Does Survey Mode Still Matter?
Findings From a 2010 Multi-Mode Comparison

Stephen Ansolabehere
Professor of Government
Harvard University
sda@gov.harvard.edu

Brian F. Schaffner
Associate Professor of Political Science
University of Massachusetts, Amherst
schaffne@polsci.umass.edu

Abstract

In this paper, we present data from a three-mode survey comparison study carried out in 2010. National surveys were fielded at the same time over the Internet (using an opt-in Internet panel), by telephone with live interviews (using a national RDD sample of landlines and cell phones), and by mail (using a national sample of residential addresses). Each survey utilized a nearly identical questionnaire soliciting information across a range of political and social indicators, many of which can be validated with government data. Comparing the findings from the modes to each other and the validated benchmarks, we demonstrate that a carefully executed opt-in Internet panel produces estimates that are as accurate as a telephone survey and that the two modes differ little in their estimates of other political indicators and their correlates.

Surveys are a staple of social science research. They are the main methods for collecting indispensible data on public opinion, political participation and voting behavior, public health, economic activity and labor markets, and a wide range of other social behaviors and attitudes. Survey methods, however, evolve constantly. Advances in communication technology continually alter the most effective way to reach people, requiring researchers to decide which approaches to sample selection and survey administration will yield data appropriate to answer important questions. In the 1970s, the debate among survey researchers was over the acceptability of random digit dial phone surveys, compared with the much more expensive face-to-face interviews of randomly selected households and mail surveys. In the 1990s and 2000s the debate was over the acceptability of computer-administered interviews. Today the challenge is how to conduct surveys in a world where the modes of communication have proliferated, where cell phones are as prevalent as land lines, where market research is common over the Internet, but where no one mode is likely to cover all people in the population equally well.

The communication revolution that has struck the United States over the past two decades has led to fragmentation in survey research methods. No one approach seems to capture the American population sufficiently. Nearly one-fourth of the adult population cannot be reached by pollsters over landlines (Blumberg and Luke 2009), and less than one-third of those who can be reached actually agree to be polled (Curtin et al. 2005). The rapid increase of Internet penetration in American homes has made web-based polling a viable and affordable alternative for students of public opinion. This has been reflected, for example, in the publication of Internet surveys in the top journals of political science. For example, between 2006 and 2010, 33 articles published in either the *American Political Science Review*, *American Journal of Political Science*, or *Journal of Politics* utilized Internet survey data produced by YouGov/Polimetrix or

Knowledge Networks. The National Science Foundation now receives and supports a large number of proposals for survey research using these newer technologies, and several consortia of scholars have begun institutionalizing Internet surveys to study political behavior, including Time Sharing Experiments for the Social Sciences (TESS), the Cooperative Congressional Election Study (CCES) and the Cooperative Campaign Analysis Project (CCAP). The American National Election Study (ANES) incorporated web-based surveys into its recent studies, along side their traditional face-to-face survey and occasional phone surveys.

Among public opinion researchers there is a lack of consensus about the acceptability of the newer survey methods. It is difficult to evaluate different methods because we lack evidence that can guide decisions. In 2010, the American Association of Public Opinion Researchers (AAPOR) commissioned a task force to consider how the survey research industry should approach the use of opt-in Internet panels. The task force reviewed a large number of studies using different modes. It expressed caution about the use of opt-in Internet panels, but stressed that relatively little research compares the different modes directly. "Despite the widespread use of online panels," the AAPOR group concluded, "there is still a great deal that is not known with confidence" (2010, p. 54). Of all of the studies reviewed by AAPOR, only a small number produced a design in which the questions and methods were directly comparable across modes (e.g. Yeager et al. 2009; Chang and Krosnick 2010).[1] Also, nearly all of the studies cited by the AAPOR report use data generated between 2000 and 2005, before the rise of cell phones and when there was less Internet penetration. Researchers are rightly cautious about accepting new methods without first comparing them with accepted approaches. Apart from the handful of

[1] Even these studies have produced findings that are sufficiently ambiguous to allow both sides in the debate to cite it as evidence supporting their contentions (e.g. Rivers 2009).

research papers relied on by AAPOR, there is very little systematic information about the alternative modes available today that would help researchers address questions encountered in designing or evaluating surveys. Accordingly, in 2011, AAPOR announced a new task force on survey modes, and out-going AAPOR President Frank Newport of Gallup called for serious consideration of opt-in Internet surveys.

This paper presents results from a study conducted in 2010 comparing opt-in Internet, telephone, and mail survey modes. These surveys were designed so that identical questionnaires were used across modes, utilizing questions that are commonly asked on existing surveys (such as those fielded by Pew, the NHIS, and ANES) many of which can provide objective baselines against which to compare the estimates. Our analysis examines three key comparisons. First, we measure the extent to which each approach produces accurate point estimates for measures on which we have validated benchmark data, such as from the U.S. Census. Second, we examine cross-mode similarities and differences in point estimates for political measures (such as attitudes and reported behavior) that cannot be validated. Third, we compare the correlation structure and regression models estimated from data from each mode (e.g. Sanders et al. 2007). This is not a matter of "construct validity" (or which mode produces the strongest correlations), but a matter of whether the correlation structure replicates across modes and matches that using data from the Current Population Survey's turnout study.

Ultimately, we find that the three modes produce remarkably similar estimates. All modes showed some slight deviation from the objective indicators, but these deviations were, on average, not far off from sampling error. The basic correlation structures were similar across modes and not distinguishable from the objective correlation structure. To the extent that there was a noticeable difference among the modes on indicators, that difference arose with knowledge

of politics. Those in the opt-in Internet panels appear somewhat more knowledgeable than those in the other modes. Interestingly, the difference appears to stem from the fact that the sample frame is the Internet population, rather than the opt-in nature of the study, which suggests that growing Internet penetration will reduce mode differences further. Overall, the mode effects we found were small and mostly insignificant for common questions about public opinion, politics, and public health.

Design of the 2010 Mode Study

In early 2010, we commissioned YouGov/Polimetrix of Palo Alto, CA to administer the same questionnaire online to an opt-in Internet panel, by phone to a combined landline/cell-phone sample, and by mail to a sample of residential addresses. The questionnaire mostly focused on politics, but also included several lifestyle measures that could be validated using government data. The full questionnaire is available as an online appendix to this paper (LINK REMOVED). Table 1 provides summary information for these surveys, and we describe each survey in greater detail below.

The Telephone Survey

The telephone survey was conducted January 28 – 30th, with 807 interviews completed with respondents reached via landline numbers and a supplement of 100 interviews conducted using cell phone numbers. Live interviewers were used to administer the questionnaire to respondents. Each telephone number was attempted up to 6 times before it was dropped from the sample. The response rate for the landline portion of the sample (RR3) was 20.9%, while the rate was 8.6% for cell phone numbers. The combined response rate for the telephone survey was

19.5%. The median time for completion of a telephone interview was 14 minutes and 20 seconds.

The Opt-in Internet Survey

The Internet sample for our study came from the YouGov/Polimetrix online panel. The selection process for this panel includes recruiting a large number of people to serve on the survey panel through various methods, including online advertising. Individuals who join the panel earn rewards (i.e. points that can be redeemed for gift certificates and other items) for every survey they complete. Not all people are equally likely to respond to recruitment efforts so YouGov/Polimetrix uses targeted advertising to focus particular attention on recruiting groups that are underrepresented on their panel, such as racial and ethnic minorities.

Since YouGov/Polimetrix does not use probability sampling to recruit panelists, they instead rely on sample matching to generate representative samples from their panel. When YouGov/Polimetrix is commissioned to conduct a survey, they begin by taking a random sample from the target population. For example, if a client is asking for a survey of 1,000 American adults, YouGov/Polimetrix might draw a random sample from the Census Bureau's American Community Survey and use this as the target for constructing a sample from their own panel.

