

Survey Attention and Self-Reported Political Behavior

R. Michael Alvarez and Yimeng Li*

California Institute of Technology

August 5, 2022

Abstract

Survey research methodology is evolving rapidly, as new technologies provide new opportunities. One of the areas of innovation regards the development of online interview best practices and the advancement of methods that allow researchers to measure the attention that subjects are devoting to the survey task. Reliable measurement of subject attention can yield important information about the quality of the survey response. In this paper, we take advantage of an innovative survey we conducted in 2018, in which we directly connect survey responses to administrative data, allowing us to assess the association between survey attention and response quality. We show that attentive survey subjects are more likely to provide accurate survey responses regarding a number of behaviors and attributes that we can validate with our administrative data. We discuss the best strategy to deal with inattentive respondents in surveys in light of our results.

*We thank the John Randolph Haynes and Dora Haynes Foundation for supporting our research, and the Orange County Registrar of Voters for providing access to administrative data. We thank Jian Cao and Ines Levin for their collaborations with us on related projects. This paper was presented at the 2021 Midwest Political Science Association annual meeting and in the UCLA Biostatistics Seminar Series; we thank participants for their comments and feedback about our research.

1. Introduction

Survey research has been in the midst of vast changes in recent years, as new technologies provide new opportunities. One of the most important changes has been the increase in self-completion polls and surveys. One possible consequence of the increasing use of self-completion surveys is that the absence of an interviewer may mean that respondents could move through a questionnaire quickly and not pay close attention to the questions or the potential responses (Curran, 2016). As answering a survey takes time and effort, it's possible that self-completion surveys might be particularly susceptible to satisficing behavior by subjects (Simon, 1956; Krosnick, 1991).

Inattentive and satisficing respondents in self-completion surveys may thus provide lower-quality data, as they might randomly answer questions, provide inaccurate responses, answer with a “don't know”, or use other tactics to get through a survey quickly (Atkeson et al., 2014; Clifford and Jerit, 2015). These concerns have led researchers to study the use of “attention checks”, “trap questions”, or “screeners”, usually in the form of instructed response items or instructional manipulation checks (in this paper, we use the term attention checks).

Recent studies have examined the use of attention checks in surveys and opinion polls, documenting how attentive and inattentive survey respondents differ, studying different types of attention checks and methods for measuring attentiveness, and examining how to deal with inattentive respondents in survey data (Read et al., 2021; Berinsky et al., 2021; Kung et al., 2018; Hauser and Schwarz, 2015; Berinsky et al., 2014; Oppenheimer et al., 2009). Other studies have looked at experimental subject attentiveness, in either convenience samples of students or from crowd-sourcing applications like MTurk (Ahler et al., 2021; Curran and

Hauser, 2019; Curran, 2016; Thomas and Clifford, 2017). What these studies lack is ground truth information, and that has prevented researchers from quantifying the performance of attention checks and comparing different approaches to dealing with inattentive respondents. The research we present here fills this gap by examining responses to factual survey questions that we can validate with external administrative data.

Our results indicate that respondents failing attention checks are more likely to misreport various factual information. Many of these inattentive respondents nonetheless provide responses in line with the information in the administrative records. For turnout histories in recent elections, which correlate with respondent attention, dropping inattentive respondents leads to an unrepresentative subsample and, hence, estimates with larger biases and variances. By contrast, for modes of voting in recent elections, which are largely uncorrelated with attention check passages, dropping inattentive respondents yields estimates with smaller biases that often outweigh the cost of larger variances.

2. Past Research and Hypotheses

Inattentive respondents may diminish the quality of survey data. For example, one recent study found that inattentive respondents offer quick answers, are more likely to give “don’t know” answers, and are more likely to “straightline” (Alvarez et al., 2019). One solution to the problems caused by inattentive respondents is to use attention checks, also known as “screeners” or “trap questions”. Past research has differentiated attention checks into two types. One type of attention check is an instructional manipulation check (IMC), where there is a deliberate change in the instructions in a survey question designed to capture

whether the respondent is reading and cognitively processing the question's instructions (Oppenheimer et al., 2009; Berinsky et al., 2014; Anduiza and Galais, 2017). An example of an IMC is adding a clause to a survey question instructing the respondent to ignore the question and provide a specific answer. The other type of attention check is an instructed response item (IRI), where the responses to a survey question are altered in a way that should elicit whether the respondent is attentive to the question's response options (Alvarez et al., 2019). An example of an IRI is adding a row in a grid instructing respondents to select 'strongly disagree' for survey quality control.

However, due to the lack of ground truth information, previous studies have relied on post-hoc approaches to evaluate the performance of attention checks.¹ These measures look at whether respondents passing and failing attention checks differ in producing canonical experimental results, empirical correlations between negatively correlated survey questions, straight-lining behavior, and response time. In our dataset, we possess ground truth information, which we argue can be used to better measure whether inattentive respondents provide less accurate answers — and which can help us to better evaluate different approaches for dealing with inattentive respondents.

Based on past research using post-hoc approaches, we hypothesize that inattentive survey respondents will provide less accurate answers to questions about their recent and past voting behavior, in particular whether they voted in the current and past elections, and the method that they used to vote in the current and past elections (Hypothesis 1). Because inattentive respondents provide less accurate answers, that generates error and noise in a survey dataset.

¹In a recent related study, Kuhn and Vivyan (2021) compared the accuracy of self-reported behavior using list experiments and direct questions and comparing the responses to administrative data.

Thus, when the data is used for modeling or testing hypotheses, including the answers from inattentive respondents will potentially generate bias in model estimates and incorrect inferences. Thus it is likely that inattentive respondents will generate systematic errors in a survey dataset, not just additional noise or variance. This leads to Hypothesis 2: inattentive respondents will provide answers that introduce bias in model estimates.

We find consistent support for Hypothesis 1, for self-reports of turnout for a number of recent elections, and for self-reports of the method that the voter used to return or cast their ballot. We also show support for Hypothesis 2, that there is evidence that inattentive respondents are generating systematic error in survey data (not random error).

These results have important corollaries. Since respondent inattention will generate error in survey self-reports of political behavior, under the assumption that respondent attention is uncorrelated with the outcome of interest, dropping inattentives from the analysis will increase variance because of information loss, but will decrease bias in the estimates. However, under the assumption that respondent attention is correlated with the outcome of interest, if a researcher drops inattentive respondents that can produce additional bias by creating an unrepresentative sample (though the direction of the bias is not clear, and will depend on the type of heuristic that inattentive respondents use in answering the survey). We find support for these corollaries, which have important implications for dealing with inattentive respondents in research using self-administered surveys.

Testing these hypotheses with ground-truth data is the gap in the literature that our research seeks to fill. In 2018 we fielded a unique survey that allowed us

to directly connect survey responses to voter registration data. In our survey, we posed a series of questions, regarding current and past electoral participation as well as other demographic information, that we could validate with the administrative data. The combination of self-reported political behavior, validated information from the administrative records, and measures of respondents' attention provides us with a larger array of different types of ground-truth information for both inattentive and attentive survey respondents, so that we can study the accuracy of self-reports by respondents' attentiveness.

In the next section of our paper, we discuss our survey design. Then we turn to various tests of our two hypotheses using data from our unique survey. We first examine how inattentive respondents answer questions about their voting participation and method of voting. We then examine our second hypothesis, and test whether inattentive respondents are generating answers that introduce bias into models using their survey responses. After comparing keeping and dropping inattentive survey respondents, we conclude by discussing the best strategy to deal with inattentive respondents in surveys in light of our results.

3. Survey and Attention Checks

Our study uses voter registration and voting history administrative data provided by the Orange County Registrar of Voters (OCROV). These datasets contain information about each registered voter in the county, including contact and demographic information, partisan registration, whether they turned out to vote in past elections, and if they did turn out in a past election whether they voted in person or by mail. Importantly, in California, voters can provide contact information (like a telephone number or email address) when they register to

vote. Of the approximately 1.6 million voter registration records in the county in the 2018 general election, over 530,000 of those records were associated with an email address. We used all of the records with email addresses for our survey.

For our research on the 2018 general election in Orange County, we developed a self-completion online survey focused on the voter's election experiences. The online questionnaire included attention checks in the form of IMCs and IRIs, questions on voter registration, turnout in recent elections, and method of voting, in addition to questions on other subjects. We invited registered voters (via email) to participate in our survey between Thursday, November 8, 2018, and Tuesday, November 13, 2018. From 531,777 invites to all registered voters with email addresses, we received 6,952 complete responses. The survey took about 12 to 15 minutes to complete (the median duration was 13 minutes) and was provided in English. More details about the survey can be found in the paper's Supplementary Materials.

