Re-Examining the Validity of Different Survey Modes for Measuring Public Opinion in the U.S.: 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 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 an indispensible tool for studying political problems ranging from political participation to voting behavior to approval of the government and support for public policy decisions. And, they are the main method for collecting essential data in other fields, such as sociology and economics. 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, and fewer 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 in the discipline. 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.

While social scientists and survey researchers now enjoy a much wider a range of options

for fielding a public opinion survey, guidance on the costs and benefits of different approaches is

sparse. For example, heated debate about the validity of Internet polls vis-à-vis telephone

surveys has failed to produce any obvious conclusions or recommendations. While the American

Association of Public Opinion Researchers (AAPOR) recently released a report that warned

against using opt-in Internet surveys to make inferences about population point estimates, the

report also noted that there not enough known about how these survey methods compare with

other approaches (AAPOR 2010, p. 54). Specifically, of all of the studies reviewed by AAPOR,

only 1 yielded consistent results across modes. As a case in point, one of the most discussed

mode studies conducted in recent years has been interpreted by some scholars as undermining

opt-in Internet surveys (Yeager et al. 2009) and by others as bolstering them (Rivers 2009).

Also, nearly all of the studies cited by the AAPOR report come from the beginning of the 2000s,

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 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.

This paper presents results from a recent study comparing opt-in Internet, telephone, and mail survey modes. Our analysis examines the extent to which each approach produces accurate point estimates for measures on which we have validated benchmark data, cross-mode similarities and differences in point estimates for political measures (such as attitudes and reported behavior) that cannot be validated, and the extent to which different modes produce datasets with different correlational structures.

Survey Research and the Changing Landscape of Communication Technologies

Two decades ago, pollsters could reach nearly the entire American adult population through a single method—random digit dialing. With cell phones still in their infancy, nearly all Americans had a landline telephone at their home, and since caller ID was not widely used, people were also compelled to answer their phones to find out who was calling. The rapid expansion of cellular service in the US has irrevocably changed the landscape of telephone-based survey research. Not only have most Americans taken to using cell phones, but a non-trivial proportion of the American public now relies exclusively on a cell phone. As recently as 2003, less than 5 percent of Americans reported that they were cell-phone-only; by the end of 2009, that figure exceeded 20 percent. In addition to this substantial portion of the American public that does not own a landline phone at all, an additional 15 to 20 percent of people who do have landline phones report that they take all or almost all of their phone calls over their cell phones (Blumberg and Luke 2009). When one combines these figures, 35 to 40 percent of the American public is now difficult or impossible to reach via landline telephones.

This trend has significant consequences for pollsters. Legally, cell phone exchanges cannot be dialed with the kind of automated technology that pollsters had grown accustomed to using to reduce the costs of conducting surveys over the phone. Having humans manually dial phone numbers requires more time, thereby increasing the cost of a cell phone interview relative to one conducted on a landline. There are other costs as well. The sample of cell phone numbers generally costs more to generate or purchase than a sample of landline numbers and unlike landlines, a large proportion of cell phone numbers belong to minors. In fact, one study reported that 42 percent of interviews initiated on cell phones had to be terminated because it was determined that the respondent was under the age of 18 (Keeter et al. 2008). Beginning so many conversations that do not result in valid interviews also costs valuable time and money for survey organizations. Finally, pollsters often offer to reimburse respondents for the cell minutes that they use in answering the survey. Altogether, an interview conducted with a respondent on a cell phone costs at least twice as much as a similar interview conducted over a landline (Keeter et al. 2008). Thus, the costs of conducting a valid telephone poll have increased significantly with the proliferation of cell phones.

Even when survey organizations can afford to sample both landline and cell phone numbers, there are still significant barriers to producing an accurate sample. The most significant barrier relates to the reluctance of contemporary Americans to answer telephone surveys. Generally speaking, response rates to contemporary RDD telephone polls are generally about 20%, meaning at least four-in-five individuals originally sampled do not actually take the survey. While much existing research suggests that lower response rates may not necessarily produce (or increase) nonresponse bias, they do place a higher premium on the specification of post-survey adjustments such as weighting (Groves 2006).

