Advance Access publication May 28, 2014

Political Analysis (2014) 22:285–303
doi:10.1093/pan/mpt025

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

Stephen Ansolabehere
Professor of Government, Harvard University, 1737 Cambridge Street, CGIS Knafel Building 410, Cambridge, MA 02138
e-mail: sda@gov.harvard.edu

Brian F. Schaffner
Professor of Political Science, University of Massachusetts, Thompson Hall 316, 200 Hicks Way, Amherst, MA 01003
e-mail: schaffne@polsci.umass.edu (corresponding author)

Edited by Jonathan Katz

In this article, 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 Random Digit Dialing (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 using a Total Survey Error approach, 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.

1 Introduction

The rapid increase of Internet penetration in American homes has made web-based polling a viable and affordable alternative for students of public opinion. A particularly affordable method for conducting Internet surveys relies on recruiting volunteers to take Internet polls and then generating representative samples of the target populations of interest from these panels using a technique called sample matching. This is the approach used by YouGov (formerly Polimetrix), which is the firm most commonly used by political scientists conducting research using Internet surveys that utilize opt-in panels.[1] The popularity of this approach has been reflected in the publication of opt-in Internet survey data in the top journals of political science. Between 2006 and 2011, thirty articles published in the *American Political Science Review*, *American Journal of Political Science*, or *Journal of Politics* utilized opt-in Internet survey data produced by YouGov.[2]

Among public opinion researchers, there is a lack of consensus about the acceptability of opt-in Internet surveys. 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 and ultimately concluded that, "despite the widespread use of online panels, there is still a great

[1]The other Internet survey firm commonly used by political scientists is Knowledge Networks. Knowledge Networks constructs its samples from a different model. They begin by using random digit dialing and other probability selection methods to recruit individuals for their panel. They then sample from this panel to generate samples for their clients.
[2]This figure is based on a search of Google Scholar and includes articles that rely on surveys conducted by Polimetrix before it merged with YouGov.

© The Author 2014. Published by Oxford University Press on behalf of the Society for Political Methodology.
All rights reserved. For Permissions, please email: journals.permissions@oup.com

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

deal that is not known with confidence" (2010, 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).[3] 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 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 article 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 National Health Indicators Survey [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 US 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).

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

2 Design of the 2010 Mode Study

In early 2010, we commissioned YouGov America 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 target population for each sample was American adults. The questionnaire mostly focused on politics, but also included several lifestyle measures that could be validated using government data. The full questionnaire is available at the replication page for this article on Dataverse, as are all data necessary to replicate the analyses that follow. Table 1 provides summary information for these surveys, and we describe each survey in greater detail below.

Table 1 Summary information about surveys

Mode	*Sample size*	*Field dates*	*Response rate*	*Median completion time (min)*
Internet	1000	1/15/10 to 2/11/10	42.9% (RR1)	8.94
Mail	1207	1/30/10 to 9/30/10	21.1% (RR3)	11.80*
Phone	907	1/28/10 to 1/30/10	19.5% (RR3)	14.33

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

[3]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).

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

2.1 *The Telephone Survey*

The telephone survey was conducted January 28–30, with 807 interviews completed with respondents reached via landline numbers and a supplement of one hundred interviews conducted using cell phone numbers. Respondents were selected using random digit dialing, with the cell phone supplement coming from random digit dialing of known cell phone exchanges. Live interviewers were used to administer the questionnaire to respondents. Each telephone number was attempted up to six times before it was dropped from the sample. The response rate for the landline portion of the sample (RR3) was 20.9%, whereas 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 min 20 s.

2.2 *The Opt-In Internet Survey*

The Internet sample for our study came from the YouGov 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 uses targeted advertising to focus particular attention on recruiting groups that are underrepresented on their panel, such as racial and ethnic minorities.

Since YouGov does not use probability sampling to recruit panelists, they instead rely on sample matching to generate representative samples from their panel. When YouGov 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 one thousand American adults, YouGov might draw a random sample from the Census Bureau's American Community Survey (ACS) and use this as the target for constructing a sample from their own panel. In addition to all of the demographic information that is part of the ACS, YouGov is also able to weight on additional factors that it has matched to the ACS. For example, data on reported voter registration and turnout from the Current Population Survey (CPS) are matched to this frame using a weighted Euclidean distance metric. Data on religion, church attendance, born-again status, interest in current affairs, party identification, and ideology were matched from the 2007 Pew US Religious Landscape Survey. Thus, once YouGov draws the target sample from the database, they know what each member of their random sample should look like on a range of demographic, political, and religious 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 2006). After matching everyone in the target sample with at least one person from the Internet panel, YouGov 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, the target sample was selected from the 2005 to 2007 ACS stratified on age, gender, race, and education. 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 15, and responses were accepted through February 11. The panel produced one thousand responses, and the within-panel response rate (RR1) for this study was 42.9%. 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 min 56 s.

2.3 *The Mail Survey*

The mail survey was generated by mailing questionnaires to 6600 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%.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

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 (Atkeson and Tafoya 2008; Atkeson et al. 2011). Indeed, we found few major differences between these two groups and, accordingly, we analyze all mail respondents together in the analyses that follow.[4] Interestingly, individuals completing the questionnaire online took nearly 3 min longer than respondents from the YouGov panel (but still shorter than the telephone interviews).

2.4 *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 22 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 (Duncan 1979).

The extensive data collection period for the mail survey is, of course, an important reality for scholars and practitioners considering the mail mode (see also Atkeson and Tafoya 2008; Atkeson et al. 2011). 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 article.

2.5 *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. This variable was used to account for the fact that survey respondents are often more politically engaged than the general population. 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.[5] 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).

[4]Using the measures presented in Table 2, we found that responses collected from mail respondents online departed by an average of 4.5 points, whereas 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 twelve percentage points more likely to own their home. Second, paper respondents were more likely to report that they voted for Obama, whereas web respondents reported more support for McCain (the difference was about ten 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.