We then match back each survey response to the corresponding registered voter. In most cases (6,816), the survey respondent can be linked back to a unique registered voter by the email address alone.² In cases of ambiguity, we further match responses to voters according to age (or, in rare cases, a combination of age and gender) from self-reports and administrative records. By matching survey respondents directly to administrative data, we can validate self-reports of voter turnout and method of voting, and add features (such as party registration) from the administrative data to our analysis. More details about the

²For each email address in the voter registration file, we generated a unique survey URL. As such, for 6,816 out of 6,952 cases where a unique record is associated with the email address, the email address is a unique identifier that unambiguously links survey responses to records in the voter registration file. The vast majority of the remaining cases concern different household members sharing the email address. Therefore, we do not face the methodological challenges addressed in Enamorado et al. (2019).

administrative data are discussed in the paper's Supplementary Materials.

Our 2018 survey contained both Instructional Manipulation Checks (IMC) and Instructed Response Items (IRI) for assessing survey attention.³ We designed the survey instrument so that respondents would receive one attention check approximately 25% of the way through the survey (attention check 1), the second when they had completed about 50% of their survey (attention check 2), and the final attention check at about 75% completed (attention check 3). In all cases, subjects who ignored the attention checks were allowed to continue to the next set of survey questions.

To avoid potential question order biases, we randomized the appearance of the attention check questions in our survey. For attention check 1, respondents were randomly assigned to receive either an IRI or IMC attention check. The IRI check asked subjects to answer “oppose” or “support” among five other questions about their opinions regarding election reform; the IMC attention check asked subjects to ignore a question on news media websites and to select two specific news organizations as their answers.

Attention check 2 followed a series of questions about the subject's voting experiences. Those who had been asked to answer the IRI for attention check 1 were then asked the IMC for attention check 2. Similarly, subjects who answered the IMC for attention check 1 were asked the IRI attention check. Again, we designed this to avoid any question order or location effects with respect to the use of the IMC or IRI format as attention checks.

Finally, the third attention check was located about 75% of the way through the survey questionnaire. Here the subjects were again randomized, with half

³We used an IMC studied in Berinsky et al. (2014) and an IRI studied in Alvarez et al. (2019) as our main attention checks.

Table 1: IMCs Screen Respondents More Aggressively Than IRIs

| | Fail | Pass | Skip |
|------------------------------|------|------|------|
| IRI - Attention check 1 or 2 | 15% | 82% | 4% |
| IRI - Attention check 1 | 16% | 81% | 4% |
| IRI - Attention check 2 | 13% | 83% | 4% |
| Additional IRI | 9% | 88% | 4% |
| IMC - Attention check 1 or 2 | 45% | 52% | 3% |
| IMC - Attention check 1 | 49% | 47% | 4% |
| IMC - Attention check 2 | 40% | 57% | 3% |
| Additional IMC | 28% | 70% | 2% |

receiving an additional IRI, while the others received an additional IMC. We included this attention check (an additional IRI or an additional IMC) for the purposes of robustness checks; we present our results from the main IRI and IMC attention checks in the text.⁴

Questions on turnout and mode of voting in the 2018 General Election appear right before attention check 1, while questions on turnout and mode of voting in previous elections appear right after attention check 2. We provide the complete wording of the attention checks and these questions in the paper’s Supplementary Materials.

The overall passage rates of our IRI and IMC attention checks are presented in Table 1. The first four rows of the table provide statistics on failure rates for the IRI questions, while the last four rows show similar statistics on the failure rates

⁴In the third attention check, the IRI asked subjects to answer either “disagree strongly” or “agree strongly” after a battery of questions about political polarization. The final attention check’s IMC involved a question about social media; subjects were asked to ignore the question and select two specific answers to show they were paying attention to the survey questions.

for the IMC questions. Consistent with previous studies on self-administered online surveys (Berinsky et al., 2014; Alvarez et al., 2019), a significant fraction of respondents failed these attention checks. As we expected, IMCs screen respondents more aggressively than IRIs, with 45% of respondents failing the IMC at either attention check 1 or 2 and 15% of respondents failing the IRI at either attention check 1 or 2.

Also, it is important to note two other results in Table 1. First, regardless of whether a respondent receives the IRI or the IMC question at attention check 2 (after they have had one of the attention checks earlier in the survey), we see that the passage rates for both are slightly higher later in the survey. Similarly, for the third attention check, again, passage rates are higher as well for both types of attention check questions. This indicates that as subjects proceed through the survey, the presence of successive attention checks may increase the cognitive focus of survey respondents.⁵

Furthermore, the passage of our IRI and IMC attention checks is correlated with important demographics, as shown in Table SM1 in the Supplementary Materials. Educational attainment is positively correlated with the passage of each type of attention check, consistent with previous studies on self-administered online surveys (Berinsky et al., 2014; Alvarez et al., 2019). Non-white respondents are less likely to pass IRI and IMC attention checks than white respondents, again consistent with previous research. Male and female voters appear equally likely to pass the attention checks. Finally, contrary to previous studies, we find that age is negatively correlated with the passage of each type of attention check.

⁵These results require further and future study, but are beyond the scope of the research we report in this paper.

4. Results

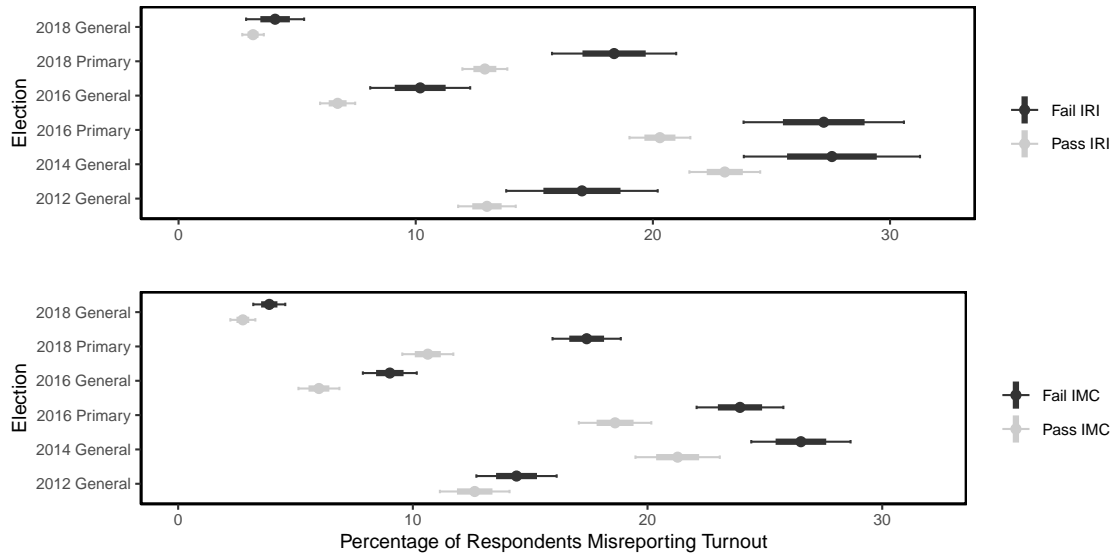
4.1 Accuracy of Self-Reports by Inattentive and Attentive Survey Respondents

We begin by analyzing three pieces of basic respondent information that we are able to validate using the administrative records (presented in Table SM2 and discussed in the Supplementary Materials): year of birth, city of residence, and voter registration before or after January 1, 2017. For each of the three pieces of basic respondent information and each of the two types of attention checks, respondents failing either the IRI or IMC attention check are more likely to misreport the information than those passing either attention check, providing initial support for Hypothesis 1.

We proceed to our main tests for Hypothesis 1, that inattentive survey respondents will provide less accurate answers to questions about their recent and past voting behavior. Here we focus on whether they voted in the current and past elections, and the method that they used to vote in the current and past elections. Our first step is to compare the accuracy of self-reported turnout between respondents passing and failing each type of attention check, as shown in Figure 1. The analysis reported in Figure 1 takes advantage of the fact that we can ask respondents for self-reports about whether they participated in a number of previous elections, and that we can then check the veracity of their self-reports against our administrative data.

It's important to note that we consistently see over-reporting of turnout in our data. But for the purposes of testing Hypothesis 1, the immediate observation from Figure 1 is that respondents failing either the IRI or the IMC attention

Figure 1: Inattentive Respondents Are More Likely to Misreport Turnout



Note: This figure displays the percentage of respondents misreporting turnout in six recent elections among those who failed the IRI (top panel, black), passed the IRI (top panel, grey), failed the IMC (bottom panel, black), or passed the IMC (bottom panel, grey). In each case, dots display the point estimates, thin intervals display 95% confidence intervals, and thick intervals display 90% confidence intervals.