The databases from which the target sample is drawn provide basic demographic information for each member of the target population. Thus, once YouGov/Polimetrix draws the target sample, they know what each member of their random sample should look like on a range of characteristics and using these characteristics an algorithm selects the closest matching individuals from their Internet panel to essentially replace each person that was randomly selected into the target sample (Rivers N.D.). After matching everyone in the target sample with

at least one person from the Internet panel, YouGov/Polimetrix fields the survey to the selected panelists and then weights the responses to ensure that the matched sample is representative of the target sample.

For this particular study, panelists were matched to the target sample using age, race, education, interest in politics, gender, party identification, ideology, voter registration status, born again status, and region. Panelists were invited to take the survey beginning on January 15th and responses were accepted through February 11th. The panel produced 1,000 responses and the within-panel response rate (RR1) for this study was 42.9 percent. Note that because this is a within-panel response rate, it is not comparable to the response rates for the other two surveys. The median completion time for an opt-in Internet respondent was 8 minutes and 56 seconds.

The Mail Survey

The mail survey was generated by mailing questionnaires to 6,600 addresses selected randomly from a list provided by a data vendor. The sample was randomly divided into different types of incentive conditions—19% received no incentive, 39% were offered $1, 39% were offered $2, and 3% were offered $5. The overall response rate for this sample was (RR3) 21.1%.

Individuals receiving the mail questionnaire were offered the opportunity to either return their survey by post, or go online to take the survey. Of those responding to the mail solicitation, 27.5% went online to complete their questionnaires. Individuals choosing to complete the survey online tended to be younger, more educated, and male; they were also much more likely to have Internet access in their homes. Mail and Internet questionnaires are both self-administered, so we would not necessarily expect to find major differences across these two platforms. Indeed, we found few major differences between these two groups and, accordingly, we analyze all mail

respondents together in the analyses that follow.[2] Interestingly, individual's completing the questionnaire online took nearly 3 minutes longer than respondents from the YouGov/Polimetrix panel (but still shorter than the telephone interviews).

Dates of Interview Completion

Field dates vary considerably across different survey modes. The Internet and mail surveys generally had a longer field time, which could be a confounding explanation for any differences detected across modes. For the telephone and Internet surveys, we attempted to produce as much overlap as possible. Specifically, the Internet responses were collected over a period of approximately four weeks and the telephone poll was conducted over three days in the middle of that four-week span.

The mail survey was executed in two waves so that we could adjust the survey based on response rates. We did not know what the response rate would be because the mailing lists were of varying quality (e.g., some names did not have complete addresses). Doing the survey in two

[2] Using the measures presented in Table 2, we found that responses collected from mail respondents online departed by an average of 4.5 points while paper responses were off by 4.9 points on average. Two differences between web and paper respondents are particularly noteworthy. First, respondents submitting their questionnaires online were about 12 percentage points more likely to own their home. Second, paper respondents were more likely to report that the voted for Obama while web respondents reported more support for McCain (the difference was about 10 percentage points in both cases). For the most part, however, the responses provided by respondents choosing to go online to complete their mail survey did not differ substantially from those sent in by post.

waves allowed us to make sure that the response rates were not unusually low in some groups. The first wave of questionnaires was mailed at the end of January and 752 responses were received through the middle of April. The second wave started on June 22nd and extended through the end of September, during which an additional 455 responses were secured. The recruitment letter did not mention a university, but instead came from the survey firm. We suspect that the response rate might have been even higher had the initial approach come under university letterhead.

The extensive data collection period for the mail survey is, of course, an important reality for scholars and practitioners considering the mail mode. Respondents often take a considerable amount of time to submit their responses, making it difficult for researchers to restrict the time frame for study. We discuss this issue in greater detail later in the paper.

Sampling Weights

Each survey was weighted using propensity score weights on age, gender, education, race, voter registration status and ideology. The weighting for ideology was simply on the proportion of individuals answering "don't know" to the question about their ideology. In addition to these factors, the telephone survey was also weighted for the number of landlines in each respondent's household. The weights were trimmed at 7 for each survey.[3] We use sampling weights in all of the analyses that follow and refrain from making comparisons on any measures that were used for weighting (or matching in the case of Internet respondents).

Comparing Modes on Validated Measures

[3] Trimming the weights at 4 did not alter our findings in any meaningful way.

While comparisons across modes can be instructive, the most valuable metric for understanding the validity of a survey mode is to compare it to a validated baseline that can be treated as a good approximation to the population parameter. Thus, we begin by examining the extent to which each mode produces accurate estimates of characteristics for which we have validated benchmarks. Since the surveys were weighted on a set of demographic variables, we do not use those measures as benchmarks. In each of our surveys, we asked respondents whether they owned their home, when was the last time they had moved, whether they had smoked 100 cigarettes during their lifetime, whether they currently smoked, and whether they had health insurance.[4] Benchmark figures for these lifestyle questions were taken from the American Community Survey (ACS), the Current Population Survey (CPS), and the National Health Indicators Survey (NHIS) using data collected during the period most proximate to our own field dates.

We are also able to validate several political measures. Specifically, we asked respondents whether they had voted in the 2008 presidential election and, if so, which candidate they voted for. We also asked respondents to each survey how they voted in 2008—whether by mail, in-person before election day, or in-person on election day. We use information from the CPS turnout study as a baseline for comparing our estimates of turnout and vote method and the

[4] We also asked individuals how many adults live in their homes, a variable that can be validated with census data. However, a non-trivial percentage of mail respondents responded 0 to this question, indicating that some respondents may have read this question as asking for how many adults lived in their home, not including themselves. Since it is impossible to know what proportion of people who entered 1 or more for this question misread it similarly, we do not analyze this variable here.

national vote tally to validate the vote choice measure.

Table 2 presents the results from this comparison. Estimates from the surveys we fielded all showed lower home ownership rates than those reported by the Current Population Survey for 2010, though the Internet panel was farthest off on this measure. The Internet survey was the only mode that did not include the validated figure within its 95% confidence interval. Estimates of residential mobility were fairly accurate. Each of the modes produced a confidence interval for the proportion of Americans that had moved within the past year (or had last moved more than five years ago) that included the validated figure. Each of the surveys produced higher smoking rates than the NHIS, and the confidence intervals for these estimates often did not include the validated parameters. Both the Internet and mail survey came fairly close to accurately estimating the proportion of adult Americans without health insurance. However, the phone survey was seven percentage points too high in its estimate, and the validated figure fell beyond the range of its confidence interval.

The final set of measures in the table relate to whether respondents voted in the 2008 election, how they voted, and which candidate they reported voting for. Since the surveys were weighted for registration status, our measure of turnout is the percentage of those registered to vote who did so in 2008. The Current Population Survey estimated this figure to be 89.6%. The estimates produced by each mode were lower than this figure, but only the mail survey produced a confidence interval that did not include the CPS figure within its bounds. With regard to vote method, the phone survey significantly underestimated the percentage of voters who cast their ballots by mail or early in-person in 2008. The mail survey was also too low in its estimate of early in-person voting in 2008.

Finally, there were also significant deviations in reported vote choice and the actual vote

for president. The telephone survey actually estimated that McCain performed nearly 5 percentage points better than Obama in 2008, while the Internet survey also significantly underestimated Obama's support. The mail survey came closest to estimating the actual support each candidate received in 2008.

To summarize the amount of error entailed in each survey approach, we calculated the average difference between each mode's point estimates and the validated figures. In other words, this measure is simply the average of the absolute difference of each survey's point estimate and the validated figure in the right-hand column. On measures we could validate, both the Internet and mail surveys were off by an average of approximately 3 percentage points, while the telephone survey was off by an average of four points. Leaving aside the comparison between the Internet and telephone surveys, it is worth noting that an average error rate of 3 percentage points is precisely what we should expect for a sample size of 1,000.