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

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

3 Total Survey Error Comparison

In this study, we use the Total Survey Error (TSE) approach to determine "among a set of alternative designs, the design that gives the smallest total survey error" (Groves and Lyberg 2010, 850). In doing so, we recognize that these different survey modes vary along several different components (including both the sampling and administration of the surveys) and that each of these differences may serve to enhance or limit the TSE of those surveys. Our interest is in providing guidance to social scientists who face several common options when contracting with survey firms. By comparing the TSE of these different approaches, we can provide information about the extent to which a particular approach varies substantially from other options. As Smith notes, "The TSE paradigm is a valuable approach for comparative studies . . . it goes beyond examining the separate components of error and provides a framework for the combining of the individual error components into their overall sum" (2011, 475). It is this overall TSE that we seek to evaluate in this analysis.

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 valid approximation to the population parameter. Thus, we begin by examining the TSE of each mode by determining the extent to which each survey 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 one hundred cigarettes during their lifetime, whether they currently smoked, and whether they had health insurance.[6] Benchmark figures for these lifestyle questions were taken from the ACS, the CPS, and the 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 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 homeownership rates than those reported by the CPS 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 CPS 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

[6]We also asked individuals how many adults live in their homes, a variable that can be validated with census data. However, a nontrivial percentage of mail respondents responded 0 to this question, indicating that some respondents may have read this question as asking 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.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

Table 2 TSE comparison of point estimates by mode

Item	*Response*	*Internet*	*Phone*	*Mail*	*Validating source*
Home ownership	Own	**0.613** (0.573, 0.653)	0.637 (0.584, 0.690)	0.636 (0.594, 0.678)	0.669 (CPS)
Mobility	Moved in past year	0.152 (0.121, 0.183)	0.148 (0.103, 0.192)	0.146 (0.112, 0.180)	0.154 (ACS)
	At address five or more years	0.555 (0.515, 0.595)	0.613 (0.561, 0.665)	0.548 (0.506, 0.590)	0.588 (ACS)
Smoked one hundred cigarettes	Yes	**0.504** (0.464, 0.544)	**0.483** (0.432, 0.534)	**0.481** (0.440, 0.523)	0.430 (NHIS)
Smoke cigarettes now	Every or some days	**0.259** (0.222, 0.296)	**0.248** (0.204, 0.293)	0.223 (0.186, 0.261)	0.203 (NHIS)
Health Ins.	None	0.157 (0.128, 0.186)	**0.237** (0.187, 0.287)	0.154 (0.120, 0.189)	0.167 (SIPP)
Voted in 2008 (if registered)	Yes	0.888 (0.865, 0.911)	0.876 (0.841, 0.911)	**0.825** (0.788, 0.861)	0.896 (CPS)
Voting method in 2008	By mail	0.191 (0.161, 0.221)	**0.122** (0.093, 0.150)	0.164 (0.133, 0.194)	0.164 (CPS)
	Early in-person	0.136 (0.111, 0.162)	**0.104** (0.078, 0.130)	**0.105** (0.082, 0.128)	0.143 (CPS)
Vote choice in 2008	Obama	**0.482** (0.444, 0.521)	**0.457** (0.405, 0.508)	0.553 (0.512, 0.593)	0.529
	McCain	0.474 (0.436, 0.513)	0.502 (0.450, 0.553)	0.432 (0.391, 0.472)	0.456
Average difference		0.031	0.041	0.029	
MSE		0.013	0.021	0.013	

Note. SIPP = Survey of Income and Program Participation. Estimates in bold had confidence intervals that did not include the validated figure.

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 five percentage points better than Obama in 2008, whereas the Internet survey also significantly underestimated Obama's support. The mail survey came closest to estimating the actual support each candidate received in 2008. In addition to estimating shares of all votes, one may estimate each candidate's share of the two-party vote. Political scientists often study just the two-party vote because nearly third parties receive a very small share of the vote in the United States. Obama's share of the two-party vote did not deviate significantly from the actual share of the two-party vote that he received in either the mail sample or the Internet sample; his estimated share of the vote in the phone sample did deviate significantly from Obama's actual share of the two-party vote.

One potential source of mode differences in these estimates would be social desirability bias—the tendency of some respondents to give responses to sensitive questions that are not entirely truthful. Specifically, research consistently demonstrates that self-administered questionnaires produce less social desirability bias than when a questionnaire is administered by an interviewer (Link and Mokdad 2005). This pattern would lead us to expect that the phone mode might underestimate rates of cigarette smoking (an undesirable activity) and overestimate rates of voter turnout (a desirable activity). Interestingly, this is not what we find in Table 2. The phone survey produced estimates of smoking rates that were higher than those from the NHIS, and the turnout rate from the phone survey was actually lower than the rate reported by the CPS. Interestingly, the largest

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

error rates for the phone poll were on questions where we would not expect social desirability bias, such as those asking about how the individual voted (by mail or in person) or whether the individual was without health insurance.

To summarize the amount of error entailed in each survey approach, we calculated the average deviation 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 three percentage points, whereas 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 three percentage points is approximately what we should expect for a sample size of one thousand.

The most common measure of TSE is mean squared error (MSE)—the squared average deviation of a survey estimate from the true value of the parameter being estimated. As Biemer notes, "A small MSE indicates that the TSE is small and under control. A large MSE indicates that one or more sources of error are adversely affecting the accuracy of the estimate" (2010, 826). The calculation is simply the average of the squared differences between the estimate produced by each survey and the validated figure. Both the Internet and mail surveys had lower MSEs than the telephone survey, but the TSE was reasonably small for each of the surveys.

3.1 *Validated Comparison of Correlational Structures*

So far, we have analyzed the TSE for each mode by assessing 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. The 2008 CPS turnout study was based on over sixty thousand interviews with American citizens during November of the election year. The study focuses exclusively on voter turnout, which limits the range of variables available for us to analyze. However, we focus on three dependent variables that tend to be of particular importance to scholars of American politics: (1) whether the individual reported voting, and among those who did vote (2) whether they voted by mail, and (3) whether they voted early in person. We limit the analysis to individuals who are registered to vote since each of the surveys was weighted based on voter registration.