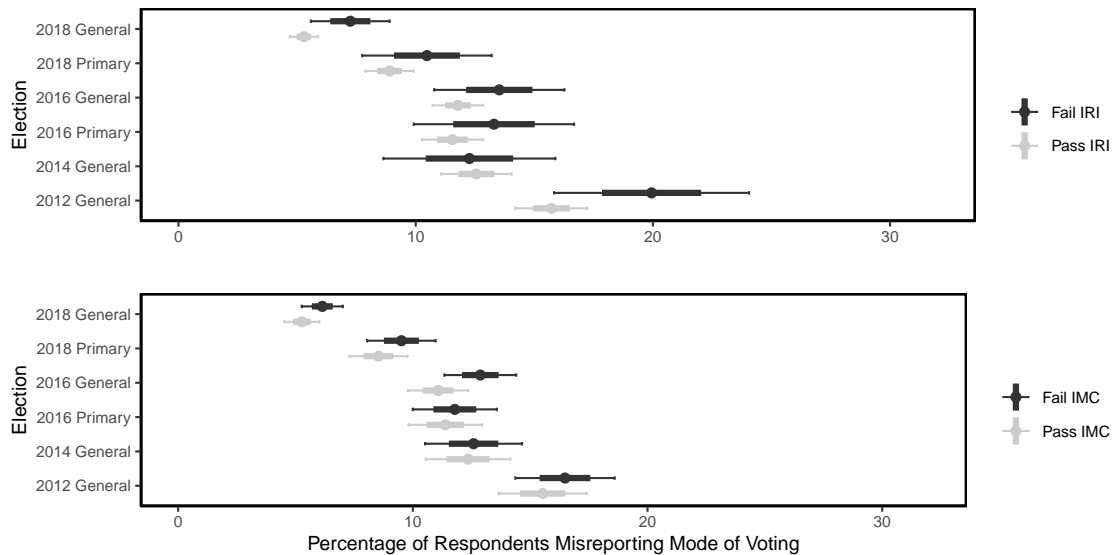
check are significantly more likely to misreport turnout than those passing either attention check. While the magnitude of the difference is modest for the reporting of turnout in the 2018 general election (a few days before the survey), the difference is generally sizable for many past elections. This is not surprising, because trying to remember participation in past elections requires greater cognitive effort (especially primary elections, which are less salient compared to general elections), as subjects need to search their memory more extensively in order to provide a correct report. Moreover, the difference in misreporting rates between inattentive and attentive respondents is larger for primary elections compared to general elections in the same year, which is expected given primary elections are less salient compared to general elections. For example, 18.4% and

17.4% of respondents failing the IRI and IMC, respectively, misreported turnout in the 2018 Primary Election, compared to 12.9% and 10.6% who passed these attention checks.

These results indicate that including inattentive respondents as measured by these attention checks will lead to bias in the estimation of turnout. Also evident from Figure 1, however, is that many respondents who failed the attention checks but nonetheless provided accurate self-reports of turnout. This observation implies that dropping inattentive respondents as measured by these attention checks will increase the variance in the estimation of turnout. Finally, notice that respondents passing the IMC provided the most accurate account of turnout, followed by respondents passing the IRI, and then all respondents. This observation is consistent with the pattern that the IMC screens out respondents more aggressively than the IRI.

Our administrative data also contains information about the method that each voter used to cast their ballot in past elections, and in our survey, we asked each respondent to recall their method of voting. Our second test of Hypothesis 1 examines the difference in terms of accurately reporting the method of voting between respondents passing and failing each type of attention check, with results presented in Figure 2. Again, here we see consistent misreporting in our self-reported survey data. While the differences seen in this figure are less pronounced than the difference shown in Figure 1 for the accuracy of reported turnout, respondents failing the IRIs are significantly more likely to misreport mode of voting than those passing the IRIs, in all of the elections except the 2014 General Election. The same is true for the IMC, but only significant for more recent elections.

Figure 2: Inattentive Respondents Are More Likely to Misreport Mode of Voting



Note: This figure displays the percentage of respondents misreporting mode of voting in six recent elections among those who failed the IRI (top panel, black), passed the IRI (top panel, grey), failed the IMC (bottom panel, black), or passed the IMC (bottom panel, grey). In each case, dots display the point estimates, thin intervals display 95% confidence intervals, and thick intervals display 90% confidence intervals.

One interesting pattern seen in Figure 2 is that generally the extent of misreporting is greater for elections further in the past, than for the most recent election. This pattern suggests that recall of the method that a voter used to obtain and return their ballot might be cognitively demanding for voters, in particular some of them who are not paying close attention as they complete the questionnaire. Compared to turnout questions, the difference in misreporting rates is smaller, which can be partly explained by the fact that method of voting tends to be persistent for many voters (Alvarez et al., 2018), making it a less cognitive demanding task to recall.

The magnitude of the accuracy difference between self-reports of mode of voting by respondents failing and passing the attention checks averages around

1.8% and 1.0% for the IRI and IMC, respectively, for the past two election cycles. This result indicates that including inattentive respondents as measured by these attention checks will lead to bias in the estimation of mode of voting, modest but statistically significant in most cases. As is the case with turnout, we find that many voters reported mode of voting accurately regardless of attentiveness. This pattern suggests that dropping inattentive respondents will lead to less precise estimates of mode of voting. Ultimately, the bias-variance trade-off will govern the best strategy to deal with inattentive respondents, which we shall explore in greater detail later.

We can also stratify our results regarding the misreporting of turnout and mode of voting by attentiveness, given the inclusion of multiple attention checks in our survey (presented in Table SM9 and Table SM10 in the Supplementary Materials). The first observation is that the misreporting rates are virtually the same between the set of respondents passing the first two attention checks and the subset passing all three attention checks. This pattern indicates that the first two attention checks, one IRI and one IMC, suffice to measure attention for a survey of our length (with a median of 13 minutes). Focusing on these two attention checks, the misreporting rates are highest for respondents failing both attention checks and lowest for respondents passing both attention checks, with respondents failing one attention check somewhere in between. Comparing respondents failing both attention checks to those passing both attention checks, the misreporting rate for turnout is 1.3% higher for the most recent election held a few days before the survey but rises to 4.6-9.9% for earlier elections. The difference is smaller for the mode of voting but still over 2.5% for three out of four general elections.

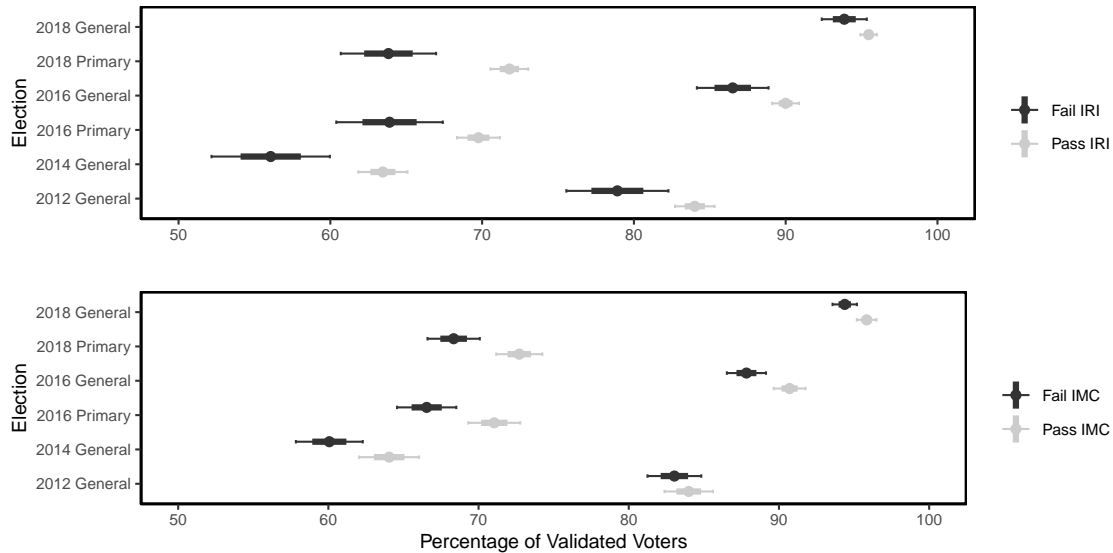
4.2 Correlation between Survey Attention and Outcomes of Interest

So far, we have used validated survey responses to establish that inattentive respondents provided less accurate reports of turnout and method of voting, both for contemporary and past elections. Our findings are consistent with previous research that found inattentive and attentive respondents provide different self-reports of key political behavior and attitudes (for example, Alvarez et al. 2019). While these results highlight the bias in the estimation of political behavior if respondent inattention is not accounted for, dealing with inattentive respondents is complicated by the fact that respondent attention may be correlated with political behavior and attitudes of interest.

Our second hypothesis regards this issue, that inattentive respondents will provide answers that can introduce bias in model estimates. We test our second hypothesis and its corollaries in a number of ways in Figures 3- 6. First, Figure 3 presents the correlation between respondent attention and turnout by comparing validated turnout between respondents passing and failing each type of attention check. Clearly, respondents failing the attention check are significantly less likely to turn out to vote than those passing the attention check, for both IRI and IMC. The magnitude of the difference ranges from 1.6% (2018 General) to 8.0% (2018 Primary) for IRI and 1.0% (2012 General) to 4.5% (2016 Primary). Our result is the first to document the positive correlation between respondent attention and political participation using validated turnout.⁶ The implication is that dropping inattentive respondents, a common way to address respondent inattention, can introduce a source of bias by creating a sample unrepresenta-

⁶Alvarez et al. (2019) found positive correlations between respondent attention and self-reported political participation.