Validated Comparison of Correlational Structures

So far, we have examined the accuracy of point estimates produced by our surveys relative to a validated baseline statistic. However, at least as important for social scientists is an understanding of how the correlational structure of the data produced by different modes compares. Fortunately, the CPS turnout study provides us with a good baseline against which to make such a comparison.

We begin by estimating three separate logit models using the CPS data. The first model uses an indicator for turnout (among those registered to vote) as the dependent variable, the second focuses on whether the voter voted by mail, and the third estimates whether the voter voted early in-person. The independent variables in the models were age, education, income,

home ownership, marital status, the respondent's tenure at the current address, and whether the respondent was female and African American.[5] We treat the coefficients estimated from the CPS data as the validated coefficients and then estimate the same models using data collected from each survey mode.

Figure 1 plots the coefficients from each mode against the valued for the CPS-generated coefficient (represented by the horizontal line). The figure also includes 95% confidence intervals for the coefficients generated by our surveys. Regardless of mode, nearly all of the coefficients estimated by the surveys were close to those estimated by the CPS model and in only a handful of cases did the CPS coefficient fall outside the range of the 95% confidence intervals. This pattern was also evident when we conducted joint F-tests for the equality in coefficients across models. The size of the F-statistic on these tests was never larger than 1.93, which meant we could not be 99% confident that any of the models estimated by one of our surveys differed from those estimated using the CPS data.

Comparison of Non-validated Political Measures

Figure 2 presents point estimates for a variety of measures of political attitudes and opinions that cannot be compared to any validated baseline. Thus, while we can look for

[5] The variables for home ownership, marital status, gender, and race were all indicators, with values of 1 if the respondent took on that characteristic. Age was simply the respondent's age in years. Education was included as an ordinal variable ranging from no high school (coded 1) to post-graduate degree (coded 6). Income took on 14 values ranging from less than $10,000 per year (coded 1) to more than $150,000 (coded 14). Tenure at the current address ranged from less than 1 month (coded 1) to more than 10 years (coded 7).

differences across modes in this figure, we are unable to determine which mode is most accurate on these measures. In general, the point estimates produced by different modes were relatively similar. However, there were five instances where the telephone and Internet surveys produced estimates that were significantly different from each other. Respondents to the phone survey were about 7 percentage points less likely than Internet respondents to say that budget cuts should come more from defense spending and they were six points less likely to say that they should come more from defense. Overall, telephone respondents were significantly more likely to take the middle position ("equally from both").

The telephone survey also produced a significantly lower estimate of the percentage of Americans who agreed more with the statement "the government is almost always wasteful and inefficient." The difference between these estimates was nearly 13 percentage points. The phone survey estimated significantly more support (approximately 12 percentage points) for affirmative action than the Internet survey. Finally, the phone and Internet generated statistically significant differences in their estimates of the percentage of the population either strongly or somewhat approving of Congress. The phone survey estimated that about 28 percent of Americans approved of the job Congress was doing while the Internet survey placed the estimate at just 19 percent.

The mail and Internet surveys differed significantly on five of the measures presented in Figure 2. The mail survey estimated higher levels of approval for Obama (about 8 percentage points higher) and Congress (12 points) as well as more support for increasing taxes on those earning more than $200,000 per year (8 points). The mail survey also registered less support for cutting more from domestic spending (7 percentage points less) and less agreement with the statement that the government is almost always wasteful and inefficient (7 points). The telephone

and mail surveys differed on three of the measures presented in the figure. The mail survey estimated 10 percentage points more support for taxing those earning more than $200,000 per year compared to the phone survey and it also estimated significantly more support for cutting more from defense (7 points). The phone survey estimated about 7 percentage points more support for affirmative action programs than the mail survey.

It is important to emphasize that while we do find some differences across the opinion and attitude measures presented in Figure 2, we cannot be sure which estimates are more "correct" given the lack of a validated baseline measure for these items. Nevertheless, understanding whether there are any patterns to the differences we uncovered may provide insight into whether any mode produces consistently more liberal or conservative estimates compared to the other modes. To answer this question, we scaled the questions asking respondents for their positions on issues like affirmative action, abortion, gay marriage, Social Security privatization, increasing taxes on incomes over $200,000, cutting government spending, and views toward government into a single standardized measure of political liberalism (Cronbach's alpha of .73). The resulting standardized measure registered higher values for respondents who gave consistently conservative responses to the issue questions and negative values for those offering more liberal answers. According to this measure, respondents to the mail survey provided the most liberal answers to the issue questions (mean of -.07), followed by the phone (-.02) and then Internet modes (.01). However, only the difference between the Internet and mail modes was statistically significant and the size of the difference was small (less than one-tenth of a standard deviation). Thus, even when the surveys provided different estimates to particular questions, the mode differences did not have a strong consistent ideological direction to them.

Mode Comparison on Political Knowledge, News Source, and Reported Contributions

In addition to the questions analyzed so far, the questionnaire also included questions asking respondents whether they made contributions to a political or religious organization during the previous year and what their primary source of news was. Respondents were also asked several factual questions about politics. The estimates produced for these questions are presented in Figure 3. Mode differences were more common on these measures. With regard to contributions, Internet respondents were significantly more likely to say that they had contributed to a political campaign, but phone respondents were more likely to report that they had made a contribution to a religious organization. News consumption also differed depending on the mode—Internet respondents were less likely to get their news from television than respondents from the mail or phone survey and more likely to report receiving their news online.

There were also consistent differences across mode for the political knowledge questions. Figure 3 presents the results from three factual questions—one asked respondents if they knew what the unemployment rate was, a second asked if they knew which party controlled the House of Representatives, and a third question asked respondents whether they knew the party that their state's governor belonged to.[6] On all three questions, the Internet survey reported the highest percentage of correct answers, followed by the mail survey and then the telephone poll. Over 50% of Internet survey respondents gave an accurate estimate of the unemployment rate compared to fewer than 40% of those answering the telephone survey. On the question asking respondents which party had a majority of seats in the House of Representatives, Internet

[6] For the unemployment rate question, respondents were coded as providing a correct answer if they gave a number between 8.7% and 10.7% (the actual figure at the time was 9.7%).

respondents answered correctly 68% of the time, compared to 58% for the mail survey and 54% for the telephone poll. Finally, 67% of Internet respondents knew the party of their state's governor, compared to 63% on the mail survey and 59% for the telephone poll.

One possible explanation for the higher levels of knowledge found on the Internet mode is that these respondents were able to look up the correct answers given that they were already online. However, web respondents were given only 30 seconds to answer each of the knowledge questions before they were pushed forward to the next question. They were also precluded from going back to those questions during the survey. Furthermore, we examined whether Internet respondents were more likely to answer these questions correctly when they took longer to complete the survey. We found no relationship between the amount of time it took a respondent to complete the survey and whether that respondent answered either question correctly.

Explaining Differences in Knowledge

One possible explanation for the mode differences in political knowledge questions is sample selection due to the opt-in nature of Internet samples; perhaps the more knowledgeable people on the Internet opt-in while the less knowledgeable go about their business. Another possible explanation is that this difference arises from differences in the sample frame: those on the Internet differ from those not on the Internet. We can address the latter possibility directly and it appears that almost all of the observed knowledge difference arises from sample frame, rather than sample selection.