Based on the dependent variables described above, we estimate three separate logit models using the CPS data. The independent variables in the models were age, education, income, homeownership, marital status, the respondent's tenure at the current address, and whether the respondent was female and African American.[7] 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).[8] The figure also includes 95% confidence intervals for the coefficients generated by our surveys. 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

[7]We did not include a variable for Latinos in this model because the phone sample included just twenty-one self-identified Hispanics. The variables for homeownership, 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 college degree (coded 5). Tenure at the current address was included as two indicator variables, one indicating whether the individual had lived at the current address for less than a year and one indicating whether they had lived at the same address for more than five years. Income was included as three dummy variables, one for those earning between $20,000 and $50,000 per year, one for those earning between $50,000 and $100,000, and one for those earning over $100,000. There was almost no missing data on the income question for Internet respondents (just two respondents did not provide this information), and fewer than 6% of respondents to the mail and phone modes failed to answer that question. When we re-estimated the models without the income question, our conclusions about minimal mode differences were unchanged.

[8]The Appendix includes a table listing the coefficients and standard errors for each of these models.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

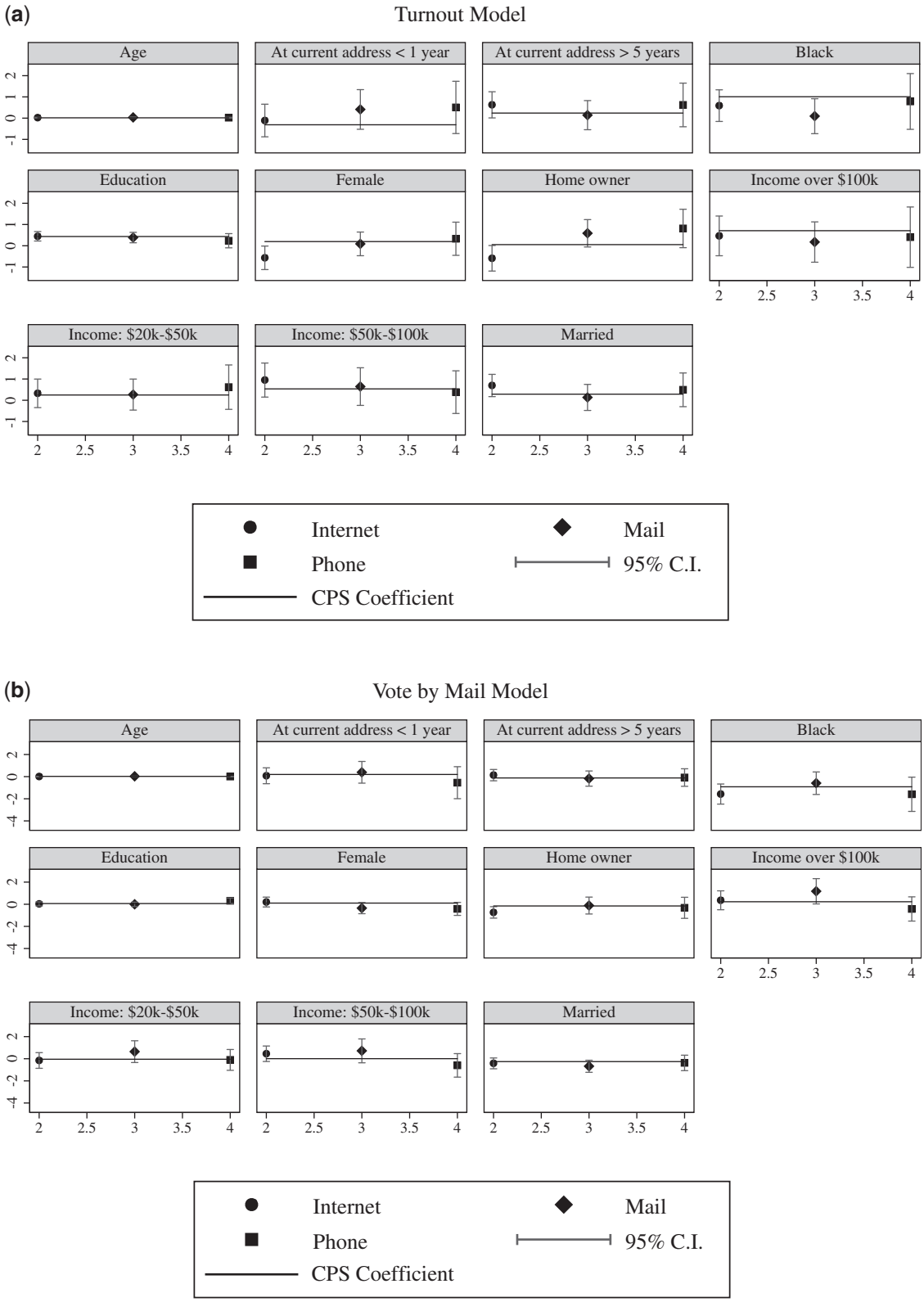


Fig. 1 Comparison of coefficients generated by each mode with cps-validated coefficients. (**a**) Turnout model. (**b**) Vote by mail model. (**c**) Vote early in-person model.

(continued)

