each response option were different between data streams, often by a large amount.

Whereas there was some evidence that collapsing agree-disagree response categories could reduce the size of some differences between the data streams, none of our coding alternatives fully accounted for the discrepancies between the surveys. Collapsing these response categories also may have had negative consequences. To the extent that researchers are interested in the strength with which respondents agreed or disagreed with a premise, collapsing categories eliminated potentially meaningful response data. In addition, reducing the number of response categories diminishes the potential for disagreement; hence, some of the reduction in differences between data streams may have been an artifact of fewer opportunities to disagree. Nonetheless, even after collapsing categories, moderate sized differences in the valance of opinions indicated that the data streams continued to lead to different inferences.

Predictors of Census Form Completion

When predicting intent to complete the Census form or actual completion of the Census form, we found a moderate number of differences between the telephone and Internet data streams. Whereas some variables predicted identically in the two data streams, others were different. We did not see a consistent pattern regarding when the data streams were similar or different. This unreliability suggests that the data streams cannot be used interchangeably.

In the larger regressions, we saw fewer distinctions between the data streams. It is unclear whether this result implies that inferences are relatively similar, ceteris paribus, or whether the relatively large standard errors of the combined regression rendered us unable to detect substantive differences.

One surprising finding in our predictions of Census form completion also deserves note. Perceptions of the Census were closely related to intent to complete the form. In contrast, these same perceptions accounted for little variation in respondents' actual form completion. Thus, the relations between these variables may be weaker than had been presumed by researchers in the past. Hence, respondents may not have been particularly accurate forecasters of future Census form completion.

Changes Over Time

Changes over time in the telephone and Internet data streams often led to different conclusions. In most cases where reliable over-time variation was apparent in at least one of the two data streams, changes were not equivalent between the two data streams. Furthermore, in some cases, changes over time in the telephone and Internet data streams were inversely related. This echoes the results of the one other analysis comparing changes over time in probability and non-probability samples (Pasek, 2010b).

Relations Between Variables

In a final comparison, systematic differences were apparent between the relations among variables documented by the telephone and Internet data streams. Relations tended to be stronger in the Internet data than in the telephone data. This result, which has also been obtained in other studies comparing probability sample Interviewer-administered surveys with non-probability sample Internet surveys (Malhotra and Krosnick, 2007), suggests that respondents in the Internet surveys were systematically different from or approached their task differently than did the telephone respondents.

Differences Between Questions

Whereas the results of our analysis were relatively robust across coding and measurement schema, it is important to note that none of the questions examined in this analysis were truly identical. "Don't know" responses were explicitly presented in all Internet questions, whereas telephone survey respondents had to volunteer that they did not know the answer for a question. Furthermore, the agree-disagree questions did not offer an explicit middle response option in the Internet surveys, whereas they did in the telephone interviews. It is important to keep in mind that these and other differences between data streams in question wordings may account for some of the differences observed here.

Maximizing Participation in the Census and Canaries in the Mine

The surveys we analyzed were designed to track purported determinants of Census form completion, so that early warning signs could identify beliefs in the public that might be impediments to form completion. When such signs appear in surveys like these prior to or during the fielding of the Census, outreach efforts can be made to correct public misunderstandings.

The results reported here suggest that remarkably few of the purported determinants of form completion actually emerged as reliable predictors of form completion in the regressions shown in Table 6. Among the apparently inconsequential beliefs are: beliefs about whether the Census can help or harm the respondent, whether the Census is used to locate illegal residents, whether the confidentiality promise can be trust, and whether the respondent's participation matters. In fact, the only two predictors that emerged as consequential were having enough time to complete the form and the belief that counting everyone is important. This evidence suggests that perhaps only these two factors are indeed causes of form completion, and the remainder of the measured variables, some of which correlated with form completion, were not causally consequential. If that is true, then efforts to monitor these beliefs and to intervene to correct misunderstandings might be misplaced.

It is especially powerful to note that if the Bureau were to rely on the Internet data to reach conclusions about which of these purported causes of form completion to monitor, the conclusion would be: none. In column 5 of Table 6, not one of the substantive beliefs was a statistically significant predictor of form completion. Clearly, then, the choice of data stream would affect whether a researcher concluded that some of these beliefs may have been consequential or that none of them were.

Choosing Between Data Collection Methods

The results reported here suggest that the probability telephone and non-probability Internet surveys were not interchangeable. The data collection method used influenced the conclusions that researchers would reach about intent to complete the Census form, actual completion, their correlations, and their over-time trends. Therefore, choice of sampling and data collection mode should be made carefully in the future.

References

American Association for Public Opinion Research. (2009). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys (6th Edition). *Online Report*. Available from: http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions1 &Template=/ CM/ContentDisplay.cfm&ContentID=1814

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

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing vs. the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly*, 73(4), 641–678.

DeBell, M., & Krosnick, J. A. (2009). Weighting Plan for the American National Election Studies. ANES Technical Report Series. No. nes012427. Available from:

ftp://ftp.electionstudies.org/ftp/nes/bibliography/documents/nes012427.pdf.

DraftFCB (2010). 2010 Census Attitudinal Tracking Study (CATS) Frequently Asked Questions (FAQ). *Unpublished Internal Document.*

Fricker, S. S., Galesic, M., Torangeau, R., & Yan, T. (2005). An experimental comparison of web and telephone surveys. *Public Opinion Quarterly*, 69, 370–392.

