Why Does the American National Election Study Overestimate Voter Turnout?

July 15, 2017

Abstract

Surveys are the key tool for understanding political behavior, but they are subject to biases that render their estimates about the frequency of socially-desirable behaviors innacurate. For decades the American National Election Studies has over-estimated voter turnout, though the causes of this persistent bias are poorly understood. The face-to-face component of the 2012 ANES produced a turnout estimate at least 13 points higher than the benchmark voting-eligible population turnout rate. We consider three explanations for this overestimate in the survey: non-response bias, over-reporting and the possibility that the ANES constitutes an inadvertent mobilization treatment. Analysis of turnout data supplied by voter file vendors allows the three phenomena to be measured for the first time in a single survey. We find that over-reporting is the largest contributor, responsible for six percentage points of the turnout over-estimate, while non-response bias and mobilization account for an additional 4 and 3 percentage points, respectively.

Word count: 9132

The Problem

Surveys are the central tool by which social scientists learn about political behavior. Survey respondents reporting on their own behaviors and attitudes formed the basis of our understanding of turnout(Wolfinger and Rosenstone 1980), partisanship(Campbell, Miller and Converse 1960) and activism (Verba, Schlozman and Brady 1995). But can survey measurements be trusted?

It is widely recognized among survey methodologists that many sources of error afflict surveys, and that these sources of error are, with properly designed studies, measurable. The Total Survey Error framework dominates the survey methods field(Weisberg 2009), giving researchers a broad conceptual framework to understand and measure the ways in which survey measurements deviate from the true levels of behaviors and attitudes in the population.

In some cases, the bias can be attributable to a single source, as an over-abundance of older respondents in a sample might be attributable to non-response bias. But with many political and behavioral survey items, the true nature of the bias can be difficult to identify: non-response bias, misreporting and survey conditioning may all play a role. Decomposing the overall bias into these various sources is vital to understanding what surveys are truly measuring and correcting these biases going forward.

It's difficult to overstate the importance of surveys to poltical research. Pollsters are at the heart of political campaigns and serve in senior White House positions. Horserace polls inform news coverage and have spurred a cottage industry of poll averaging for election prediction. Most importandly for political scientists, they underlie virtually everything the field knows about voters' individual behaviors. Given the profound importance of this methodolgical tool, understanding with the greatest specificity the sources of biases in polls is essential.

From its origins in 1952, the American National Election Study has over-estimated voter turnout by sizable margins. The difference between the ANES, survey-based estimate of voter turnout and estimates based on official statistics was 11 points in 1952 and grew reasonably steadily thereafter, reaching 24 points in 1996 (Burden 2000, Figure 1).

The ANES over-estimate of turnout persisted in 2012. Using official vote tallies and careful counts of eligible voters, McDonald (2013) reports 2012 voter turnout as 58.2% of the voting eligible population (VEP) and 53.6% of the voting age population (VAP). The 2012 American National Election Study was administered using two survey modes — in-person interviewing and self-complete, web surveys — yielding self-reported turnout estimates of 71.0% and 73.2%, respectively.[1]

The ANES sampling frame excludes non-citizens, so ANES turnout estimates ought to correspond to the VEP rate rather than the lower, voting age population VAP rates. The in-person arm of ANES has a slight degree of under-coverage of the voting eligible population[2], but this can hardly account for the large discrepancy between the ANES turnout estimate and the estimated VEP rate.[3] Nor can sampling

[1]These estimates are actually lower bounds, computed by treating all respondents with missing data as non-voters. 6.1% of respondents generate missing data in 2012. Missing data includes "don't know" responses, refusals and respondents not taking the post-election survey who did not report early voting on the pre-election survey. Assuming (somewhat implausibly) that all respondents generating missing data would have reported turning out sees the estimated turnout rate among respondents interviewed in-person rise to 77.1% (an upper bound). Unless otherwise indicated, all estimates and the analysis throughout use the post-stratification weights on the ANES data file.