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

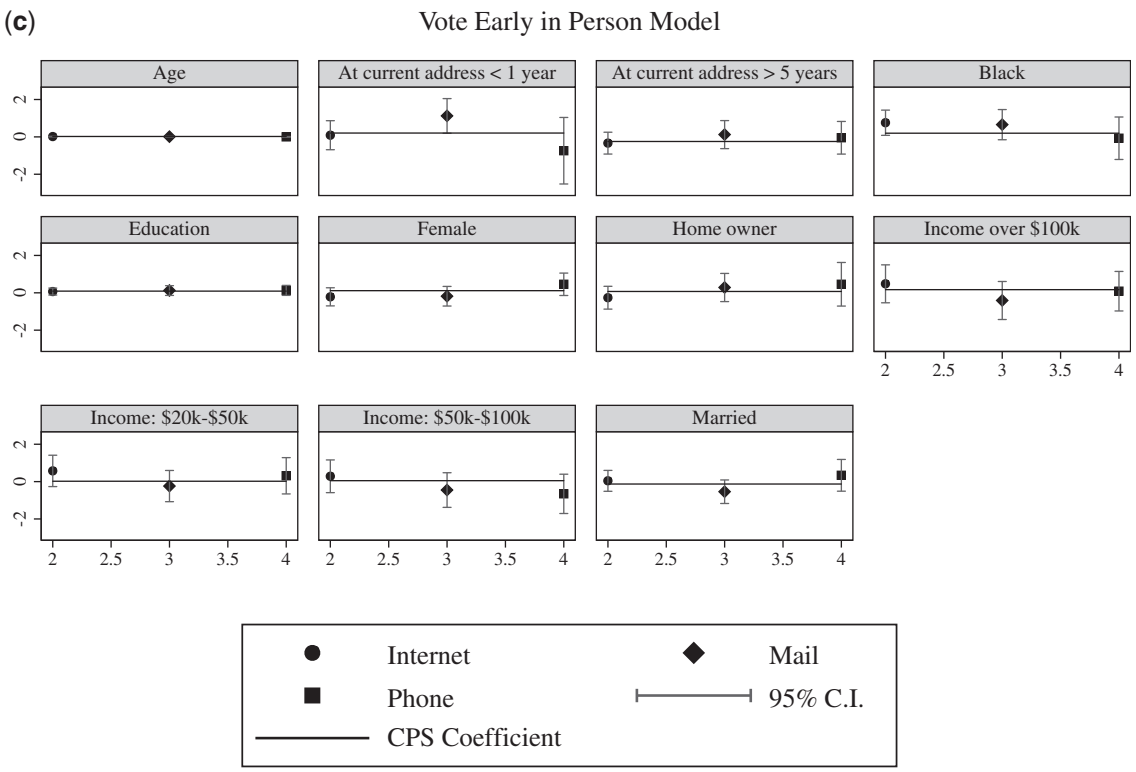


Fig. 1 Continued.

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.

Additionally, since we have coefficients that we can treat as valid measures of the true coefficients for our turnout model, we can take a TSE approach to comparing the correlational structure of the three modes. Specifically, we take the average of the squared difference between the coefficients produced by each mode and the baseline coefficients produced by the model using the CPS data. For the turnout model, the Internet mode generated an MSE for its coefficients of 0.161. By comparison, the MSE for the coefficients for the phone mode was 0.162 and for the mail mode the MSE was 0.183. Thus, the three modes fared quite similarly in terms of the correlational structure estimated for the turnout model. In the other two models estimating factors affecting how individuals voted, the Internet mode fared better than the other two modes. For the model estimating whether voters cast their ballots by mail, the Internet mode produced an MSE of 0.101, compared to 0.201 for the phone survey and 0.216 for the mail mode. Finally, in the model estimating whether individuals voted early in person, the Internet survey produced an MSE of 0.094, the phone survey an MSE of 0.199, and the mail survey had an MSE of 0.194.

4 Comparison of Non-Validated Political Measures

To this point in our analysis, we have focused on measures for which we have validated baselines. This allows us to construct a relatively straightforward TSE comparison of the different modes. However, each survey also collected a substantial amount of data on questions regarding opinions and attitudes for which we lack validated baseline values. In analyzing the surveys on these measures, we take what Smith refers to as a total survey measurement variation approach (2011). This approach involves estimating the extent to which the surveys produce similar (or

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

Table 3 Comparison of point estimates on attitudinal and opinion questions across modes

	Proportions			*Difference in proportions*		
Variable	*Internet*	*Phone*	*Mail*	*Internet-phone*	*Internet-mail*	*Phone-mail*
Approve of Obama	0.473	0.522	0.555	−0.049	−0.082*	−0.033
	(0.020)	(0.026)	(0.021)			
Approve of Congress	0.192	0.283	0.319	−0.091*	−0.128*	−0.037
	(0.017)	(0.025)	(0.022)			
Right track	0.327	0.328	0.314	−0.001	0.014	0.015
	(0.021)	(0.024)	(0.021)			
Support aff. action	0.400	0.519	0.453	−0.119*	−0.053	0.066*
	(0.021)	(0.026)	(0.021)			
Abortion always legal	0.167	0.167	0.180	0.001	−0.014	−0.014
	(0.014)	(0.020)	(0.016)			
Favor gay marriage	0.405	0.379	0.412	0.026	−0.006	−0.032
	(0.021)	(0.028)	(0.022)			
Favor SS privatization	0.549	0.573	0.540	−0.024	0.009	0.032
	(0.023)	(0.025)	(0.024)			
Favor tax on >$200 k	0.617	0.596	0.700	0.021	−0.084*	−0.105*
	(0.021)	(0.025)	(0.020)			
Cut defense	0.221	0.149	0.222	0.073*	0.001	−0.073*
	(0.017)	(0.019)	(0.017)			
Cut domestic	0.299	0.238	0.228	0.061*	0.072*	0.011
	(0.017)	(0.021)	(0.016)			
Govt. wasteful & inefficient	0.688	0.563	0.620	0.126*	0.068*	−0.057
	(0.019)	(0.026)	(0.021)			
Abs. standards right and wrong	0.504	0.455	0.491	0.049	0.013	−0.036
	(0.020)	(0.026)	(0.021)			
Average absolute difference				0.053	0.045	0.043

Note. Entries are weighted proportions of respondents in each category after excluding those responding "don't know" or "not sure." *$p < 0.05$ *F*-test. Standard errors in parentheses.

dissimilar) measures of each concept, but differs from the TSE approach because we are not able to judge which value is closest to the true parameter.

Table 3 presents point estimates for a variety of measures of political attitudes and opinions that cannot be compared to any validated baseline. 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 seven 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 thirteen percentage points. The phone survey estimated significantly more support (approximately twelve 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% of Americans approved of the job Congress was doing, whereas the Internet survey placed the estimate at just 19%.

The mail and Internet surveys differed significantly on five of the measures presented in Table 3. The mail survey estimated higher levels of approval for Obama (about eight percentage points higher) and Congress (twelve points) as well as more support for increasing taxes on those earning more than $200,000 per year (eight points). The mail survey also registered less support for cutting

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

more from domestic spending (seven percentage points less) and less agreement with the statement that the government is almost always wasteful and inefficient (seven points). The telephone and mail surveys differed on three of the measures presented in the figure. The mail survey estimated ten 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 (seven points). The phone survey estimated about seven percentage points more support for affirmative action programs than the mail survey.

It is important to emphasize that although we do find some differences across the opinion and attitude measures presented in Table 3, 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 0.73). 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 −0.07), followed by the phone (−0.02) and then Internet modes (0.01). 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.