While the use of landline telephones has been in steady decline over the past decade, Internet penetration has been continually on the rise. By 2009, two-thirds of American households and three out of every four adults had access to the Internet on a daily basis. Americans send private emails and conduct financial transactions over the Internet, so they are unlikely to be concerned about privacy issues when answering survey questions online. The spread of Internet usage has provided an opportunity for social science research. Indeed, as many social scientists have already discovered, conducting surveys over an Internet platform has significant potential for reducing the cost of conducting surveys, for making the interview process more dynamic, and for potentially yielding higher quality responses. Despite these potential benefits, claims to the validity of Internet survey approaches persist.

Recent Mode Studies

Most recent mode studies have focused on the question of whether it is possible to use statistical methods to generate samples of respondents that are representative of the target population (Pasek and Krosnick 2010; Chang and Krosnick 2010; Malhotra and Krosnick 2007; Sanders et al. 2007). However, there is presently a lack of comprehensive and unambiguous data on how opt-in Internet surveys compare to other studies conducted with different sampling strategies and through different modes. Even the studies that have been conducted thus far have been interpreted differently in different research papers. For example, one study conducted in 2004 compared seven respected opt-in Internet panels to the probability-based Knowledge Networks panel and a probability-based telephone survey. Each survey firm was asked to administer a similar questionnaire to 1,000 adults who were representative of the American adult population. The questions used were chosen because they could be validated with government

data. The nine surveys were then compared to these population benchmarks to determine how close they were to the actual values they were attempting to estimate. Despite the strong design of this study, the findings are sufficiently ambiguous to allow both sides in the debate to cite it as evidence supporting their contentions. As Rivers (2009) concluded, there were some statistically significant differences in this comparison, but the practical error rate was only about 2%.

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. While AAPOR expressed caution about the use of opt-in Internet panels, the task force also stressed the relative lack of research that has been conducted on different approaches to conducting Internet surveys: "Despite the widespread use of online panels there is still a great deal that is not known with confidence" (2010, p. 54). Furthermore, much of what we know about this technology is already out-dated, as the technologies used to produce representative samples from these panels are continually evolving. Take, for example, the Hill et al. (2007) evaluation of the 2006 CCES, which found that respondents tended to have higher levels of political interests than other survey modes. As the authors noted in their conclusion, YouGov/Polimetrix was already adapting their technology at the time they were writing their paper to weight for items such as political interest and partisanship. By successfully weighting for these factors and purposefully recruiting less politically engaged individuals, future YouGov/Polimetrix samples may be less distinctive on these factors.

Overall, research comparing survey modes suffers from several drawbacks. First, techniques for conducting both telephone and Internet surveys have been rapidly evolving, and studies based on data collected several years ago will not be particularly valid for judging the current state of survey research. Indeed, most reputable telephone survey firms now conduct

interviews on both landlines and cell phones and opt-in Internet panels use increasingly

sophisticated weighting methods to create representative samples. Thus, even studies conducted

five years ago produce outdated comparisons.

Second, existing mode studies have been more focused on pitting different survey types

against each other than on unpacking the different components of survey data. Our aim is not to

pit different approaches against each other to determine a "winner" or to advocate for one

method over another, but rather to investigate the advantages and disadvantages (and costs and

benefits) of taking different approaches for different types of goals. For example, are the point

estimates of one mode significantly more accurate than those produced by other modes? How

would any gains in accuracy be weighed against any additional costs? Do some modes produce

consistently more liberal/conservative responses compared to others? And does the correlational

structure of the data vary from mode-to-mode?

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

(http://people.umass.edu/schaffne/mode_questionnaire.pdf). Table 1 provides summary

information for these surveys, and we describe each survey in greater detail below.

The Telephone Survey

8

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. Since not all people are equally likely to respond to recruitment efforts, 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.

9

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. Nonetheless, as we note below, we tested for any potential differences across platforms when conducting each of the analyses. 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 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. In addition to these factors, the telephone survey was also weighted for the number of landlines in each respondent's household. 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

We begin our analysis 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, and whether they currently smoked. Benchmark figures for these lifestyle questions were taken from the

12

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 also asked respondents whether they had voted in the 2008 presidential election and, if so, which candidate they voted for. We use information from the CPS turnout study to validate the turnout measure and the 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 also the only mode that did not include the validated figure within its 95% confidence interval. Estimates of residential mobility were also fairly accurate. Each of the modes produced a confidence interval for the proportion of Americans that had moved within the past year that included the validated figure. When it came to estimating the proportion of the population that had been at the same address for at least five years, only the mail survey was significantly off.