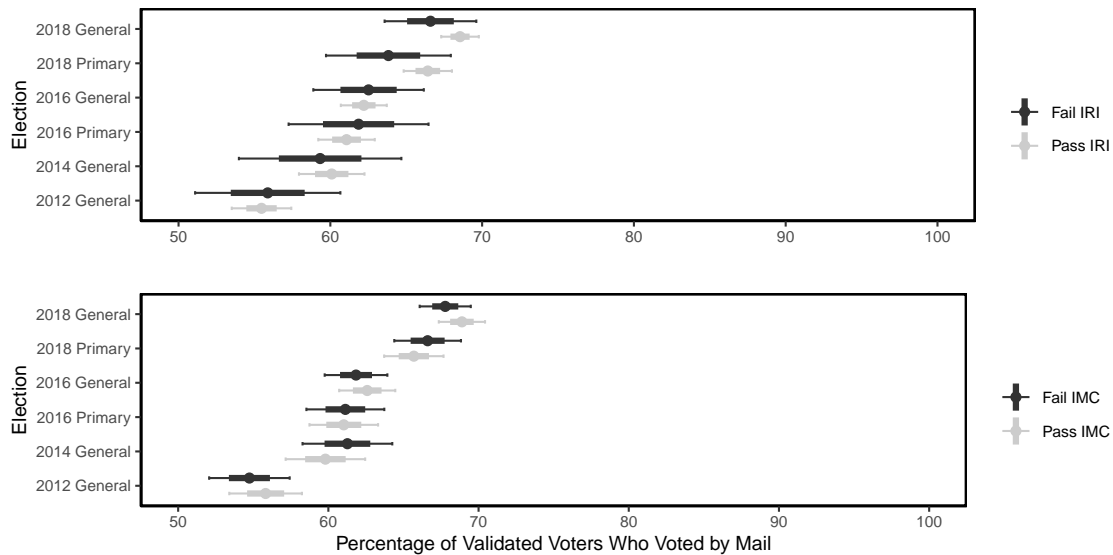
Figure 3: Respondent Attention Is Positively Correlated with Validated Turnout



Note: This figure displays the percentage of respondents confirmed to have participated in six recent elections among those who failed the IRI (top panel, black), passed the IRI (top panel, grey), failed the IMC (bottom panel, black), or passed the IMC (bottom panel, grey). In each case, dots display the point estimates, thin intervals display 95% confidence intervals, and thick intervals display 90% confidence intervals.

tive of the population in terms of political participation. Again, consistent with the fact that the IMC screens out respondents more aggressively than the IRI, respondents passing the IMC had the highest level of turnout, followed by the IRI, and then all respondents.

We next examine the correlation between respondent attention and mode of voting by comparing validated mode of voting between respondents passing and failing each type of attention check in Figure 4. In contrast with turnout, respondents failing the attention check are not different from those passing the check in terms of their actual choice of voting method. This result indicates that dropping inattentive respondents will not introduce bias that comes from an unrepresentative sample in terms of mode of voting. As we discuss in detail

Figure 4: Respondent Attention Is *Not* Correlated with Validated Mode of Voting

Note: This figure displays the percentage of respondents confirmed to have voted by mail in six recent elections among those who failed the IRI (top panel, black), passed the IRI (top panel, grey), failed the IMC (bottom panel, black), or passed the IMC (bottom panel, grey). In each case, dots display the point estimates, thin intervals display 95% confidence intervals, and thick intervals display 90% confidence intervals.

below, the presence or absence of correlation between respondent attention and political behavior or attitudes of interest dictates the magnitude of this important source of bias and ultimately factors heavily into the consideration of strategies to deal with respondent inattention.

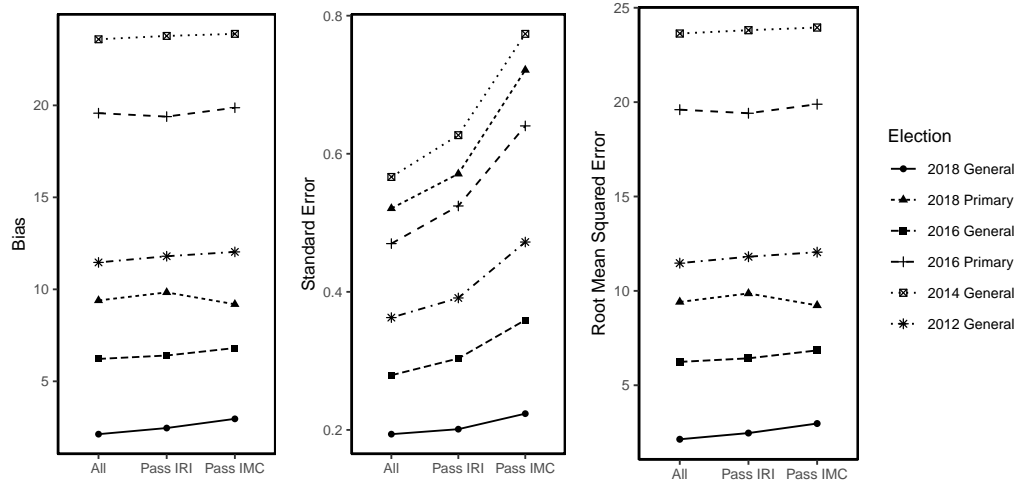
4.3 Keeping versus Dropping Inattentive Survey Respondents

With these results in hand, we can now work toward developing the best strategy to deal with inattentive respondents in different scenarios. To illustrate the considerations going into such decisions, we consider the simplest possible strategy, dropping inattentive respondents identified by the attention checks.

We first compare the performance of turnout estimates based on all respondents, respondents passing the IRI, and respondents passing the IMC, in terms of bias, standard error, and root mean squared error, as shown in Figure 5. While respondents failing the attention checks are significantly more likely to misreport turnout (Figure 1), the reduction in bias from dropping these inattentive respondents is countervailed by the increase in bias from the unrepresentativeness of attentive respondents in terms of turnout (Figure 3). As a result, dropping inattentive respondents does not reduce bias in turnout estimates (left panel). Respondents failing the attention checks, moreover, often provide correct self-reports of turnout nonetheless, leading to the pattern that keeping all respondents yields turnout estimates with the smallest standard errors (middle panel). Given the patterns present in terms of bias and variance, it's unsurprising that we find that the turnout estimates based on all respondents have the smallest root mean squared errors in most cases (right panel).

We next look at the difference in the performance of estimates of proportions of by-mail voters based on all respondents as well as respondents passing each type of attention check, with results presented in Figure 6. In contrast to the results for turnout, since respondent attention as measured by attention checks is uncorrelated with their mode of voting, dropping inattentive respondents does not create an unrepresentative sub-sample that would lead to an increase in bias. Dropping respondents failing attention checks, therefore, leads to a reduction of bias (left panel, with the sole exception of the 2012 General) as these respondents are more likely to misreport their mode of voting. While such procedures still increase imprecision of the estimates compared to keeping all respondents (middle panel), its effect is dwarfed by the reduction of bias.

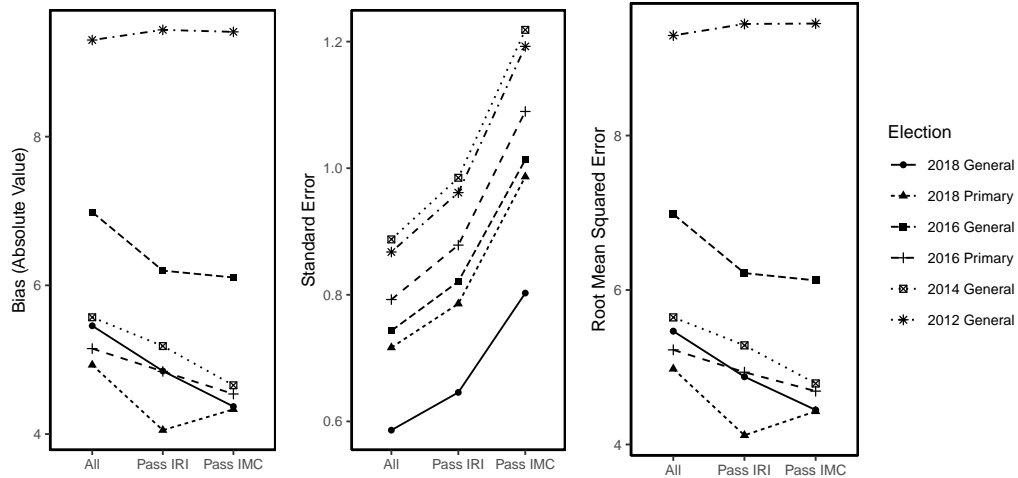
Figure 5: Dropping Inattentive Respondents Does *Not* Reduce Bias in Turnout Estimates



Note: This figure displays the bias (left panel), standard error (middle panel), and root mean squared error (right panel) of turnout estimates (percentage) based on all respondents, respondents passing the IRI, and respondents passing the IMC, for six recent elections. Details are in the Supplementary Materials.