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

In addition to the questions analyzed so far, the survey also asked about basic facts about politics and ascertained different forms of nonvoting political engagement, including media viewership and reported contributions. Respondents were asked 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 what the unemployment rate was at the time, which party controlled the House of Representatives, and the party affiliation of their state's governor.[9] Figure 2 compares responses to these questions across mode. It is on these measures that we see some substantial differences across modes, but the differences are no more than 10%.

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. On all three knowledge items, 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 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 s to answer each of the knowledge questions before

[9]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%).

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

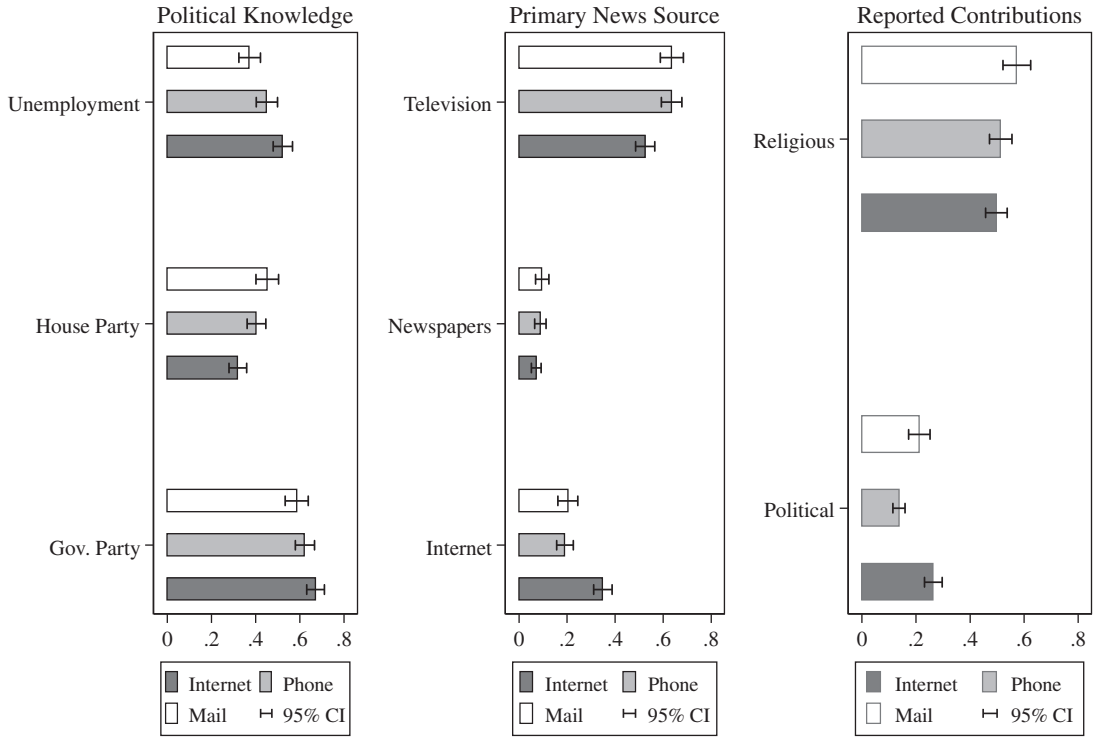


Fig. 2 Comparison of point estimates on knowledge, news source, and reported contributions. Entries are weighted proportions of respondents in each category after excluding those responding "don't know" or "not sure," except for the knowledge questions, where "don't know" or "not sure" was coded as incorrect.

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. In addition, respondents to the mail survey had even greater opportunities to search for answers to these questions because they did not face time constraints. But mail survey respondents did not show a higher propensity to get the questions right.

5.1 *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, whereas 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. Whereas 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. Although some studies have found an association between Internet use and political knowledge, the effects are often small (see Kenski and Stroud 2006). In Fig. 3, we use the telephone and mail respondents to compare the political knowledge of those with and without Internet access in their homes. The estimates

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

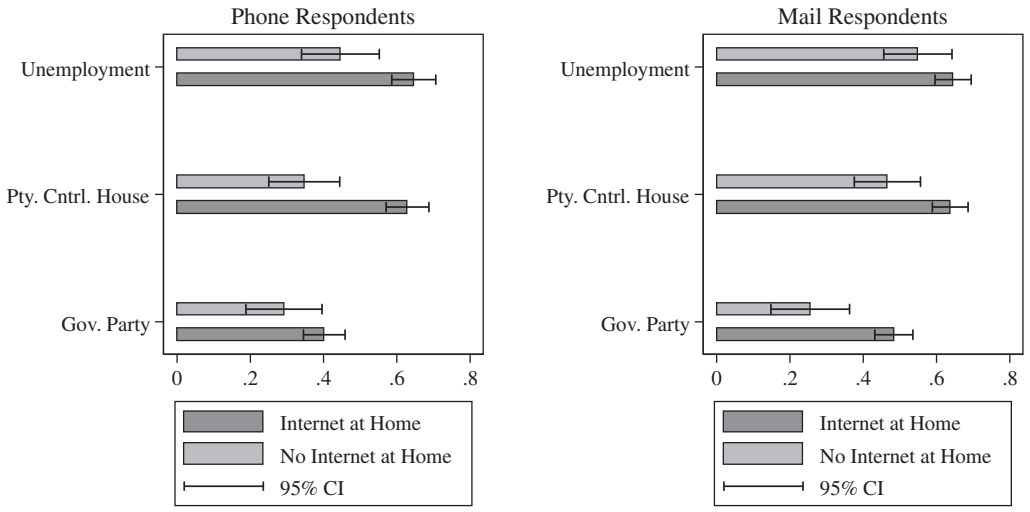


Fig. 3 Comparison of point estimates on knowledge between respondents with and without Internet access at home. Entries are weighted proportions of respondents answering each question correctly; "don't know" or "not sure" were coded as incorrect.