[2]e.g., citizens voting from outside the United States are not in the ANES frame; ANES does not sample face-to-face cases in Alaska and Hawaii, nor does ANES seek in-person interviews in institutional settings such as prisons, hospitals, nursing homes, college dormitories, or military barracks.

[3]The possibility that coverage error was a substantial source of bias in the ANES turnout estimate was considered — and dismissed — by Clausen (1968) and Traugott and Katosh (1979).

error be the culprit. ANES sample sizes are large enough to produce 95% credible intervals of about $\pm$ 2.6 percentage points around the turnout point estimate. The only available conclusion is that the ANES estimates of turnout are biased upwards by about 13 to 19 percentage points in 2012, a bias of over 10 standard errors.

This bias is no small matter. ANES is a long-standing, well resourced and high quality survey project. The face-to-face component of ANES utilizes rigorous sampling methods, implemented by teams of highly qualified and experienced sampling statisticians, with field operations conducted by carefully managed, trained personnel. Accordingly, ANES is widely regarded as a benchmark , a near-canonical record of the American electorate. That the world's "gold standard" election survey (e.g., Aldrich and McGraw 2011) carries such a large and enduring bias with respect to a fundamental political behavior clearly warrants investigation.

Getting the turnout estimate right isn't just a matter of prestige. As a survey that measures political attitudes in an election year, being able to use survey respondents to assess the prevalence of attitudes or characteristics among voters is crucial. To understand the differences between Romney and Obama voters, one needs to know who actually turned out to vote for the two candidates. One obvious benchmark is whether the ANES gets the two-party vote share correct. Among self-reported voters choosing Obama or Romney, 53.2% report voting for Obama, 1.2 percentage points above the true result of 52%, or about one standard error. But among validated voters (see Table 2), 51.9% supported Obama.[4] While this improvement might be due to sampling variability alone, it is an indication that turnout validation might lead to more accurate results on other survey items.

[4]The difference in Obama support between over-reporters and validated voters is 6.4 p.p., the weighted t-test for the difference of means has a p-value of .058.

The unusual amount of control and visibility that is built into the administration of the ANES allows us to measure quantities that would be hard to measure in a typical phone or internet survey. We use this information to make a number of contributions to the understanding of the ANES overestimate of voter turnout. First, we match ANES respondents and sample households against multiple national databases of registered voters and commercial records. These databases let us validate respondents' self-reports of turnout and to compare responder households to non-responder households, informing our estimate of non-response bias. We exploit the random selection of a single respondent in multiple person households to identify the intent to treat effect of being the selected respondent, conditional on living in a rostered household. For the first time, we are able to rigorously estimate the causal effect of being an ANES respondent on voter turnout, long thought to be a possible side-effect of taking the survey and contributing to the resulting over-estimate of voter turnout.

To our knowledge, no modern research effort has managed to measure and decompose these various sources of turnout bias in a single survey. While the individual components of the turnout estimate have been measured in the past in various studies, it hasn't been possible to compare their relative magnitude. By measuring them all, we are able to ascertain for the first time how much of the turnout over-estimate is attributable to each of these three sources of bias.

Hypotheses

We investigate the following three determinants of the over-estimate of turnout in the 2012 ANES.

Over reporting

Social desirability — the tendency of research subjects to present a false or exaggerated self-presentation when asked about socially desirable (or undesirable) attitudes or behaviors[5] — leads some respondents to report being registered to vote and turning out to vote when in fact they are not. In-person interviewing may exacerbate social-desirability effects (Tourangeau and Smith 1996; Tourangeau and Yan 2007). The literature distinguishes between social desirability as a trait of a given individual (e.g., a higher than average need for social approval) and as a response to a specific survey item. Both mechanisms are plausible drivers of overreports of registration and turnout in ANES. Having been asked for opinions on an array of political issues, it is conceivable that some respondents over-report voter registration and turnout, providing responses consistent with the presentation of self offered up through that part of the survey. Interactions between social desirability and non-response bias are quite plausible: those who consent to be interviewed may well inhabit a social milieu where civic engagement and voting are highly valued and likely to overreport voting (Silver, Anderson and Abramson 1986; Bernstein, Chadha and Montjoy 2001).[6]

Over-reporting is the most studied driver of the ANES overestimate of turnout, facilitated by ANES' vote validation studies in 1964 (Clausen 1968) and in seven election years between 1976 and 1990, checking respondents' self-reports of registration and turnout against official reports compiled by state and local election officials; see Traugott (1989).[7] Belli, Traugott and Beckmann (2001) pool the ANES vote

[5]See the review in Krumpal (2013).