Dropping inattentive respondents, as a result, yields smaller root mean squared errors in estimates of proportions of by-mail voters (right panel). Notice that dropping inattentive respondents reduce bias due to the higher likelihood of misreporting from these respondents and increase variance due to a reduced sample size in both turnout estimates and the estimates of proportions of by-mail voters. The difference in the best strategies for these two quantities of interest comes from the correlation, or the lack thereof, between respondent attention and outcomes of interest, or equivalently, the representativeness of inattentive respondents for the entire sample in terms of the variable of interest.

Figure 6: Dropping Inattentive Respondents Reduces Bias in Voting-by-Mail Estimates



Note: This figure displays the bias (left panel), standard error (middle panel), and root mean squared error (right panel) of estimates of the percentage of by-mail voters based on all respondents, respondents passing the IRI, and respondents passing the IMC, for six recent elections. Details are in the Supplementary Materials.

5. Discussion and Conclusion

The implications of our analysis are important for dealing with inattentive respondents in research using self-administered surveys. We find consistent support for Hypothesis 1, that attentive survey subjects are more likely to provide accurate survey responses, for self-reports of turnout for a number of recent elections, and for self-reports of the method that the voter used to return or cast their ballot. We also show support for Hypothesis 2, that there is evidence that inattentive respondents are generating systematic (not random) error in survey data. Our results also suggest that the difference between inattentive and attentive respondents is larger for questions requiring more cognitive effort.

We obtained these results using a large sample from a diverse population,

and our findings are consistent with the results in studies that have focused on attentiveness in small samples (for example, student or MTurk studies). Our paper, meanwhile, advances survey research in this area by validating survey responses with external administrative data at the individual level, and to the best of our knowledge is the first paper to do. Thus we argue that our results have general and important implications for survey research, in particular for large-scale surveys and polls. Our focus on specific factual questions (voter registration, turnout, mode of voting) is driven by our ability to validate the responses to these questions in our administrative data. We leave it to future research to explore other factual questions that can be validated using other forms of external ground-truth information. Our research design does not allow us to tackle attitudinal questions or survey experiments, unfortunately, for which ground-truth information does not exist.⁷

In the end, our prescriptive advice for researchers begins by recommending that researchers include attention measures in their survey designs, especially in online surveys. There are different ways to measure survey attention, ranging from measuring response time to the types of attention questions we studied in this paper. At this point, we argue that a best practice for online surveys is to measure survey attention.

Next, researchers should check their primary substantive results to determine whether they are robust with respect to survey attention. For example, if the outcome of interest is the prevalence of specific political behavior or attitudes,

⁷In addition to attention checks (including instructional manipulation checks), traditional manipulation checks directly assessing treatment receipt are often used in survey experiments. Some scholars argue that attention checks are preferable to traditional manipulation checks because attention checks are usually not directly related to the experimental randomization, and in any case, can be placed before the experimental randomization to avoid posttreatment bias (Montgomery et al. 2018, but see Kane and Barabas 2019).

the researcher should run the estimation for the entire sample they are using, and then for the subsample of attentive respondents. Similarly, if a survey-based research design involves estimating multivariate models using survey responses, the researcher should test whether the models using the entire sample and the attentive subsample provide homogeneous results using standard methodologies.⁸ We argue that survey research should report whether the results are statistically and substantively similar as a best practice.

If the robustness test for sensitivity to survey attention indicates that attentiveness does not influence the substantive results, the researcher can merely report this result. However, if the substantive results are sensitive to survey attention, more work needs to be done. While – unlike in this paper – the researcher usually cannot determine whether the attentive respondents are representative of the entire sample in terms of the outcome of interest, demographic and other individual characteristics relevant to the outcome of interest can provide useful indications. In particular, the researcher should examine whether there are clear differences between the sample of attentive and inattentive respondents using all of the demographic features and other relevant individual characteristics available in their dataset. If the two samples are very similar overall and few differences are present, our results in this paper suggest that the researcher can drop the inattentive respondents from the estimation sample. The primary consequence of dropping the inattentive respondents will be increased variance in the estimated results, given the loss of information due to dropping some respondents.

But in the situation where the modeled results are not robust with respect to

⁸For example, the researcher can include the attention measure and its interactions with other covariates and test the joint statistical significance of the interactive effects.

survey attention and the inattentive and attentive samples are imbalanced, our results suggest that unless additional action is taken by the researcher, there is likely to be bias in their modeled estimates. In these situations, the researcher will need to include the attentiveness in their modeled results, for example, by including the direct and interactive effects of attentiveness in their model. In the simplest case, this would reduce to stratifying the results by level of attentiveness.

Our research design also allowed us to compare different types of attention checks. Between instructed response items and instructional manipulation checks, two commonly used types of attention checks, we find that instructional manipulation checks screen respondents more aggressively than instructed response items. As a result, while using instructional manipulation checks as screeners further reduces biases compared to instructed response items in many (but not all) cases, it comes at the cost of larger variances for our estimates. For applied researchers, we suggest considering multiple factors. First, while the variance consideration is secondary for a large survey like ours (over 6,900 respondents), it may be a primary consideration for surveys of small or moderate sizes. Secondly, many survey instruments already contain grid/list questions, making instructed response items much less costly than instructional manipulation checks, which require additional standalone questions just for quality control purposes. Lastly, since multiple attention checks are often recommended for online surveys longer than a few minutes, different types of attention checks can be employed to ensure the robustness of results to various levels of respondent attention screening.

References

- Ahler, D. J., C. E. Roush, and G. Sood (2021). The micro-task market for lemons: data quality on amazon's mechanical turk. *Political Science Research and Methods*, 1–20.
- Alvarez, R. M., L. R. Atkeson, I. Levin, and Y. Li (2019). Paying attention to inattentive survey respondents. *Political Analysis* 27(2), 145–162.
- Alvarez, R. M., I. Levin, and Y. Li (2018). Fraud, convenience, and e-voting: how voting experience shapes opinions about voting technology. *Journal of Information Technology & Politics* 15(2), 94–105.
- Anduiza, E. and C. Galais (2017). Answering without reading: Imcs and strong satisficing in online surveys. *International Journal of Public Opinion Research* 29(3), 497–519.
- Atkeson, L. R., A. N. Adams, and R. M. Alvarez (2014). Nonresponse and mode effects in self- and interviewer-administered surveys. *Political Analysis* 22(3), 304–320.
- Berinsky, A. J., M. F. Margolis, and M. W. Sances (2014). Separating the shirkers from the workers? making sure respondents pay attention on self-administered surveys. *American Journal of Political Science* 58(3), 739–753.
- Berinsky, A. J., M. F. Margolis, M. W. Sances, and C. Warshaw (2021). Using screeners to measure respondent attention on self-administered surveys: Which items and how many? *Political Science Research and Methods* 9(2), 430–437.
- Clifford, S. and J. Jerit (2015). Do attempts to improve respondent attention increase social desirability bias? *Public Opinion Quarterly* 79(3), 790–802.
- Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology* 66, 4–19.

- Curran, P. G. and K. A. Hauser (2019). I'm paid biweekly, just not by leprechauns: Evaluating valid-but-incorrect response rates to attention check items. *Journal of Research in Personality* 82, 103849.
- Enamorado, T., B. Fifield, and K. Imai (2019). Using a probabilistic model to assist merging of large-scale administrative records. *American Political Science Review* 113(2), 353–371.
- Hauser, D. J. and N. Schwarz (2015). It's a trap! instructional manipulation checks prompt systematic thinking on “tricky” tasks. *Sage Open* 5(2), 2158244015584617.
- Kane, J. V. and J. Barabas (2019). No harm in checking: Using factual manipulation checks to assess attentiveness in experiments. *American Journal of Political Science* 63(1), 234–249.
- Kim, S.-y. S., S. Schneider, and R. M. Alvarez (2020). Evaluating the quality of changes in voter registration databases. *American Politics Research* 48(6), 670–676.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology* 5(3), 213–236.
- Kuhn, P. M. and N. Vivyan (2021). The misreporting trade-off between list experiments and direct questions in practice: Partition validation evidence from two countries. *Political Analysis*, 1–22.
- Kung, F. Y., N. Kwok, and D. J. Brown (2018). Are attention check questions a threat to scale validity? *Applied Psychology* 67(2), 264–283.
- Montgomery, J. M., B. Nyhan, and M. Torres (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science* 62(3), 760–775.

- Oppenheimer, D. M., T. Meyvis, and N. Davidenko (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of experimental social psychology* 45(4), 867–872.
- Read, B., L. Wolters, and A. J. Berinsky (2021). Racing the clock: Using response time as a proxy for attentiveness on self-administered surveys. *Political Analysis*.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review* 63(2), 129.
- Thomas, K. A. and S. Clifford (2017). Validity and mechanical turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior* 77, 184–197.