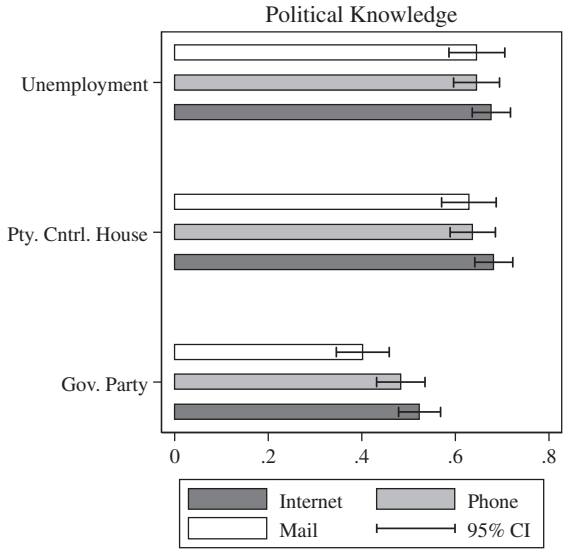


Fig. 4 Comparison of point estimates on knowledge among respondents with Internet access at home. Entries are weighted proportions of respondents answering each question correctly; "don't know" or "not sure" were coded as incorrect.

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 Fig. 4) repeats the knowledge comparisons presented in Fig. 2, but restricts the analysis to only those respondents who report that they have Internet access

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

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; the immediate lesson, though, is that the opt-in nature of the Internet survey does not seem to be a major source of the mode difference in knowledge.

6 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 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 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 were relatively small and failed to attain statistical significance, indicating that we could not be confident that the coefficients differed across the modes. There were, however, two *F*-statistics that were significant at $p < 0.05$ and one that was significant at $p < 0.01$.

Table 4 OLS estimates of factors affecting Obama approval ratings

	Coefficients (SEs)			*F-test statistics*		
Variable	*Internet*	*Phone*	*Mail*	*Internet versus phone*	*Internet versus mail*	*Phone versus mail*
Ideology	−0.210***	−0.186***	−0.167**	0.10	0.26	0.06
	(0.058)	(0.048)	(0.061)			
Right track	0.864***	1.078***	0.833***	1.80	0.04	2.47
	(0.115)	(0.109)	(0.111)			
Economy	−0.249***	−0.283***	−0.302***	0.15	0.39	0.04
	(0.059)	(0.066)	(0.061)			
Democrat	0.518***	0.409***	0.496***	0.42	0.02	0.30
	(0.129)	(0.108)	(0.114)			
Republican	−0.181**	−0.136	−0.385**	0.13	2.27	2.45
	(0.068)	(0.108)	(0.117)			
Age	−0.002	−0.008**	0.002	3.43	1.84	7.33**
	(0.002)	(0.002)	(0.003)			
Female	−0.070	0.143	0.220**	3.35	6.64*	0.40
	(0.075)	(0.089)	(0.084)			
White	−0.140	−0.073	−0.167	0.20	0.03	0.30
	(0.085)	(0.125)	(0.120)			
Education	−0.006	−0.053	0.010	1.58	0.17	2.01
	(0.021)	(0.031)	(0.031)			
Income	−0.041***	−0.030*	−0.007	0.36	3.97*	1.52
	(0.011)	(0.014)	(0.013)			
Born again	−0.025	−0.073	−0.123	0.02	0.73	0.86
	(0.078)	(0.093)	(0.083)			
Intercept	3.594***	3.943***	3.197***			
	(0.269)	(0.243)	(0.322)			
N	676	566	687			
R^2	0.668	0.622	0.617			

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$. Standard errors in parentheses.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

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, whereas 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, whereas 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 4 is that the correlational structure of the data did not differ substantially across the modes. In fact, of thirty-three *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 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 $200,000. Table 5 includes *F*-statistics for tests determining whether the regression coefficients estimated from one mode are jointly different from those estimated by another mode. For example, the first row of *F*-statistics indicates that in the models for Obama's approval rating, the set of coefficients estimated for phone respondents was statistically different from those estimated for mail respondents ($F = 1.82$, $p < 0.05$). 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 from the Internet survey were never statistically distinguishable from those estimated from the phone survey. Each case of a statistically significant joint *F*-statistic involves the mail survey, in comparison to either the Internet or phone mode. It is also important to note that although the *F*-statistics are significant in four cases in Table 4, in none of these instances is the *F*-statistic very large. Thus, although we can be confident that there are four instances where the correlational structure of the data differs, the size of this difference is not necessarily large.

As a final test of the correlational structure of the data, we returned to our measures of political knowledge. Since these generated different point estimates across modes, it is important to determine whether there are also mode differences in correlations involving knowledge questions. In this test, the dependent variable is the number of factual questions that each respondent answered correctly. Thus, this variable ranges from 0 to 3. We regressed on this dependent

Table 5 Results from *F*-tests comparing models across modes

Dependent variable	*Internet versus phone*	*Internet versus mail*	*Phone versus mail*
Obama approval	1.08	1.62	1.82*
Congressional approval	1.03	0.50	0.40
2008 presidential vote	1.66	1.13	2.23*
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 >$200 k support	1.38	2.36**	1.72
Cut domestic or defense spending more	1.65	1.12	0.94
Is government wasteful	0.59	1.63	2.38**

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

Table 6 OLS estimates of factors affecting political knowledge

	Coefficients (SEs)			*F-test statistics*		
Variable	*Internet*	*Phone*	*Mail*	*Internet versus phone*	*Internet versus mail*	*Phone versus mail*
Liberal	0.146	0.198	0.176	0.08	0.04	0.01
	(0.099)	(0.160)	(0.114)			
Conservative	0.054	0.241*	0.273*	1.35	1.62	0.03
	(0.105)	(0.122)	(0.136)			
Democrat	0.117	0.107	0.186	0.00	0.21	0.19
	(0.094)	(0.139)	(0.116)			
Republican	0.243*	0.101	−0.111	0.70	4.07*	1.28
	(0.111)	(0.129)	(0.136)			
Age	0.006**	0.017***	0.013***	8.94**	3.76	0.85
	(0.002)	(0.003)	(0.003)			
Female	−0.505***	−0.357***	−0.224*	1.27	5.14*	0.84
	(0.076)	(0.107)	(0.098)			
White	0.004	0.235	0.243*	1.96	2.57	0.00
	(0.086)	(0.141)	(0.122)			
Education	0.091***	0.174***	0.186***	2.78	4.55*	0.05
	(0.026)	(0.042)	(0.036)			
Income	−0.051***	0.051***	0.045***	0.00	0.09	0.08
	(0.011)	(0.016)	(0.014)			
Intercept	1.291***	−0.263	0.053***			
	(0.164)	(0.239)	(0.267)			
N	632	538	614			
R^2	0.221	0.331	0.241			

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$. Standard errors in parentheses.

variable a set of covariates that political scientists have found to be correlated with knowledge (Delli Carpini and Keeter 1996). These variables include age, gender, race, education, income, identification with one of the two major parties, and identification as a liberal or conservative.