The Gallup Organization (2010). G1K Methodology for Census. *Unpublished Internal Document.*

Gelman, A. (2007). Struggles with Survey Weighting and Regression Modeling. *Statistical Science*, 22(2), 153-164.

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an Econometric Evaluation Estimator. *The Review of Economic Studies*, 65(2), 261-294.

Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly*, 74, 37-67.

Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P. (2006). Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. *Public Opinion Quarterly*, 70, 759–779.

Malhotra, N., & Krosnick, J. A. (2007). The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Non-Probability Samples. *Political Analysis*, 15, 286-323.

Pasek, J. (2010a). Anesrake: ANES Raking Implementation. *CRAN*. R package version 0.2. Available from: http://cran.r-project.org/web/packages/anesrake/index.html.

Pasek, J. (2010b). Trends Over Time in Public Opinion: Do Probability Sample Telephone Surveys and Non-Probability Internet Surveys Produce the Same Results? *Unpublished Manuscript.*

Yeager, D., Krosnick, J. A., Chang, L., Javitz, H. S., Levindusky, M. S., Simpser, A., and Wang, R. (In Press). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Public Opinion Quarterly.* Available from: http://www.knowledgenetworks.com/insights/docs/Mode-04_2.pdf.

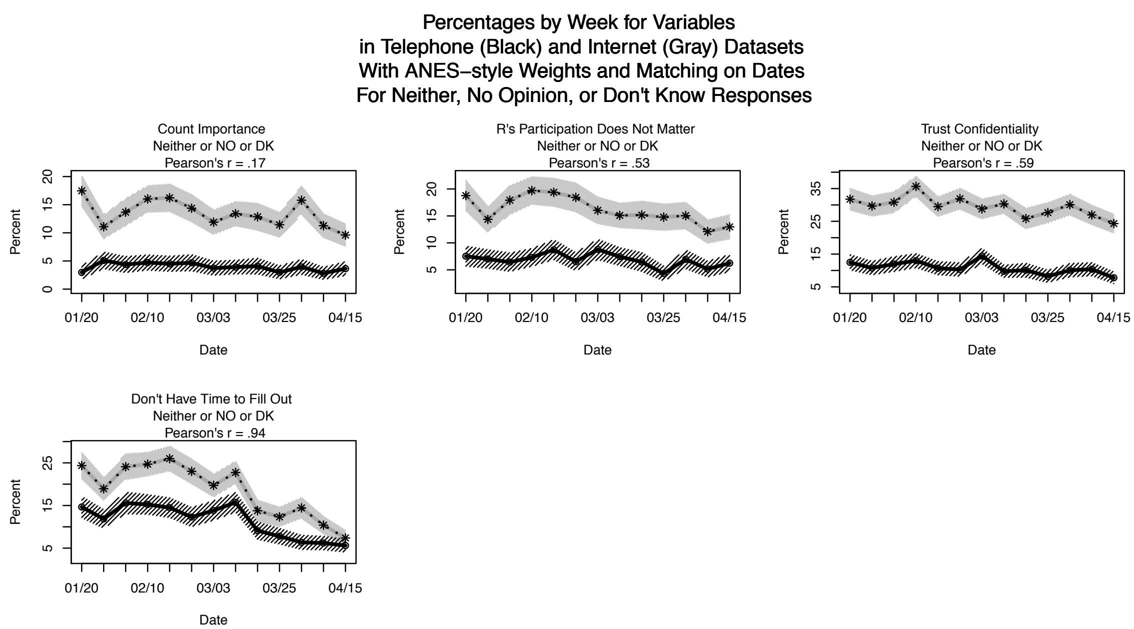
Appendix A

Table A1: Changes Over Time in Telephone and Internet Data Streams and Interaction Between Changes With ANES-Style Weights and Matching on Survey Dates For Neither, No Opinion, and Don't Know Responses

Variable	Variation Across Weeks – Telephone Data	Variation Across Weeks – Internet Data	Difference Between the Telephone and Internet Data
Trust Confidentiality – Neither or No Opinion or DK	1.35***	2.13***	1.79
Don't Have Time to Fill Out - Neither or No Opinion or DK	3.41***	5.34***	2.31
Count Importance - Neither or No Opinion or DK	.57	1.96***	1.77
R's Participation Does Not Matter - Neither or No Opinion or DK	.91*	2.08***	1.90

Notes: Columns 1 and 2 display the average absolute percentage point differences between the global mean and each week's mean, averaged across weeks. Column 3 displays the average weekly difference between means in the two data streams after controlling for the global average difference. χ^2 tests of each outcome variable by weeks are used to test statistical significance in the two data streams. * $p<.05$ | ** $p<.01$ | *** $p<.001$.

Figure A1


Percentages by Week for Variables
in Telephone (Black) and Internet (Gray) Datasets
With ANES–style Weights and Matching on Dates
For Neither, No Opinion, or Don't Know Responses
Count Importance
Neither or NO or DK
Pearson's r = .17
R's Participation Does Not Matter
Neither or NO or DK
Pearson's r = .53
Trust Confidentiality
Neither or NO or DK
Pearson's r = .59
Don't Have Time to Fill Out
Neither or NO or DK
Pearson's r = .94
Percent
Date
01/20
02/10
03/03
03/25
04/15