[6]This may help account for the pattern noted by Burden (2000), in which the ANES turnout overestimate is negatively correlated with the ANES response rate.

[7]Although validating survey respondents' self-reports of voter turnout predates ANES (e.g., Parry and Crossley 1950).

validation studies, reporting an average overreport of 10.2 percentage points, and an average underreport of 0.7 percentage points. A large literature focuses on the determinants of over-reporting, and how rates of over-reporting vary across demographic groups[8] but this variation is not our focus here. Ansolabehere and Hersh (2012) in particular, are able to use the larger sample size of the Cooperative Congressional Election Study to describe the demographic and political characteristics with greater nuisance than is possible with the smaller-sample ANES. We would point readers interested in the question of who over-reports to that paper.

ANES uses question-wording and response options that attempt to reduce the risk of the respondent providing a socially desirable over-report (e.g., Abelson, Loftus and Greenwald 1992; Belli et al. 1999).[9] In the 2012 ANES, respondents were first asked about their voter registration via a question about the address at which they are registered, with "not currently registered" as one of the response options. Only respondents who report being registered are asked if they voted.[10]

Until recently, vote validation was extremely labor intensive, especially for a survey project with almost complete national coverage like ANES. But the advent of national data bases on voter registration and turnout has made it much easier to

[8] e.g., Traugott and Katosh (1979); Hill and Hurley (1984); Silver, Anderson and Abramson (1986); Anderson and Silver (1986); Kanazawa (2005); Ansolabehere and Hersh (2012).

[9] In 2012 ANES respondents were asked: "In talking to people about elections, we often find that a lot of people were not able to vote because they weren't registered, they were sick, or they just didn't have time. Which of the following statements best describes you?" (1) I did not vote (in the election this November); (2) I thought about voting this time, but didn't; (3) I usually vote, but didn't this time; or (4) I am sure I voted. The long question stem listing reasons why people don't turn out has appeared on ANES studies since 1952. The three response options applicable to non-voting — giving non-voting respondents a set of "softer" options than the binary, "yes" or "no" responses — have been used by ANES since 2000.

[10] Prior to 2012 ANES respondents were only asked about registration *after* the turnout question. A plausible conjecture is that part of the ANES over-report of turnout was due to non-registrants being presented with the turnout question, at which point social-desirability pressures helped generate an over-report. In 2012, the ANES reversed this procedure, asking about registration before turnout.

validate self-reports from survey respondents (Ansolabehere and Hersh 2010). Moreover, the responses from the ANES registration battery (e.g., alternate versions of the respondent's name and place of registration) and help us locate respondents in these national data bases. We present the results of validating the registration and turnout self-reports from the 2012 ANES, below.

Non-response bias

Although face-to-face ANES 2012 respondents are located and contacted via random sampling, sampled individuals choose to take the survey or not. Non-response bias with respect to voter turnout will arise if respondents (or, *compliers*, and we will use the terms interchangeably) are more likely to turn out than non-respondents (*non-compliers*). This is entirely plausible: compliers may well agree to participate in the survey for essentially the same reasons that make them more likely to turn out to vote than non-compliers.[11] Reviews of the literatures on survey participation (Groves and Couper 1998; Dillman et al. 2002; National Research Council 2013) and political participation (Verba, Schlozman and Brady 1995; Blais 2000) point to an overlapping set of motivations and attributes driving both behaviors: e.g., civic-mindedness, social capital, efficacy, and interest in politics.[12]

Non-response bias can also result from the process by which interviewers are able to find respondents, say, if contact with respondents is correlated with registration and turnout. Factors predicting contact include living in easily reached households:

[11] Groves (2006) refers to this as the "common cause" model of survey non-response, noting that in general, a correlation between a variable of interest and survey participation means that nonresponse is non-ignorable (Little and Rubin 2002).