The results are presented in Table 6. On the right side of the table are the tests for equality in the coefficients across the modes. When comparing the coefficients from the Internet to the phone mode, there is one statistically significant difference in coefficients—the coefficient for age is nearly three times larger in the phone mode than it is for the Internet survey. There are several significant differences in coefficients when comparing the Internet and mail surveys. The coefficient for education was significantly smaller in the Internet survey compared to the mail mode, whereas the Internet mode produced larger effects for the gender variable and the dummy variable identifying Republican respondents. Notably, there were no significant differences in the coefficients produced by the mail survey and the phone mode.

7 Discussion

Our goal in comparing modes is not to crown one approach as the "winner" over others, but to provide guidance to researchers about the costs and benefits of each survey approach. Indeed, in this study there is not a clear winner, because the modes produce quite similar results. In our TSE 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 MSE, whereas the web and mail surveys produced MSEs that were quite similar. Our comparisons of the correlational structures of the data also uncovered few cross-mode differences. And although our other analyses could not be tied to validated baselines, we typically found negligible differences across modes here as well. Specifically, we found no strong tendency for any particular mode to produce consistently more liberal or conservative responses to questions about policy issues.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

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 calculation may be a bit closer when comparing telephone and Internet surveys, but given that its error rate was about what it should have been given the sample size, the Internet survey would likely still be preferred by many researchers. Not only was the Internet survey half as expensive as the telephone poll, but it also took significantly less time to administer. Indeed, Table 1 shows that the median completion time for a telephone interview was about 60% longer than the median for the Internet mode. 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., Malhotra and Krosnick 2007; Chang and Krosnick 2009; Yeager et al. 2009; Pasek and Krosnick 2010) but is consistent with other work (Chatt et al. 2003; Rivers 2007; Ansolabehere 2009). We believe there are two main reasons for why some scholars find larger mode differences than we have uncovered here. First, many studies 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 look remarkably like those 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 as easily as a poorly designed telephone, mail, or in-person poll.

Thus, we caution researchers to take care in differentiating between different approaches to producing Internet surveys. When judging the quality of an opt-in Internet survey, we think several related factors are particularly important. First, the use of sample matching based on a target sample appears to be crucial (Rivers 2007). The use of a target sample not only provides a baseline for building a representative sample from the pool of Internet panelists, but it also allows one to use defensibly statistical tests that rely on assumptions about probability sampling (since these assumptions are met by the target sample). Second, attention should be paid to the types of variables used for matching the target sample. When these variables are more likely to address the biases known to exist among opt-in panelists, the matched sample should prove to be more representative of the target sample (and, therefore, of the target population). Thus, for the Internet sample we commissioned, the use of variables such as age, ideology, and news interest in the matching algorithm helps account for the fact that Internet panelists are likely to be younger and more interest in politics. Third, it also appears to be crucial that the set of potential panelists be sufficiently large and diverse to allow for reasonably close matches to be achieved (Rivers 2007). Fourth, propensity score weights should be calculated to account for the remaining unrepresentativeness of the matched sample. Our findings indicate that when these conditions are met, an opt-in Internet survey 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.

Thus, the real lesson is that the criteria we require of good opt-in samples are similar to those we would want for all samples. Even face-to-face probability surveys, such as the American National Election Survey, now report post-stratification weights instead of true probability sample weights. The shift to post-stratification weighting, now the norm among survey researchers, means that, regardless of mode, the results derived from samples are highly contingent on the quality of the weighting scheme developed by the survey firm.

Funding

Funding for this study came from the Dean of the Faculty of Arts and Sciences, Harvard University.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press