The surveys asked respondents whether they had access to the Internet in their homes. Not surprisingly, there are major mode differences in Internet access. While 97% of respondents to the Internet survey had access to the Internet in their homes, the figure was just 75% for

respondents to the phone and mail surveys. To measure how political knowledge may be influenced by this difference in sampling frames, we present two analyses. First, it is important to understand whether political knowledge is affected by having Internet access at home. While some studies have found an association between Internet use and political knowledge, the effects are often small (see Kenski and Stroud 2006). In Figure 4, we use the telephone and mail respondents to compare the political knowledge of those with and without Internet access in their homes. The estimates in the figure show that those with Internet access were significantly more likely to answer the knowledge questions correctly, and that these differences were sometimes quite large. Most of these differences also persisted when we estimated logit models controlling for education, gender, race, age, party affiliation, and income. This indicates that simply weighting on these demographic and socioeconomic factors would not be sufficient to account for the independent effect of Internet access on political knowledge.

If the knowledge differences we observe across modes are related to the fact that the Internet mode over-samples individuals with Internet access, then we should find few mode-based knowledge differences when comparing knowledge for only respondents with Internet access. Thus, our second analysis (presented in Figure 5) repeats the knowledge comparisons presented in Figure 3, but restricts the analysis to only those respondents who report that they have Internet access in their homes. The figure indicates that once we control for Internet access, the mode-based differences in knowledge largely vanish. In fact, there are no statistically significant differences in the proportion of correct answers for any of the three questions. Why Internet users differ from non-Internet users is a subject for more extensive study. Using the web requires somewhat more basic literacy than using a phone or responding to a mail survey. Also, web use may make people more knowledgeable because of the amount of

information people are exposed to on-line. Understanding why these differences arise will help survey researchers understand differences in mode and make appropriate corrections. It will also help with the further development of survey methods and sampling techniques. The immediate lesson, though, is that the opt-in nature of the Internet survey we examined does not seem to be a major source of the mode difference in knowledge. That difference, which is a modest difference, appears to arise primarily from differences in sample frame.

Correlational Comparison for Non-Validated Dependent Variables

In this section, we analyze the extent to which mode differences appeared for regression coefficient estimates produced from models that cannot be compared to a validated baseline. Table 3 presents the first part of our analysis in this vein. For each mode, we regressed Obama's job approval on a standard set of demographic and political measures. The approval measure ranges from "strongly disapprove" (coded 1) to "strongly approve" (coded 4). The coefficients and standard errors for each mode's regression model are presented in the first three columns of the table. The latter three columns present F-statistics for tests of equality between these coefficients. Significant F-statistics for a particular variable would indicate that we can be more than 95% confident that the coefficient estimates differ across the modes. For almost every variable in the model, the F-statistics failed to attain statistical significance, indicating that there was little difference in the effects of these variables depending on mode. There were, however, two F-statistics that were significant at $p<.05$ and one that was significant at $p<.01$.

The most significant difference in coefficients came for the coefficient on age in the phone versus mail mode. In the phone survey, age had a statistically significant and negative effect on approval of Obama while the mail survey estimated the relationship to be positive and

insignificant. The other two cases of significant F-tests involved the model for the Internet survey compared to the mail survey. The regression model run with mail respondents produced a statistically significant and positive relationship between being female and approving of Obama, while the coefficient in the Internet model was negative and statistically indistinguishable from 0. The coefficient for income was statistically significant and negative for the Internet mode but took on a smaller (and statistically insignificant) negative value for the mail survey.

Despite these three instances of significantly different coefficients, the overall conclusion suggested by Table 3 is that the correlational structure of the data is altered very little across modes. In fact, of 33 F-tests between coefficients, only three attained statistical significance. It is also worth noting that there was not a single statistically significant difference in coefficients between the Internet and phone surveys.

Table 4 presents an extension of the type of tests we included in Table 4. Using the same independent variables listed in Table 3, we estimated OLS models for an additional set of dependent variables. These dependent variables included approval of Congress, attitudes on abortion restrictions, and support for affirmative action, gay marriage, social security privatization, and increasing taxes on individuals earning over $200,000. In each case, we estimated a different model for respondents to each survey mode. Thus, the table includes F-statistics for tests determining whether the regression coefficients estimated from one mode are jointly different from those estimated by another mode. For example, the first row of F-statistics indicates that in the models for Obama's approval rating, the set of coefficients estimated for phone respondents was statistically different from those estimated for mail respondents ($F = 1.82$, $p<.05$). However, significant differences did not exist between the Internet and phone survey or the Internet and mail survey.

The most notable pattern to arise from Table 4 is that the coefficients generated from Internet respondents were never statistically distinguishable from those generated from the phone survey. Each case of a statistically significant joint F-statistic involves the mail survey, either in comparison to the Internet or phone mode. Thus, to the extent that we find differences in the correlational structure of the data across modes, those differences are confined to the mail survey. These differences may be the result of the different sampling frame or manner in which the survey is administered. Alternatively, these differences may have resulted from the extended time frame over which respondents answered the mail survey versus the Internet of phone poll. Regardless of the cause, it is important to note that while the F-statistics are significant in four cases in Table 4, in none of these instances is the F-statistic very large. Thus, while we can be confident there are four instances where the correlational structure of the data differs, the size of this difference is not necessarily large.

Discussion

As noted earlier in the paper, our goal in this mode comparison is not to crown one approach as the "winner" over others, but to provide detail on the costs and benefits of each survey approach. Understanding whether researchers will reach different or less accurate conclusions by relying on one type of survey mode over another is of primary importance in such an accounting. In our analysis of indicators for which we had validated baseline values for comparison, none of the survey modes performed particularly poorly. The telephone survey produced the highest average error rate, at 4.1 percentage points, while the web and mail surveys produced error rates right around 3 percentage points. It is worth noting that a 3 percentage point error rate is precisely what sampling theory would lead us to expect from a survey with

approximately 1,000 respondents.

While our other analyses could not be tied to validated baselines, we typically found negligible differences across modes here as well. We examined a long list of questions gauging citizens' evaluations of political leaders, opinions on issues, reported behavior and news consumption, and answers to factual questions. The biggest cross-mode differences appeared on the latter measures, with Internet respondents demonstrating higher levels of political knowledge than respondents in the other two surveys (see also Hill et al. 2007).[7] Our analysis of the knowledge differences indicates that it is largely a sample frame issue—citizens with Internet access simply know more about politics and since Internet surveys over-represent such individuals in their sampling frame, they are bound to produce higher levels of knowledge. Indeed, once we limited our comparisons to only those with Internet access, the mode differences in knowledge disappeared.

While there were a few instances of significant differences in opinions and attitudes across modes, generally these differences were not large, particularly on the attitudinal measures. Furthermore, we found no strong tendency for any particular mode to produce consistently more liberal or conservative responses to questions about policy issues. Thus, even when there were differences, they did not appear to be systematic.

Our comparisons of the correlational structures of the data also uncovered few cross-mode differences. In our strongest test, we examined how logit models estimated from each survey stacked up against baseline models estimated with the CPS turnout study. Our analysis demonstrated that coefficients generated from each survey mode generally provided good

[7] Internet respondents were also less likely to report receiving their news from television and more likely to chose the Internet as their news source.

approximations of those generated from the CPS data. Furthermore, in comparing regression models we could not validate, we consistently found an absence of statistically significant differences in the coefficients produced by regression models estimated with the Internet survey compared to the telephone poll. There were some statistically significant differences in coefficients generated from the mail survey compared to the other two modes, but these differences were generally small and may have been primarily the result of the extended time period during which the mail survey was in the field.

Overall, it appears that researchers will not consistently get more accurate results, nor reach substantially different conclusions, when using one mode relative to another. That said, costs are undoubtedly an important consideration for most researchers. The mail mode was particularly expensive, both in terms of actual costs per completed interview and in terms of the extensive time period required to collect an adequate number of responses. The cost per interview for the mail mode was approximately 5 times greater than it was for the Internet survey and more than twice the cost of the telephone poll. Given these considerations and the point that the mail survey was slightly less accurate than the other two modes, we expect that it will not be a reasonable option for most researchers or practitioners.