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. Only the phone survey produced a confidence interval of the proportion of Americans who had smoked 100 cigarettes in their lives that included the NHIS rate within its bounds. Similarly, the combined cell/landline survey was the only one to produce an accurate confidence interval for the proportion of the public who currently identify as smokers.

The final set of measures in the table relate to whether respondents voted in the 2008 election and which candidate they reported voting for. The Current Population Survey estimated that 89.6% of Americans who are registered to vote actually did vote in the 2008 election. 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. There were also significant deviations in reported vote choice and the actual vote for president. Specifically, the Internet and phone modes reported less support for Obama than he actually received.

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 values. 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. Overall, we found relatively minimal differences in accuracy across modes. This is especially the case when comparing the telephone and Internet surveys. The estimates from the telephone survey were off by an average of 3.5 points while the average error for the Internet survey was 3.6 percentage points. Despite the much longer field dates, the average error for the mail survey was less than a point higher at 4.3 percentage points.

Comparison of Non-validated Political Measures

Table 3 presents point estimates for a variety of political measures that cannot be compared to any validated baseline. Thus, while we can look for differences across modes in this table, we are unable to determine which mode is most accurate on these measures. The last column in the table indicates the range between the lowest and highest estimate for each response for ease in determining which measures created the largest cross-mode differences. In general, this table provides mixed results for the consistency of estimates produced across the modes. Of the 42 responses shown in Table 3, 15 produced differences smaller than 5 percentage points across the three modes, while there were 13 instances where the cross-mode differences exceeded 10 percentage points.

Evaluations of the country's general direction ("right track") and the economic situation were relatively similar regardless of mode. There were also few major differences across modes for respondents' positions on policy issues such as abortion, gay marriage, and Social Security privatization, though the mail survey showed significantly more support for taxing individuals making more than $200,000 per year. However, larger cross-mode differences were evident on some opinion items. For example, the Internet survey produced a substantially higher proportion of people saying that the government was "almost always wasteful and inefficient" than either the phone or mail survey.

There were several differences with regard to reported behavior as well. The phone survey reported a much higher proportion of respondents who said that they had voted in person on election day. Internet respondents were 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.

Some of the largest differences across modes appeared for the political knowledge questions. Table 3 presents the results from two questions—one asked respondents if they knew what the unemployment rate was and the other asked if they knew which party controlled the House of Representatives.[1] For the former, 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%). On both 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

---

[1] Two additional knowledge questions were asked in the questionnaire but we have not yet been able to analyze these questions.

accurate estimate of the unemployment rate compared to fewer than 40% of those answering the telephone surveys. On the question asking respondents which party had a majority of seats in the House of Representatives, Internet respondents answered correctly 68% of the time, compared to 58% for the mail survey and 54% for the telephone poll.

The differences we find on the knowledge questions may indicate that Internet panelists are more knowledgeable about politics, or it may simply be the case that some Internet respondents took the time to look up the information to make sure they answered correctly. To adjudicate between these explanations, 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. Thus, individuals who are captured by the Internet survey's sampling frame may simply be more knowledgeable about politics, even after weighting for education and other factors. We discuss this possibility in more detail below.

The final two rows in Table 3 shows the average difference in point estimates across each mode pairing. For example, the Internet and phone surveys produced estimates that were, on average, off by 6.2 percentage points. The phone mode was off by an average of 4.2 points from the mail survey, and the mail survey estimates differed from those produced by the Internet mode by 5.1 points. It is also worth noting that when the calculation included only the attitudinal measures in the table (the rows ranging from economic evaluations through the "right and wrong" question) the Internet mode was less distinct from the telephone and mail surveys. Across these measures, the Internet estimates differed from the phone estimates by an average of 5.2 percentage points and were off from the mail estimates by an average of 4.4 points.

Finally, we examined whether any of the modes produced estimates that were consistently more liberal or conservative than the others. To do so, 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 .72). According to this measure, respondents to the mail survey provided the most conservative answers to the issue questions, followed by the phone and then Internet modes. However, only the difference between the Internet and mail modes was statistically significant and the size of the effect was small (about one-tenth of a standard deviation).

Comparing the Correlational Structure of the Data

The analyses presented above focus exclusively on comparing point estimates across the survey modes. However, point estimates are just part of what social scientists and practitioners produce from their analyses of survey data. To what extent may different survey modes affect our analyses of the relationships between variables of interest? In this section, we focus on answering this question by examining the extent to which different modes provide different estimates of the relationships between variables in regression models.