Supplementary Materials

Details about Administrative Data

Administrative Data

Given the importance of administrative data for the work we report in this paper, here we provide additional details for interested readers. Administrative data, like voter registration records, have long been used by political science researchers. Some use voter registration records to provide contact information for field experiments (e.g., Gerber and Green 2000), as sampling frames to improve the accuracy of electoral polling (e.g. Green and Gerber 2006), or to study the misreporting of voter turnout in surveys (e.g., Ansolabehere and Hersch 2017). These studies all assume that the administrative records they use are accurate.

In the United States, states are now generally required to have a statewide voter registration database due to the *Help America Vote Act* (HAVA, 2002). These statewide voter registration datasets are used for many purposes, other than academic research. Campaigns and political parties use them for voter mobilization and persuasion (Hersch 2015). State, county, and municipal election officials use these data for voter information activities, to send vote-by-mail ballots, to allocate resources for in-person voting, and to authenticate eligible voters during an election. Clearly election officials have important incentives to keep these data as accurate as possible.

The procedures that states can use to maintain the accuracy of their voter registration datasets is regulated by the *National Voter Registration Act* (NVRA, 1993). Each state then issues its own mechanisms for compliance with NVRA, and California's rules are provided by the Secretary of State (<https://www.sos.ca>).

[gov/elections/voter-registration/nvra/laws-standards/nvra-manual](https://www.elections.ca.gov/elections/voter-registration/nvra/laws-standards/nvra-manual)). In general, the NVRA regulations provide guidance about how potentially ineligible voters can be moved to an inactive status or removed from the file, usually based on changes like moving outside of the jurisdiction or death. In our paper we use data from Orange County, and their registration list maintenance procedures are provided online (<https://www.ocvote.com/registration/maintaining-an-accurate-voter-list>).

As has been noted in recent research using these administrative data from Orange County (Kim et al., 2020), in a jurisdiction of this size, there will be record changes, new records, and removed records. However, this recent research has shown that these changes in the administrative data are relatively small in relation to the overall file, and there is no research that we are aware of that would indicate that file maintenance or inaccuracies in the administrative data more generally should have any effect on our estimate of the quantities we examine in our paper. In particular, the components of the voter registration data that are most subject to change and possible error are the fields with voter identifying information — their name, address, and other voter-provided information. The information from the administrative files we used in our paper, each registered voter's participation in current and past elections, comes from the county's election management system and we have every reason to believe that these components of the administrative record are accurate.

By focusing our research on a large and diverse county, we argue that our results should generalize. Orange County, California, is a very large and diverse election jurisdiction located in Southern California. With about 3 million residents, in the 2018 general election Orange County had just over 2 million citizens

eligible to vote, with nearly 1.6 million registered voters.⁹. It is one of the largest election jurisdictions in the United States, closely divided in partisan registration: in 2018, the county had 541,665 registered Republicans, 523,624 registered Democrats, and 429,675 registered No Party Preference voters.

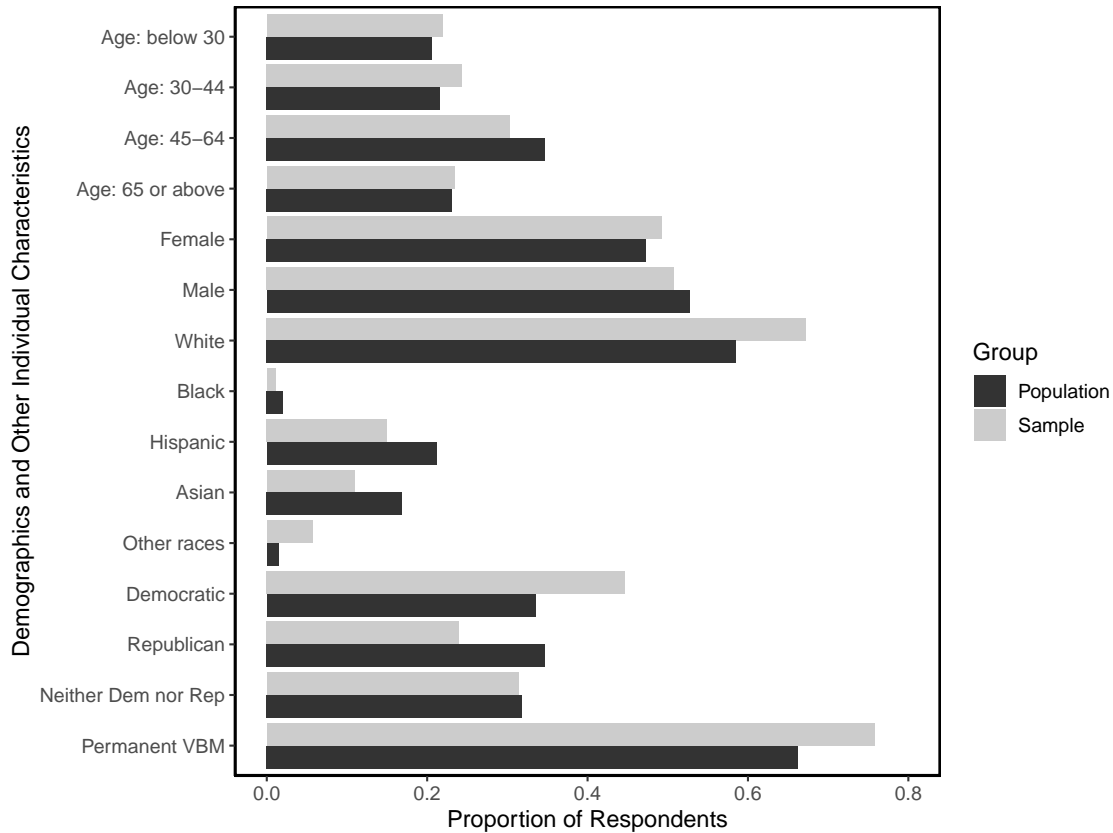
Details about the Survey

Among 6,952 respondents who completed our survey, we are able to determine basic demographic information for 6,912 respondents. Figure 7 shows the demographic compositions and the distributions of other individual characteristics for our survey respondents and the population of Orange County registered voters before the November 2018 general election.

Registered voters of different ages and genders are well represented in our survey sample. There are slightly more (2.7%) respondents between 30 and 44 years old, fewer (4.4%) respondents from age group 45-64, and fewer (2.0%) female respondents, compared to the population of registered voters. Our survey sample exhibits more imbalances in race/ethnicity and party registration. More white voters and fewer Hispanic or Latino and Asian American voters participated in the survey (by 8.7%, 6.1% and 6.0% respectively). Meanwhile, while 34.7% of voters registered with the Republican party in Orange County, only 24.0% of those who completed our survey are Republican voters. On the other hand, 44.6% of respondents are Democratic voters, compared to 33.5% in the population. The disparities in terms of race/ethnicity and party registration in our survey sample are expected given our knowledge about survey participation in general and consistent with other surveys with voluntary participation. Fi-

⁹<https://elections.cdn.sos.ca.gov/sov/2018-general/sov/02-county-voter-reg-stats-by-county.pdf>

Figure 7: Respondent Composition of the Survey



nally, 66.2% of registered voters in Orange County are permanent vote-by-mail voters, whereas the percentage is 75.8% in our sample; the distribution of cities of residence for our sample tracks the population well.

Of course, we could use calibration weighting to adjust our sample to population parameters (Caughey et al. 2020). We decided here to use unweighted data, as we did not want to make population estimates specific to Orange County. That said, our unweighted estimates should generalize to other similar large-scale surveys and polls. Weighting to Orange County population parameters would not change the general results we see in our analysis, and as Orange County is

a large and diverse population, our results should apply to surveys and polls of other large and diverse populations.

Attention Check Questions

Main IRI: Do you support or oppose any of the following proposals for new ways of voting or conducting elections?

[For each of the following statements, respondents can select *Support strongly*, *Support somewhat*, *Neither support nor oppose*, *Oppose somewhat*, *Oppose strongly*]

- Allow absentee voting over the Internet
- Run all elections by mail
- Allow people to register on Election Day at the polls
- Require all people to show government issued photo ID when they vote
- Require electronic voting machines to print a paper backup of the ballot
- *For half of the respondents:* For survey quality control, select Oppose strongly
- *For half of the respondents:* For survey quality control, select Support strongly

Note: To minimize order effects, response options are presented in ascending order for half of the respondents, and in descending order for the other half of the respondents.

Main IMC: When a big news story breaks people often go online to get up-to-the-minute details on what is going on. People differ in which websites they trust to get this information. For survey quality control, please ignore the question and select ABC News and Reuters as your two answers.

When there is a big news story, which websites you would visit first?

- New York Times website
- The Drudge Report
- The Associated Press (AP) website
- Huffington Post
- Google News
- NBC News website
- Washington Post website
- ABC News website
- National Public Radio (NPR) website
- CNN.com
- CBS News website
- USA Today website
- FoxNews.com
- Reuters website
- New York Post Online

- MSNBC.com
- Yahoo! News
- None of these websites

Note: The position of prompt *For survey quality control, please ignore the question and select ABC News and Reuters as your two answers* is randomized between the middle of the question and the end of the question.