The calculation may be a bit closer when comparing telephone and Internet surveys, but given that its error rate was precisely what it should have been given the sample size, the Internet survey would likely still be preferred by most researchers. Not only was the Internet survey half as expensive as the telephone poll, but it also took significantly less time to administer to respondents. Indeed, Table 1 shows that the median completion time for a telephone interview was about 60% longer than the median for an opt-in Internet respondent. Even individuals in the mail sample who chose to complete their survey online (people we might think of as Internet

survey novices) took nearly 3 minutes less to complete their questionnaires online than the median phone respondent took to complete a telephone interview. Thus, researchers interested in asking more questions during a survey may find the Internet mode more attractive.

Our finding that the Internet and telephone surveys performed so similarly runs counter to several recent papers (e.g. Yeager et al. 2009; Pasek and Krosnick 2010). We believe there are two main reasons for why some scholars find larger mode differences than we have uncovered here. First, as noted above, many studies, including those that have attracted significant attention recently, are based on data collected five or more years ago. The science of constructing, matching and weighting opt-in Internet panels has developed rapidly over the past decade at the same time that Internet use among the public has continued to increase. Second, and perhaps more importantly, our findings indicate that opt-in Internet panels *can* produce data that looks remarkably like that from a landline/cell telephone survey. However, just as with surveys executed through any mode, not all opt-in Internet panels are created equal and a poorly constructed Internet survey may produce inaccurate and biased results just as easily as a poorly designed telephone, mail, or in-person poll. Overall, our findings indicate that an opt-in Internet survey produced by a respected firm can produce results that are as accurate as those generated by a quality telephone poll and that these modes will produce few, if any, differences in the types of conclusions researchers and practitioners will draw in the realm of American public opinion.

Works Cited

AAPOR Opt-In Online Panel Task Force. 2010. “AAPOR Report on Online Panels.” Available at: http://www.aapor.org/AAPOR_Releases_Report_on_Online_Survey_Panels/2263.htm.

Blumberg, Stephen J., and Julian V. Luke. 2009. “Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July–December 2008.” Centers or Disease Control. Available at http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200905.pdf.

Chang, Linchiat and Jon A. Krosnick. 2009. “National Surveys via RDD Telephone Interviewing versus the Internet: Comparing Sample Representativeness and Response Quality.” *Public Opinion Quarterly* 2010 73(4): 641-678.

Chang, Linchiat and Jon A. Krosnick. 2010. “Comparing Oral Interviewing With Self-Administered Computerized Questionnaires: An Experiment.” *Public Opinion Quarterly* 74(1): 154-167.

Curtin, R., S. Presser, and E. Singer. 2005. ``Changes in Telephone Survey Nonresponse Over the Past Quarter Century." *Public Opinion Quarterly* 69: 87-98.

Groves, Robert M. 2006. “Nonresponse Rates and Nonresponse Bias in Household Surveys.” *Public Opinion Quarterly* 70: 646-675.

Hill, Seth J., James Lo, Lynn Vavreck, and John Zaller. 2007. “The Opt-in Internet Panel: Survey Mode, Sample Methodology, and the Implications for Political Research.” Unpublished Manuscript.

Keeter, Scott, Michael Dimock, and Leah Christian. 2008. Calling Cell Phones in ’08 Pre-Election Polls.” Available at http://people-press.org/reports/pdf/cell-phone-commentary.pdf.

Kenski, Kate and Natalie Jomini Stroud. 2006. "Connections Between Internet Use and Political Efficacy, Knowledge, and Participation." *Journal of Broadcasting & Electronic Media* 50(2): 173 – 192.

Krosnick, Jon A. 1999. "Survey Research." *Annual Review of Psychology* 50: 537-567.

Krosnick, Jon, Norman Nie, and Douglas Rivers (2005) Web Survey Methodologies: A Comparison of Survey Accuracy. Paper presented at the 60th Annual Conference of the American Association for Public Opinion Research in Miami Beach, Florida, May 12-15.

Malhotra, Neil and Jon A. Krosnick. 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 Nonprobability Samples." *Political Analysis* 15: 286-323.

Pasek, Josh and Jon A. Krosnick. 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." Statistical Research Division of the U.S. Census Bureau (#2010-15).

Rivers, Doug. 2009. ``Second Thoughts about Internet Surveys." Pollster.com (September 6). Available at: http://www.pollster.com/blogs/doug_rivers.php.

Rivers, Douglas. N.D. Sample Matching: Representative Sampling from Internet Panels. Available at: http://www.polimetrix.com/documents/YGPolimetrixSampleMatching.pdf.

Sanders, David, Harold D. Clarke, Marianne C. Stewart, and Paul Whiteley. 2007. "Does Mode Matter for Modeling Political Choice? Evidence from the 2005 British Election Study." *Political Analysis* 15: 257-285.

Yeager, David S., Jon A. Krosnick, LinChiat Chang, Harold S. Javitz, Matthew S. Levendusky, Alberto Simpser, Rui Wang. 2009. "Comparing the Accuracy of RDD Telephone

Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples." Unpublished Manuscript.

Table 1: Summary Information About Surveys

Mode	Sample Size	Field Dates	Response Rate	Median Completion Time
Internet	1,000	1/15/10 – 2/11/10	42.9% (RR1)	8.94 minutes
Mail	1,207	1/30/10 - 9/30/10	21.1% (RR3)	11.80 minutes*
Phone	907	1/28/10 – 1/30/10	19.5% (RR3)	14.33 minutes

* Timing only for mail recruits who took the survey online.

Table 2: Validation Comparison of Point Estimates by Mode

Item	Response	Internet	Phone	Mail	Validating Source
Home Ownership	Own	.613 (.573, .653)	.637 (.584 .690)	.636 (.594, .678)	.669 (CPS)
Mobility	Moved in past year	.152 (.121, .183)	.148 (.103, .192)	.146 (.112, .180)	.154 (ACS)
	At address 5 or more years	.555 (.515, .595)	.613 (.561, .665)	.548 (.506, .590)	.588 (ACS)
Smoked 100 Cigarettes	Yes	.504 (.464, .544)	.483 (.432, .534)	.481 (.440, .523)	.430 (NHIS)
Smoke Cigarettes Now	Every or some days	.259 (.222, .296)	.248 (.204, .293)	.223 (.186, .261)	.203 (NHIS)
Health Ins.	None	.157 (.128, .186)	.237 (.187, .287)	.154 (.120, .189)	.167 (SIPP)
Voted in 2008 (if registered)	Yes	.888 (.865, .911)	.876 (.841, .911)	.825 (.788, .861)	.896 (CPS)
Voting Method in 2008	By Mail	.191 (.161, .221)	.122 (.093, .150)	.164 (.133, .194)	.164 (CPS)
	Early In-Person	.136 (.111, .162)	.104 (.078, .130)	.105 (.082, .128)	.143 (CPS)
Vote choice in 2008	Obama	.482 (.444, .521)	.457 (.405, .508)	.553 (.512, .593)	.529
	McCain	.474 (.436, .513)	.502 (.450, .553)	.432 (.391, .472)	.456
Average difference		.031	.041	.029	

ACS = American Community Survey. NHIS = National Health Indicators Survey. CPS = Current Population Survey. SIPP = Survey of Income and Program Participation.