Table 4 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 is presented in the first three columns of the table. The latter three columns present F-statistics for tests of equality between

17

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 5 presents an extension of the type of tests we included in Table 4. Using the same independent variables listed in Table 4, 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

18

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.

Two notable patterns arise from Table 5. First, the coefficients generated from Internet respondents were never statistically distinguishable from those generated from the phone survey. Additionally, 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.

Questionnaire Administration in the Mail Survey

As noted above, individuals solicited for participation in the mail survey were invited to

return their questionnaire by mail or by completing it online. Since over one-fourth of the respondents ultimately chose to fill out the questionnaire online, we are able to determine whether the act of taking the survey online produced different response patterns. Of course, since respondents selected their own response method, any differences between these groups could be attributed either to the platform itself or the propensity of certain types of people to prefer taking surveys on the web rather than on paper.

Table 6 compares responses from the web versus those received by mail for the measures on which we have validated benchmarks. Recall from Table 2 that, overall, the mail survey departed from the validated measures by an average of 4.3 percentage points. Table 6 shows that responses collected online departed by an average of 5 points while paper responses were off by 5.9 points on average. Two differences between web and paper respondents are particularly noteworthy. First, respondents submitting their questionnaires online were significantly 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.

Using the same response options listed in Table 3, we also constructed a measure of the average difference between mail respondents who completed their form online and those who mailed their questionnaire back. The average difference between these groups was just 5.1 percentage points on this set of variables and there were only three instances where the difference was greater than 10 points. Support for social security privatization was 11 percentage points higher among those answering online compared to respondents returning their questionnaires by mail. Web respondents were also 11.1 percentage points less likely to "somewhat approve" of Congress and 12.1 points more likely to "somewhat support" affirmative action. Overall, however, the differences between web and paper respondents were relatively

small.

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 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, there was no approach that stood out as being substantially more or less accurate. Indeed, the average error in the telephone survey was just .1 percentage points smaller than that for the Internet survey and only .8 points smaller than the mail mode.

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 the economy and 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 significantly higher levels of political knowledge and indicating a greater propensity to make political contributions than respondents in the other two surveys (see also Hill et al. 2007).[2] As we noted above, the differences in political knowledge do not seem to be driven by some Internet respondents looking up the right answers while taking the survey online. One potential explanation is that citizens with Internet access simply know more about politics; since Internet surveys over-represent such individuals in their sampling frame, they are bound to produce higher levels of

---

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

knowledge. Indeed, in preliminary analyses we have conducted on the mail and phone modes, we find that individuals with Internet access at home are significantly more likely to answer the knowledge questions correctly, even after controlling for factors (like education, age, and income) that are used to weight the surveys.

While there were a few instances of larger differences across modes, generally these differences were not large, particularly on the attitudinal measures. The average difference between the Internet and phone surveys on these measures was 5.2 percentage points while it was 4.4 points between the Internet and mail survey. The average difference between the phone and mail surveys was 4.2 points. 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 fact, 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 similar accuracy rates, 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 studies (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 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.

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)	.632 (.575 .688)	.632 (.582, .681)	.669 (CPS)
Mobility	Moved in past year	.152 (.121, .183)	.155 (.105, .204)	.162 (.115, .209)	.154 (ACS)
	At address 5 or more years	.555 (.515, .595)	.609 (.553, .664)	.519 (.469, .569)	.588 (ACS)
Smoked 100 Cigarettes	Yes	.504 (.464, .544)	.471 (.417, .524)	.497 (.448, .547)	.430 (NHIS)
Smoke Cigarettes Now	Every or some days	.259 (.222, .296)	.242 (.197, .287)	.241 (.191, .291)	.203 (NHIS)
Voted in 2008 (if registered)	Yes	.888 (.865, .911)	.876 (.842, .911)	.821 (.777, .864)	.896 (CPS)
Vote choice in 2008	Obama	.482 (.444, .521)	.454 (.401, .507)	.553 (.505, .598)	.529
	McCain	.469 (.436, .513)	.505 (.451, .558)	.431 (.386, .480)	.456
Average difference		.036	.035	.043	

Shaded areas indicate that the validated figure is not within the bounds of the confidence interval.
ACS = American Community Survey. NHIS = National Health Indicators Survey. CPS = Current Population Survey.