Attention and Accuracy of Self-Reports About Birth Year, City of Residence and Voter Registration

In Table SM2 we present a comparison of self-reported birth year, city of residence and voter registration by failure or passage of the IRI/IMC attention questions. We see systematic evidence in this table that respondents who fail the attention checks are more likely to report incorrect information for these types of questions. The differences in the incorrect reporting rates between attentive and inattentive respondents are slight, usually a difference of 1% to 3%. It is important to note that these are questions that should not be difficult for respondents to answer, they are simple factual questions that should not require much cognition, which is why it not surprising that we do not see differences of larger magnitudes.

Turnout and Mode of Voting Survey Questions

Note: Turnout Question 1 (2018 General Turnout), Turnout Question 2a, and Mode of Voting Question 1 (2018 General Mode of Voting) appear right before the first attention check, while Turnout Question 2b, Turnout Question 3 (Turnout in Previous Elections) and Mode of Voting Question 2 (Mode of Voting in Previous Elections) appear right after the second attention check.

Turnout Question 1: Which of the following statements best describes you?

- I did not vote in the election this November
- I thought about voting this time, but didn't
- I usually vote, but didn't this time
- I tried to vote, but was not allowed to when I tried
- I tried to vote, but it ended up being too much trouble
- I definitely voted in the November General Election

[If the respondent chose one of the last three response options in Turnout Question 1, they receive the following two questions.]

Turnout Question 2a Was this your first time voting, or have you voted in elections before? [Randomize first two options.]

- I am a first time voter
- I have voted in elections before
- I don't know

Mode of Voting Question 1 How did you vote this election?

- Voted in person on Election Day (at a polling place or precinct)
- Voted in person before Election Day
- Voted by mail or absentee ballot by mail
- I don't know

[If the respondent chose one of the first three response options in Turnout Question 1, they receive the following question.]

Turnout Question 2b Have you voted in elections before? [Randomize first two options.]

- Yes
- No
- I don't know

[If the respondent chose I have voted in elections before in Turnout Question 2a or Yes in Turnout Question 2b, they receive the following questions.]

Turnout Question 3 Earlier you mentioned that you have voted in elections before. Did you vote in the following general elections? [For each of the following statements, respondents can select *Yes*, *No*, or *I don't remember*.]

- November 2016 Presidential General Election
- November 2014 Midterm General Election
- November 2012 Presidential General Election

... in the following primary elections?

- June 2018 Statewide Primary Election

- June 2016 Presidential Primary Election

Mode of Voting Question 2 How did you vote in these general elections you voted in? [For each of the following statements, respondents can select *Voted in person*, *Voted by mail or absentee ballot by mail*, or *I don't know*. Only those elections that a respondent answers Yes to in Turnout Question 3 are included.]

- November 2016 Presidential General Election
- November 2014 Midterm General Election
- November 2012 Presidential General Election

How did you vote in the primary elections you voted in?

- June 2018 Statewide Primary Election
- June 2016 Presidential Primary Election

Table SM1: Demographics and Passage of Attention Checks (Logistic Regression)

| | IRI | | IMC | |
|----------------|---------|------------|---------|------------|
| | Avg. ME | Std. Error | Avg. ME | Std. Error |
| Below 30 | | | | |
| 30 - 44 | -0.03 | 0.01 | -0.03 | 0.02 |
| 45 - 64 | -0.07 | 0.01 | -0.11 | 0.02 |
| 65 or above | -0.12 | 0.01 | -0.16 | 0.02 |
| Female | | | | |
| Male | 0.00 | 0.01 | -0.02 | 0.01 |
| Rather not say | -0.04 | 0.05 | -0.08 | 0.06 |
| Self-describe | 0.02 | 0.05 | -0.06 | 0.07 |
| HS or less | | | | |
| Some college | 0.06 | 0.03 | 0.13 | 0.03 |
| 2-year college | 0.03 | 0.03 | 0.16 | 0.04 |
| 4-year college | 0.08 | 0.03 | 0.19 | 0.03 |
| Postgraduate | 0.11 | 0.03 | 0.24 | 0.03 |
| White | | | | |
| Hispanic | -0.12 | 0.02 | -0.15 | 0.02 |
| Asian | -0.11 | 0.02 | -0.09 | 0.02 |
| Black | -0.19 | 0.05 | -0.17 | 0.06 |
| Other races | -0.08 | 0.02 | -0.23 | 0.03 |

Table SM2: Accuracy of Self-Reported Birth Year, City of Residence and Voter Registration

| | Incorrect | Correct | Skip | N |
|---------------------|-----------|---------|------|------|
| Birth Year: | | | | |
| Fail IRI | 6.3% | 92.9% | 0.8% | 999 |
| Pass IRI | 3.4% | 96.0% | 0.6% | 5648 |
| Fail IMC | 4.8% | 94.8% | 0.4% | 3072 |
| Pass IMC | 3.2% | 96.2% | 0.6% | 3618 |
| City of Residence: | | | | |
| Fail IRI | 8.4% | 90.9% | 0.6% | 971 |
| Pass IRI | 6.8% | 92.8% | 0.4% | 5474 |
| Fail IMC | 7.9% | 91.7% | 0.4% | 2994 |
| Pass IMC | 6.8% | 92.9% | 0.3% | 3496 |
| Voter Registration: | | | | |
| Fail IRI | 8.5% | 88.0% | 3.4% | 986 |
| Pass IRI | 6.9% | 90.9% | 2.2% | 5608 |
| Fail IMC | 8.3% | 88.7% | 3.0% | 3037 |
| Pass IMC | 6.3% | 92.0% | 1.8% | 3595 |

Table SM3: Inattentive Respondents Are More Likely to Misreport Turnout

| | All | Fail IRI | Pass IRI | Fail IMC | Pass IMC |
|--------------|---------------|---------------|---------------|---------------|---------------|
| 2018 General | 3.3 (0.2) | 4.1 (0.6) | 3.1 (0.2) | 3.9 (0.3) | 2.8 (0.3) |
| 2018 Primary | 13.7 (0.4) | 18.4 (1.3) | 12.9 (0.5) | 17.4 (0.7) | 10.6 (0.6) |
| 2016 General | 7.3 (0.4) | 10.2 (1.1) | 6.7 (0.4) | 9.0 (0.6) | 6.0 (0.4) |
| 2016 Primary | 21.2 (0.6) | 27.2 (1.7) | 20.3 (0.7) | 23.9 (0.9) | 18.6 (0.8) |
| 2014 General | 23.6 (0.7) | 27.5 (1.9) | 23.0 (0.8) | 26.5 (1.1) | 21.3 (0.9) |
| 2012 General | 13.5 (0.6) | 17.0 (1.6) | 13.0 (0.6) | 14.4 (0.9) | 12.6 (0.8) |

Note: This table displays the percentage of respondents misreporting turnout in six recent elections among those who failed the IRI, passed the IRI, failed the IMC, or passed the IMC.

Table SM4: Inattentive Respondents Are More Likely to Misreport Mode of Voting

| | All | Fail IRI | Pass IRI | Fail IMC | Pass IMC |
|--------------|---------------|---------------|---------------|---------------|---------------|
| 2018 General | 5.5 (0.3) | 7.2 (0.8) | 5.3 (0.3) | 6.1 (0.4) | 5.3 (0.4) |
| 2018 Primary | 9.1 (0.5) | 10.5 (1.4) | 8.9 (0.5) | 9.5 (0.7) | 8.5 (0.6) |
| 2016 General | 11.9 (0.5) | 13.5 (1.4) | 11.8 (0.5) | 12.9 (0.8) | 11.1 (0.7) |
| 2016 Primary | 11.6 (0.6) | 13.3 (1.7) | 11.5 (0.7) | 11.8 (0.9) | 11.4 (0.8) |
| 2014 General | 12.4 (0.7) | 12.3 (1.8) | 12.6 (0.8) | 12.6 (1.1) | 12.4 (0.9) |
| 2012 General | 16.1 (0.7) | 19.9 (2.1) | 15.7 (0.8) | 16.5 (1.1) | 15.5 (1.0) |

Note: This table displays the percentage of respondents misreporting mode of voting in six recent elections among those who failed the IRI, passed the IRI, failed the IMC, or passed the IMC.