Table 3: OLS Estimates of Factors Affecting Obama Approval Ratings

	Coefficients (SEs)			F-Test Statistics		
Variable	Internet	Phone	Mail	Internet vs. Phone	Internet vs. Mail	Phone vs. Mail
Ideology	-.210***	-.186***	-.167**	0.10	0.26	0.06
	(.058)	(.048)	(.061)			
Right Track	.864***	1.078***	.833***	1.80	0.04	2.47
	(.115)	(.109)	(.111)			
Economy	-.249***	-.283***	-.302***	0.15	0.39	0.04
	(.059)	(.066)	(.061)			
Democrat	.518***	.409***	.496***	0.42	0.02	0.30
	(.129)	(.108)	(.114)			
Republican	-.181**	-.136	-.385**	0.13	2.27	2.45
	(.068)	(.108)	(.117)			
Age	-.002	-.008**	.002	3.43	1.84	7.33**
	(.002)	(.002)	(.003)			
Female	-.070	.143	.220**	3.35	6.64*	0.40
	(.075)	(.089)	(.084)			
White	-.140	-.073	-.167	0.20	0.03	0.30
	(.085)	(.125)	(.120)			
Education	-.006	-.053	.010	1.58	0.17	2.01
	(.021)	(.031)	(.031)			
Income	-.041***	-.030*	-.007	0.36	3.97*	1.52
	(.011)	(.014)	(.013)			
Born Again	-.025	-.073	-.123	0.02	0.73	0.86
	(.078)	(.093)	(.083)			
Intercept	3.594***	3.943***	3.197***			
	(.269)	(.243)	(.322)			
N	676	566	687			
R-squared	.668	.622	.617			

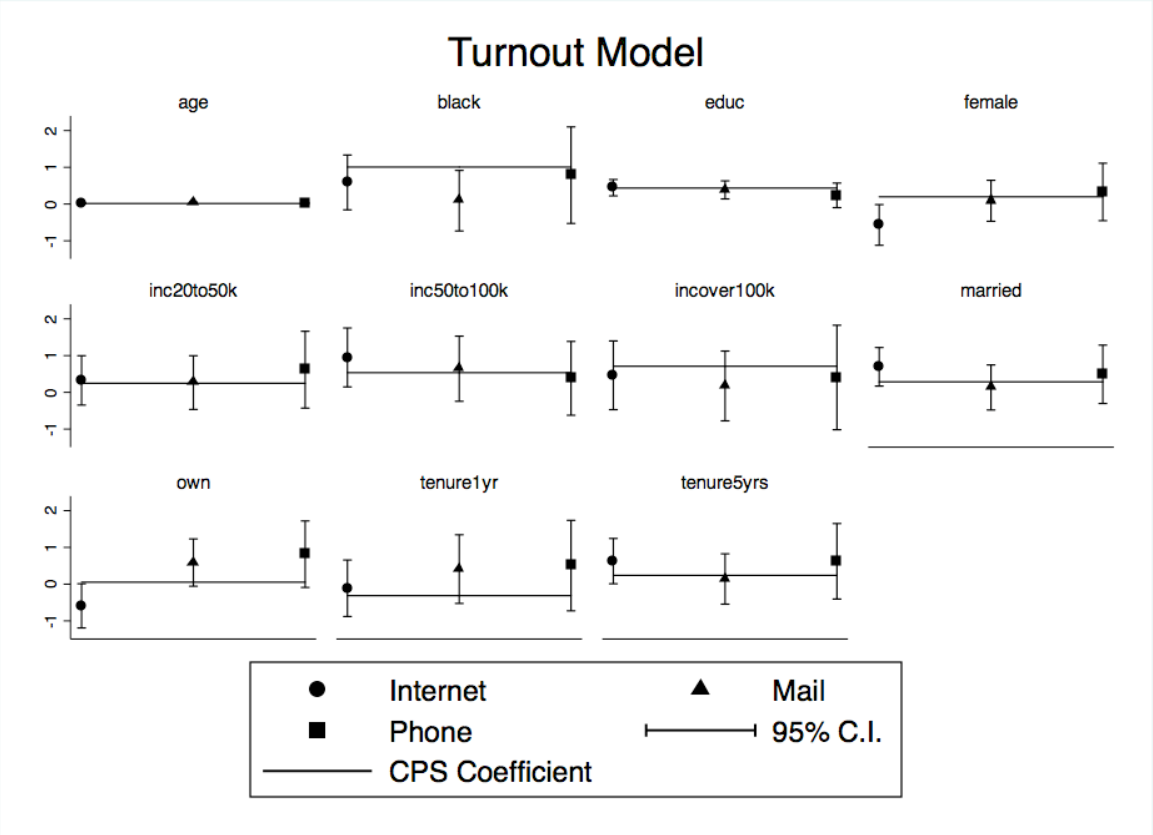
***p<.001, **p<.01, *p<.05.

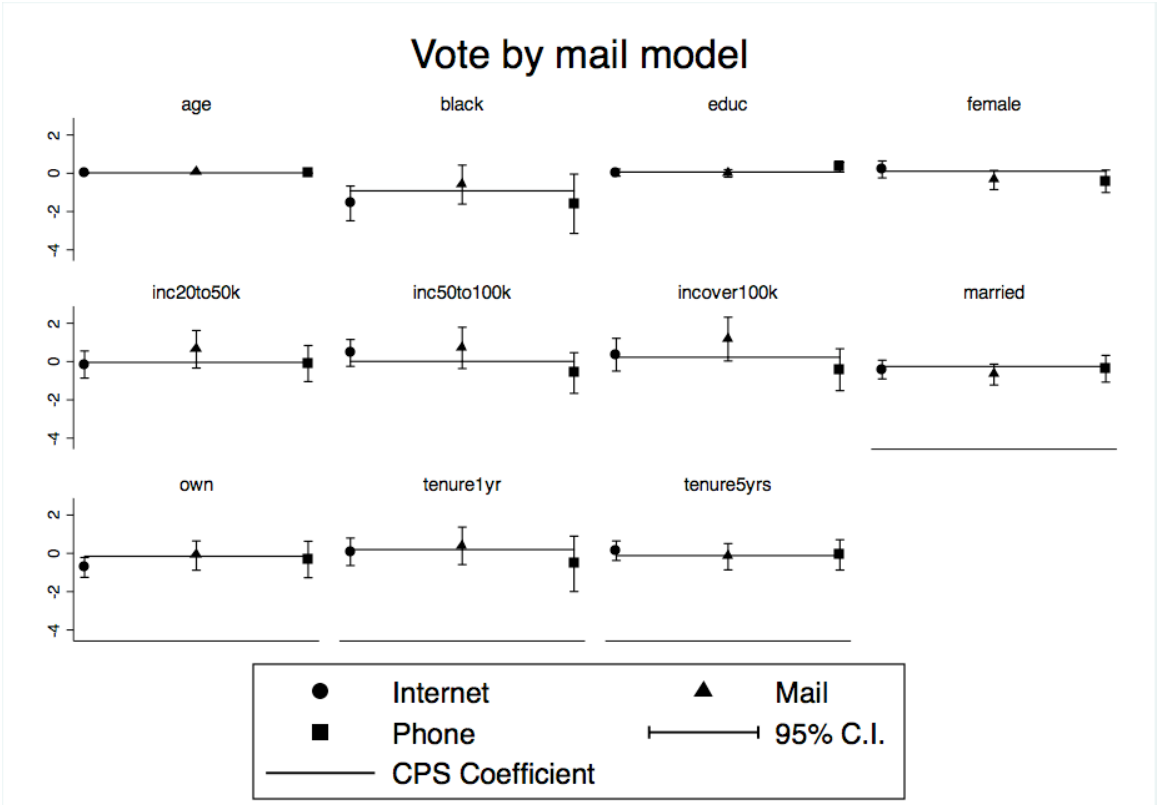
Table 4: Results from F-Tests Comparing Models Across Modes

Dependent Variable	Internet vs. Phone	Internet vs. Mail	Phone vs. Mail
Obama Approval	1.08	1.62	1.82*
Congressional Approval	1.03	0.50	0.40
Abortion Attitudes	1.69	1.63	1.10
Affirmative Action Support	1.20	1.50	1.46
Gay Marriage Support	1.32	0.87	1.93*
Social Security Privatization Support	0.95	2.16*	1.31
Increase Taxes on >$200k Support	1.38	2.36**	1.72