Appendix

Table A1 Comparison of CPS coefficients with estimates from surveys

	DV: Voted				*DV: Voted by mail*				*DV: Voted early in-person*			
Variable	*CPS*	*Internet*	*Phone*	*Mail*	*CPS*	*Internet*	*Phone*	*Mail*	*CPS*	*Internet*	*Phone*	*Mail*
Education	0.43	0.45	0.24	0.39	0.06	0.03	0.31	−0.01	0.09	0.07	0.13	0.12
	(0.01)	(0.11)	(0.17)	(0.13)	(0.01)	(0.09)	(0.13)	(0.1)	(0.01)	(0.09)	(0.14)	(0.13)
Black	1.01	0.59	0.79	0.09	−0.92	−1.57	−1.59	−0.59	0.2	0.76	−0.07	0.65
	(0.06)	(0.38)	(0.67)	(0.42)	(0.06)	(0.46)	(0.79)	(0.52)	(0.04)	(0.34)	(0.58)	(0.41)
Age	0.02	0.02	0.02	0.03	0.02	0.01	0.01	0.03	0.02	0.01	0	0.01
	(0)	(0.01)	(0.01)	(0.01)	(0)	(0.01)	(0.01)	(0.01)	(0)	(0.01)	(0.01)	(0.01)
Female	0.2	−0.57	0.33	0.09	0.11	0.2	−0.42	−0.35	0.11	−0.22	0.46	−0.18
	(0.03)	(0.28)	(0.4)	(0.28)	(0.03)	(0.23)	(0.3)	(0.25)	(0.02)	(0.25)	(0.31)	(0.27)
Home owner	0.05	−0.59	0.81	0.59	−0.15	−0.74	−0.32	−0.12	0.07	−0.26	0.46	0.28
	(0.04)	(0.31)	(0.46)	(0.33)	(0.04)	(0.26)	(0.48)	(0.39)	(0.03)	(0.31)	(0.59)	(0.38)
Married	0.29	0.69	0.49	0.13	−0.26	−0.42	−0.38	−0.68	−0.13	0.04	0.34	−0.54
	(0.04)	(0.27)	(0.4)	(0.31)	(0.03)	(0.25)	(0.36)	(0.28)	(0.03)	(0.28)	(0.43)	(0.32)
Income: 20–50 k	0.25	0.33	0.62	0.27	−0.04	−0.16	−0.1	0.64	0.02	0.57	0.31	−0.24
	(0.05)	(0.34)	(0.53)	(0.37)	(0.05)	(0.36)	(0.48)	(0.5)	(0.04)	(0.43)	(0.49)	(0.43)
Income: 50–100 k	0.54	0.95	0.38	0.64	0.01	0.45	−0.6	0.71	0.05	0.28	−0.66	−0.45
	(0.05)	(0.41)	(0.51)	(0.45)	(0.05)	(0.36)	(0.54)	(0.55)	(0.04)	(0.45)	(0.54)	(0.47)
Income: Over 100 k	0.71	0.46	0.4	0.17	0.23	0.36	−0.43	1.17	0.17	0.48	0.09	−0.41
	(0.07)	(0.48)	(0.72)	(0.48)	(0.06)	(0.44)	(0.56)	(0.58)	(0.05)	(0.52)	(0.54)	(0.52)
At address <1 year	−0.31	−0.11	0.5	0.41	0.2	0.08	−0.55	0.39	0.2	0.09	−0.74	1.12
	(0.05)	(0.39)	(0.63)	(0.48)	(0.05)	(0.37)	(0.74)	(0.5)	(0.04)	(0.4)	(0.91)	(0.47)
At address >5 years	0.24	0.63	0.62	0.14	−0.12	0.14	−0.08	−0.18	−0.25	−0.34	−0.05	0.12
	(0.04)	(0.31)	(0.52)	(0.35)	(0.03)	(0.26)	(0.4)	(0.35)	(0.03)	(0.3)	(0.44)	(0.38)
Intercept	−0.78	−0.78	−1.41	−2.08	−2.66	−1.58	−2.34	−2.98	−2.19	−2.85	−3.37	−2.97
	(0.08)	(0.62)	(0.88)	(0.79)	(0.08)	(0.67)	(1.09)	(0.72)	(0.07)	(0.61)	(0.95)	(0.77)
N	57180	768	702	907	51085	683	648	792	51085	683	648	792
MSE of coefficients		0.161	0.162	0.183		0.101	0.201	0.215		0.094	0.190	0.194

Notes. Standard errors in parentheses. MSE of coefficients is the mean squared difference between the coefficients estimated in each model and the baseline coefficients produced by the models estimated with CPS data.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

References

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

Ansolabehere, Stephen. 2009. *Guide to the 2008 Cooperative Congressional Election Survey*. http://projects.iq.harvard.edu/cces/data?dvn_subpage=/faces/study/StudyPage.xhtml?globalId=hdl:1902.1/14003.

Ansolabehere, Stephen, and Brian F. Schaffner. 2013. *Replication Data for: Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison*. http://dx.doi.org/10.7910/DVN/23025 UNF:5:lIok9HT/VfayP4ULEfvB4A== IQSS Dataverse Network [Distributor] V1 [Version].

Atkeson, Lonna Rae, Lisa A. Bryant, Alex N. Adams, Luciana Zilberman, and Kyle L. Saunders. 2011. Considering mixed mode surveys for questions in political behavior: Using the Internet and mail to get quality data at reasonable costs. *Political Behavior* 33:161–78.

Atkeson, Lonna Rae, and Lori Tafoya. 2008. Surveying political activists: An examination of the effectiveness of a mixed mode (Internet and mail) survey design. *Journal of Elections, Public Opinion, and Parties* 18(4):367–86.

Biemer, Paul P. 2010. Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly* 74(5):817–48.

Blumberg, Stephen J., and Julian V. Luke. 2009. *Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, January–June 2011*. Centers for Disease Control. http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201112.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* 73(4):641–78.

———. 2010. Comparing oral interviewing with self-administered computerized questionnaires: An experiment. *Public Opinion Quarterly* 74(1):154–67.

Chatt, Cindy, J. Michael Dennis, Rick Li, and Paul Pulliam. 2003. Data collection mode effects controlling for sample origins in a panel survey: Telephone versus Internet. Paper presented at the Annual Meeting of the Midwest Chapter of the American Association for Public Opinion Research, Chicago.

Curtin, R., S. Presser, and E. Singer. 2005. Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly* 69:87–98.

Delli Carpini, Michael X., and Scott Keeter. 1997. *What Americans know about politics and why it matters*. New Haven, CT: Yale University Press.

Duncan, W. Jack. 1979. Mail questionnaires in survey research: A review of response inducement techniques. *Journal of Management* 5:39–55.

Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, and Eleanor Singer. 2009. *Survey Methodology*. 2nd ed. Hoboken, NJ: J. Wiley.

Groves, Robert M., and Lars Lyberg. 2010. Total survey error: Past, present, and future. *Public Opinion Quarterly* 74(5):849–79.

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.

Kenski, Kate, and Natalie Jomini Stroud. 2006. Connections between Internet use and political efficacy, knowledge, and participation. *Journal of Broadcasting & Electronic Media* 50(2):173–92.

Link, Michael W., and Ali H. Mokdad. 2005. Alternative models for health surveillance surveys: An experiment with web, mail, and telephone. *Epidemiology* 16(5):701–4.

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 US Census Bureau (#2010-15).

Rivers, Douglas. 2006. *Sample matching: Representative sampling from Internet panels*. http://www.websm.org/db/12/16408/Web%20Survey%20Bibliography/Sample_matching_Representative_sampling_from_Internet_panels/.

———. 2007. Sampling for web surveys. Paper prepared for the 2007 Join Statistical Meetings, Salt Lake City, UT.

———. 2009. *Second thoughts about Internet surveys*. Pollster.com (September 6). http://www.pollster.com/blogs/doug_rivers.php.

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* 15:257–85.

Smith, Tom W. 2011. Refining the total survey error perspective. *International Journal of Public Opinion Research* 23(4):464–84.

Yeager, David S., Jon A. Krosnick, LinChiat Chang, Harold S. Javitz, Matthew S. Levendusky, Alberto Simpser, and Rui Wang. 2009. Comparing the accuracy of RDD telephone surveys and Internet surveys conducted with probability and non-probability samples. Unpublished manuscript.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt025 Published online by Cambridge University Press