Table SM5: Respondent Attention Is Positively Correlated with Validated Turnout

| | All | Fail IRI | Pass IRI | Fail IMC | Pass IMC |
|--------------|---------------|---------------|---------------|---------------|---------------|
| 2018 General | 95.2 (0.3) | 93.9 (0.8) | 95.5 (0.3) | 94.4 (0.4) | 95.8 (0.3) |
| 2018 Primary | 70.7 (0.6) | 63.8 (1.6) | 71.8 (0.6) | 68.3 (0.9) | 72.7 (0.8) |
| 2016 General | 89.4 (0.4) | 86.5 (1.2) | 90.0 (0.5) | 87.8 (0.7) | 90.7 (0.5) |
| 2016 Primary | 68.8 (0.7) | 63.9 (1.8) | 69.8 (0.7) | 66.5 (1.0) | 71.0 (0.9) |
| 2014 General | 62.2 (0.7) | 56.1 (2.0) | 63.5 (0.8) | 60.1 (1.1) | 64.0 (1.0) |
| 2012 General | 83.5 (0.6) | 78.9 (1.7) | 84.0 (0.7) | 83.0 (0.9) | 84.0 (0.8) |

Note: This table displays the percentage of respondents confirmed to have participated in six recent elections among those who failed the IRI, passed the IRI, failed the IMC, or passed the IMC.

Table SM6: Respondent Attention Is *Not* Correlated with Validated Mode of Voting

| | All | Fail IRI | Pass IRI | Fail IMC | Pass IMC |
|--------------|---------------|---------------|---------------|---------------|---------------|
| 2018 General | 68.1 (0.6) | 66.6 (1.5) | 68.5 (0.6) | 67.8 (0.9) | 68.9 (0.8) |
| 2018 Primary | 66.1 (0.7) | 63.8 (2.1) | 66.4 (0.8) | 66.6 (1.1) | 65.7 (1.0) |
| 2016 General | 62.0 (0.7) | 62.5 (1.9) | 62.2 (0.8) | 61.8 (1.1) | 62.6 (1.0) |
| 2016 Primary | 61.3 (0.9) | 61.9 (2.3) | 61.1 (0.9) | 61.1 (1.3) | 61.0 (1.2) |
| 2014 General | 60.3 (1.0) | 59.3 (2.7) | 60.1 (1.1) | 61.3 (1.5) | 59.8 (1.3) |
| 2012 General | 55.6 (0.9) | 55.9 (2.4) | 55.5 (1.0) | 54.7 (1.4) | 55.8 (1.2) |

Note: This table displays the percentage of respondents confirmed to have voted by mail in six recent elections among those who failed the IRI, passed the IRI, failed the IMC, or passed the IMC.

Table SM7: Dropping Inattentive Respondents Does *Not* Reduce Bias in Turnout Estimates

| | All | Pass IRI | Pass IMC |
|--------------|------|----------|----------|
| Bias: | | | |
| 2018 General | 2.1 | 2.5 | 3.0 |
| 2018 Primary | 9.4 | 9.8 | 9.2 |
| 2016 General | 6.2 | 6.4 | 6.8 |
| 2016 Primary | 19.6 | 19.4 | 19.9 |
| 2014 General | 23.6 | 23.8 | 23.9 |
| 2012 General | 11.5 | 11.8 | 12.0 |
| Std. Error: | | | |
| 2018 General | 0.2 | 0.2 | 0.2 |
| 2018 Primary | 0.5 | 0.6 | 0.7 |
| 2016 General | 0.3 | 0.3 | 0.4 |
| 2016 Primary | 0.5 | 0.5 | 0.6 |
| 2014 General | 0.6 | 0.6 | 0.8 |
| 2012 General | 0.4 | 0.4 | 0.5 |
| RMSE: | | | |
| 2018 General | 2.1 | 2.5 | 3.0 |
| 2018 Primary | 9.4 | 9.9 | 9.2 |
| 2016 General | 6.2 | 6.4 | 6.8 |
| 2016 Primary | 19.6 | 19.4 | 19.9 |
| 2014 General | 23.6 | 23.8 | 24.0 |
| 2012 General | 11.5 | 11.8 | 12.1 |

Note: Bias is calculated as the percentage of voters according to the self-reports minus the percentage of voters according to administrative records. Std. Error is obtained from the standard error associated with the intercept by running a linear regression of turnout containing only the intercept. RMSE is calculated using 1,000 bootstrap samples with replacements.

Table SM8: Dropping Inattentive Respondents Reduces Bias in Voting-by-Mail Estimates

| | All | Pass IRI | Pass IMC |
|--------------|------|----------|----------|
| Bias: | | | |
| 2018 General | -5.5 | -4.8 | -4.4 |
| 2018 Primary | -4.9 | -4.1 | -4.3 |
| 2016 General | -7.0 | -6.2 | -6.1 |
| 2016 Primary | -5.1 | -4.8 | -4.5 |
| 2014 General | -5.6 | -5.2 | -4.7 |
| 2012 General | -9.3 | -9.4 | -9.4 |
| Std. Error: | | | |
| 2018 General | 0.6 | 0.6 | 0.8 |
| 2018 Primary | 0.7 | 0.8 | 1.0 |
| 2016 General | 0.7 | 0.8 | 1.0 |
| 2016 Primary | 0.8 | 0.9 | 1.1 |
| 2014 General | 0.9 | 1.0 | 1.2 |
| 2012 General | 0.9 | 1.0 | 1.2 |
| RMSE: | | | |
| 2018 General | 5.5 | 4.9 | 4.4 |
| 2018 Primary | 5.0 | 4.1 | 4.4 |
| 2016 General | 7.0 | 6.2 | 6.1 |
| 2016 Primary | 5.2 | 4.9 | 4.7 |
| 2014 General | 5.6 | 5.3 | 4.8 |
| 2012 General | 9.3 | 9.4 | 9.5 |

Note: Bias is calculated as the percentage of by-mail voters according to the self-reports minus the percentage of voters according to administrative records. Std. Error is obtained from the standard error associated with the intercept by running an linear regression of voting-by-mail containing only the intercept. RMSE is calculated using 1,000 bootstrap samples with replacements.

Table SM9: Turnout Misreporting by Level of Attentiveness

| | Fail Both | Fail One Pass One | Pass Both | Fail Three | Pass Three |
|--------------|---------------|----------------------|---------------|---------------|---------------|
| 2018 General | 4.0 (0.7) | 4.0 (0.4) | 2.7 (0.3) | 3.9 (0.9) | 2.6 (0.3) |
| 2018 Primary | 20.4 (1.5) | 16.4 (0.8) | 10.6 (0.6) | 19.8 (2.0) | 10.4 (0.6) |
| 2016 General | 10.7 (1.2) | 8.2 (0.7) | 5.9 (0.5) | 10.8 (1.6) | 6.0 (0.5) |
| 2016 Primary | 28.4 (1.9) | 22.7 (1.1) | 18.5 (0.8) | 27.2 (2.5) | 18.3 (0.8) |
| 2014 General | 28.7 (2.1) | 26.0 (1.3) | 21.2 (0.9) | 24.6 (2.6) | 21.1 (1.0) |
| 2012 General | 17.2 (1.8) | 13.8 (1.0) | 12.6 (0.8) | 15.0 (2.3) | 12.5 (0.8) |

Note: This table displays the percentage of respondents misreporting turnout by the number of attention checks passed/failed.

Table SM10: Mode of Voting Misreporting by Level of Attentiveness

| | Fail Both | Fail One Pass One | Pass Both | Fail Three | Pass Three |
|--------------|---------------|----------------------|---------------|---------------|---------------|
| 2018 General | 8.0 (1.0) | 5.3 (0.5) | 5.4 (0.4) | 7.7 (1.3) | 5.4 (0.4) |
| 2018 Primary | 8.6 (1.4) | 10.5 (0.9) | 8.2 (0.6) | 7.9 (1.8) | 8.2 (0.6) |
| 2016 General | 13.9 (1.6) | 13.0 (0.9) | 11.0 (0.7) | 12.8 (2.0) | 10.9 (0.7) |
| 2016 Primary | 12.6 (1.9) | 12.0 (1.1) | 11.3 (0.8) | 12.6 (2.4) | 11.2 (0.8) |
| 2014 General | 13.1 (2.1) | 12.9 (1.2) | 12.4 (0.9) | 13.9 (2.7) | 12.4 (1.0) |
| 2012 General | 20.9 (2.4) | 15.4 (1.2) | 15.7 (1.0) | 18.9 (2.9) | 15.6 (1.0) |

Note: This table displays the percentage of respondents misreporting mode of voting by the number of attention checks passed/failed.

Supplementary Materials References

- Ansolabehere, S., and Hersh, E. 2017. Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate. *Political Analysis*, 20(4), 437-459.
- Caughey, D., Berinsky, A. J., Chatfield, S., Hartman, E., Schickler, E., and Sekhon, J. S. 2020. Target Estimation and Adjustment Weighting for Survey Nonresponse and Sampling Bias New York: Cambridge University Press.
- Gerber, A. S., and Green, D. P. 2000. The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review*, 94(3), 653-663.
- Green, D. P. and Gerber, A. S. 2006. *Can Registration-Based Sampling Improve the Accuracy of Midterm Election Forecasts?* Public Opinion Quarterly, 70(2), 197-223.
- Hersh, E. D. 2015. *Hacking the Electorate: How Campaigns Perceive Voters..* New York: Cambridge University Press.