***p<.001, **p<.01, *p<.05

Figure 1: Comparison of Coefficients Generated by Each Mode with CPS-Validated Coefficients




Vote by mail model
age
black
educ
female
inc20to50k
inc50to100k
incover100k
married
own
tenure1yr
tenure5yrs
Internet
Mail
Phone
95% C.I.
CPS Coefficient

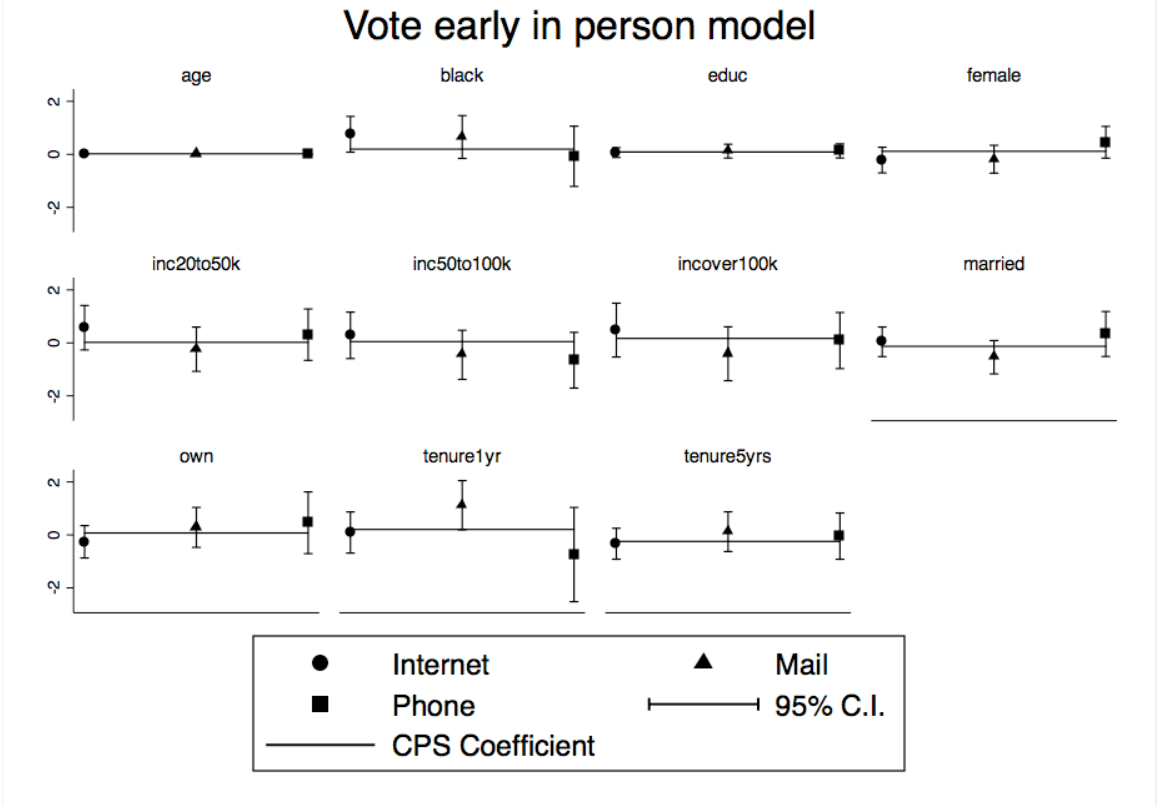
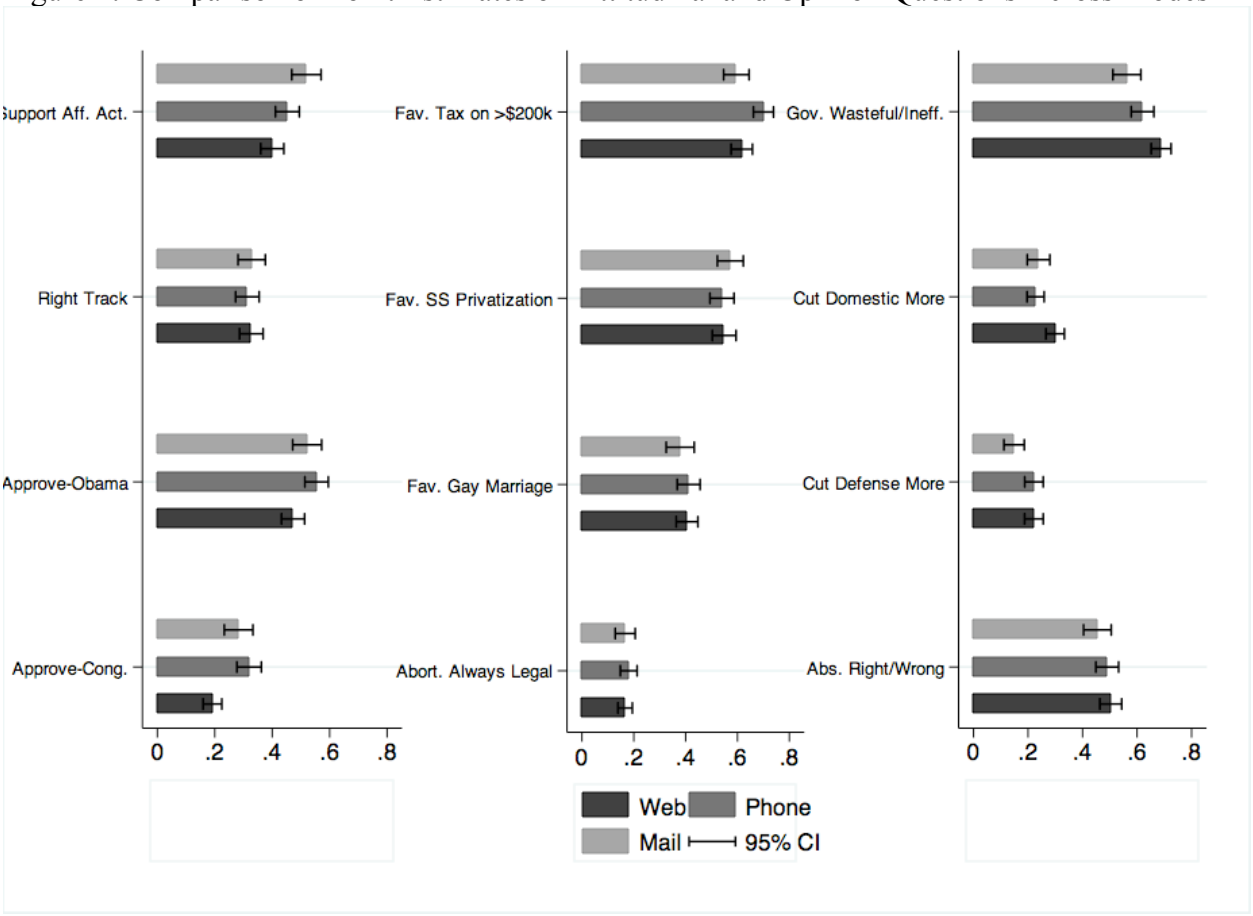
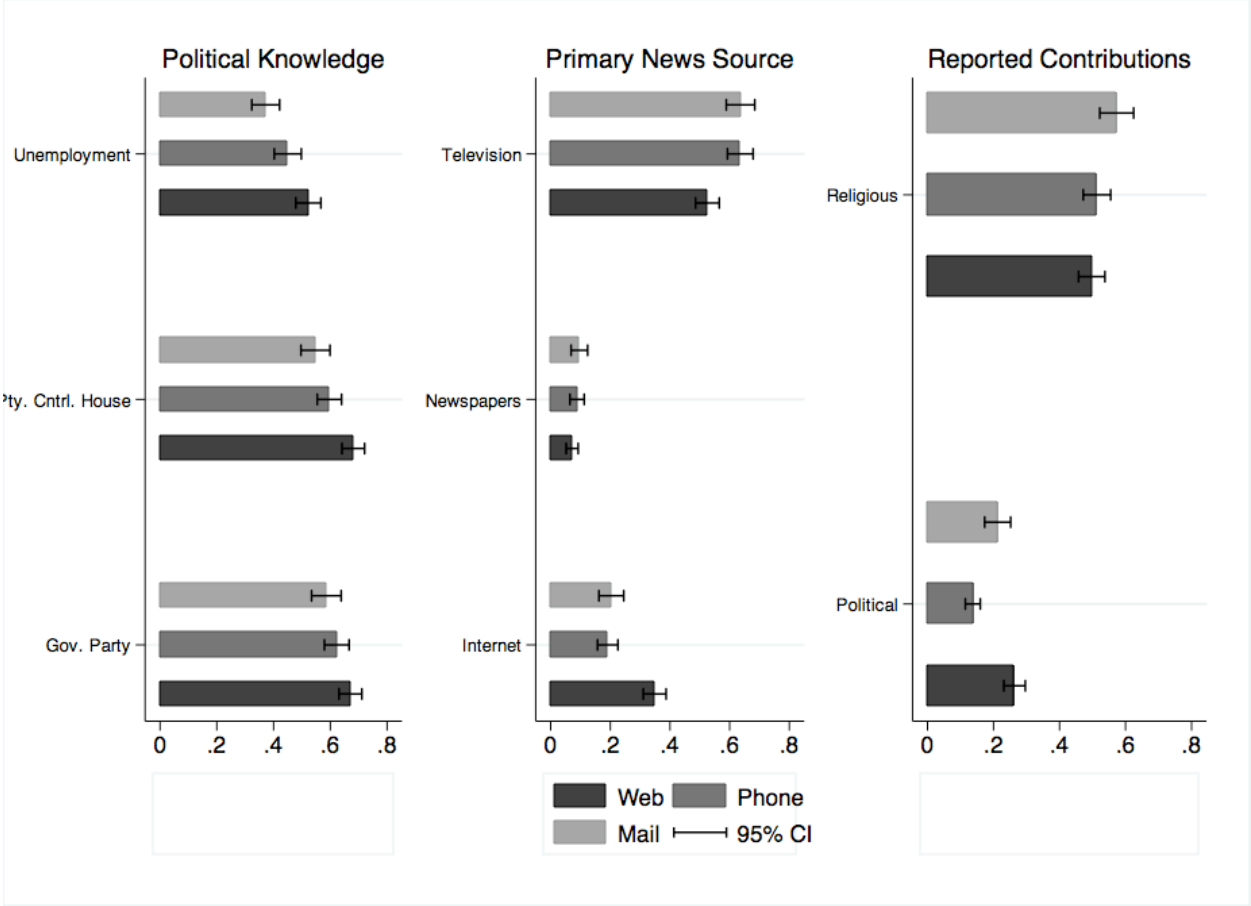