Table 3: Comparison of Non-Validatable Political Point Estimates by Mode

Item	Response	Internet	Phone	Mail	Range
Economy	Getting Better	0.233	0.182	0.184	0.051
Track	Right Track	0.327	0.329	0.312	0.017
Obama	Strongly Approve	0.176	0.251	0.234	0.075
	Somewhat Approve	0.296	0.277	0.330	0.053
	Somewhat Disapprove	0.161	0.176	0.159	0.017
	Strongly Disapprove	0.367	0.296	0.277	0.090
Congress	Strongly Approve	0.021	0.047	0.040	0.026
	Somewhat Approve	0.171	0.243	0.296	0.125
	Somewhat Disapprove	0.300	0.252	0.333	0.081
	Strongly Disapprove	0.508	0.457	0.330	0.178
Member	Strongly Approve	0.114	0.131	0.100	0.031
	Somewhat Approve	0.374	0.441	0.435	0.067
	Somewhat Disapprove	0.274	0.239	0.253	0.035
	Strongly Disapprove	0.237	0.189	0.211	0.048
Abortion	Always Legal	0.167	0.164	0.177	0.013
	Legal with some restrictions	0.327	0.259	0.354	0.095
	Legal only in special circumstances	0.368	0.402	0.335	0.067
	Always illegal	0.138	0.175	0.134	0.041
Affirmative Action	Strongly support	0.113	0.231	0.146	0.118
	Somewhat support	0.287	0.294	0.306	0.019
	Somewhat oppose	0.263	0.211	0.281	0.070
	Strongly oppose	0.336	0.265	0.267	0.071
Gay Marriage	Favor	0.406	0.381	0.426	0.045
Invest SS	Support	0.549	0.578	0.555	0.029
Tax over $200k	Favor	0.617	0.599	0.701	0.102
Cut Spending	More from defense	0.222	0.149	0.216	0.073
	About equally from both	0.479	0.612	0.563	0.133
	More from domestic	0.300	0.240	0.221	0.079
Government	Almost always wasteful and inefficient	0.689	0.560	0.614	0.129
	Does a better job than given credit for	0.312	0.441	0.386	0.129
Right and Wrong	Everyone has to decide themselves	0.497	0.543	0.509	0.046
	Absolute standards	0.504	0.458	0.491	0.046
Voting Method	In person	0.675	0.779	0.641	0.138
	Early in person	0.135	0.101	0.088	0.047
	Mail	0.189	0.118	0.140	0.071
Religious Contrib.	Yes	0.498	0.566	0.510	0.068
Political Contrib.	Yes	0.264	0.211	0.130	0.134
Knowledge 1	Correct House Party Control	0.681	0.541	0.577	0.140
Knowledge 2	Correct Unemployment Rate	0.523	0.371	0.432	0.152
News Source	Television	0.525	0.642	0.650	0.125
	Newspapers	0.072	0.094	0.083	0.022
	Internet	0.349	0.202	0.187	0.162
		Internet vs. Phone	Phone vs. Mail	Internet vs. Mail	
Average Difference	(All measures)	.062	.042	.051	
Average Difference	(Attitudinal measures only)	.052	.042	.044	

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

Table 4: OLS Estimates of Factors Affecting Obama Approval Ratings

	Coefficients (SEs)			F-Test Statistics		
				Internet vs.	Internet vs.	Phone vs.
Variable	Internet	Phone	Mail	Phone	Mail	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 5: 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

Table 6: Validation Comparison of Point Estimates by Mode of Response for Mail Survey

Item	Response	Web	Paper	Validating Source
Home Ownership	Own	.720 (.622, .817)	.598 (.541 .655)	.669 (CPS)
Mobility	Moved in past year	.200 (.092, .308)	.147 (.097, .197)	.154 (ACS)
	At address 5 or more years	.494 (.388, .599)	.528 (.472, .584)	.588 (ACS)
Smoked 100 Cigarettes	Yes	.437 (.328, .546)	.520 (.465, .574)	.430 (NHIS)
Smoke Cigarettes Now	Every or some days	.171 (.064, .278)	.268 (.211, .324)	.203 (NHIS)
Voted in 2008 (if registered)	Yes	.825 (.713, .936)	.819 (.779, .859)	.896 (CPS)
Vote choice in 2008	Obama	.477 (.377, .578)	.581 (.532, .630)	.529
	McCain	.504 (.401, .606)	.405 (.357, .453)	.456
Average difference		.050	.059	

Shaded areas indicate that the validated figure is not within the bounds of the confidence interval.
ACS = American Community Survey. NHIS = National Health Indicators Survey. CPS = Current Population Survey.