Vote early in person model
age
black
educ
female
inc20to50k
inc50to100k
incover100k
married
own
tenure1yr
tenure5yrs
Internet
Mail
Phone
95% C.I.
CPS Coefficient

Figure 2: Comparison of Point Estimates on Attitudinal and Opinion Questions Across Modes



NOTE: Entries are weighted proportions of respondents in each category after excluding those responding "don't know" or "not sure."

Figure 3: Comparison of Point Estimates on Knowledge, News Source, and Reported Contributions

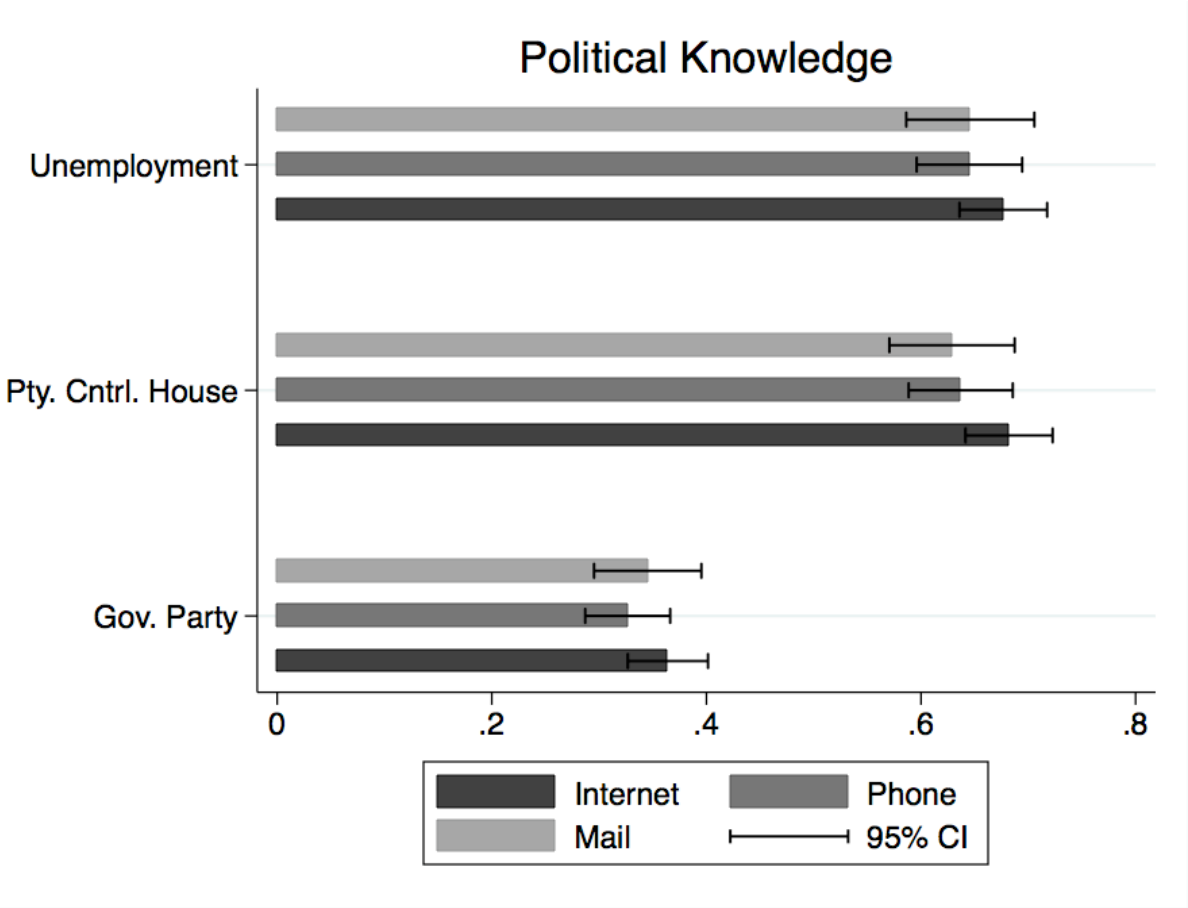


NOTE: Entries are weighted proportions of respondents in each category after excluding those responding "don't know" or "not sure" except for the knowledge questions where "don't know" or "not sure" was coded as incorrect.

Figure 4: Comparison of Point Estimates on Knowledge Between Respondents With and Without Internet Access at Home

NOTE: Entries are weighted proportions of respondents answering each question correctly; "don't know" or "not sure" were coded as incorrect.

Figure 5: Comparison of Point Estimates on Knowledge Among Respondents With Internet Access at Home



NOTE: Entries are weighted proportions of respondents answering each question correctly; "don't know" or "not sure" were coded as incorrect.