[12] ANES does attempt to reduce the correlation between survey participation and interest in politics by using a cover name for the survey that omits references to politics or elections in the pre-mailers sent to sampled households, and in any subsequent correspondence with respondents.

e.g., stand alone, single household dwellings are usually easier to reach than apartments; having just one place of residence; working regular hours close to the sampled dwelling or otherwise spending significant periods of time at the sampled dwelling (e.g., retirees), and the number of people residing at the household; see Brehm (1993, ch3) or the reviews in Groves and Couper (1998, ch5) and Groves (2006). Most of these characteristics tend to describe a population of surveys responders that skews whiter, older, and more suburban, and hence more likely to be registered to vote and to turn out than non-responders.

In addition, the two-wave, pre/post-election design of the ANES may exacerbate non-response bias via sample attrition. The interval between the two interviews gives non-response mechanisms a second opportunity to intervene. One of the stronger predictors of attrition from the post-election wave of ANES is interest in politics(Olson and Witt 2011). Respondents agreeing to the post-election interview are likely to have higher rates of turnout than the full sample.

Clausen's seminal (1968) study of the 1964 ANES includes an admittedly imprecise estimate of the effect of non-response, restricted to pre-election wave respondents who do not respond to the post-election study; had they been interviewed, Clausen estimates that post-election non-respondents would have reported a turnout rate 10 percentage points below the rate of post-election respondents. Brehm (1993, 135-139) adjusts for non-response in the 1986 and 1988 ANES studies, estimating models of validated turnout with selectivity corrections, concluding that as much as 18 percentage points of the turnout overestimate in the 1986 ANES study might be due to non-response bias. Burden (2000) notes that the gap between ANES turnout estimates and official estimates has grown as ANES response rates have fallen, conjecturing that non-response bias is a growing component of the turnout over-estimate.

ANES is a GOTV Treatment

Another intriguing possibility is that taking the ANES survey is itself a GOTV treatment, with respondents being stimulated to register and turnout at higher rates than we would have observed otherwise. This "stimulus hypothesis" (Clausen 1968) is part of the folklore of ANES: the (presumably facetious) description of ANES as "the most expensive voter mobilization project in American history" is usually attributed to Warren Miller.

There is some plausibility to this argument. For many respondents, the ANES pre-election in-person interview — well over an hour long — would be the more intense and one-sided conversations about politics they have ever experienced. Stimulating interest in politics and voter turnout could well be the unintended effects of submitting to this long and intense interview. Hawthorne and interviewer-demand effects are likely at work too (e.g., Persson 2014), perhaps particularly given that respondents are told ANES will be seeking a post-election interview.

Experiments assessing the effects of pre-election interviewing on turnout include work by Kraut and McConahay (1973) and Yalch (1976); both studies found sizeable effects of survey participation on turnout. Mann (2005) reviews these original studies as well as later, more methodologically rigorous, attempts to measure the treatment effect. After discounting studies with methodological problems or excessively small samples, he finds no evidence of a stimulus effect for most surveys. However, Mann primarily examines phone surveys, which differ from the ANES in that they are shorter in duration and require less interaction with the interviewer, both qualities that might lead to a larger treatment effect of the ANES than what is measured on phone surveyes.

A long line of literature in psychology examines the effect of self-prediction on subsequent behaviors and has spurred GOTV studies assessing the efficacy of this survey-like treatment (e.g., Greenwald et al. 1987; Smith, Gerber and Orlich 2003). Nickerson and Rogers (2010) report that self-prediction of turnout in a short phone survey had a two point effect on validated turnout ($t = 1.33$) while an "implementation" treatment — follow-up questions about when would the respondent vote, where would they be coming from, and what they would be doing beforehand — yielded a 4.1 point effect ($t = 2.4$). Green and Gerber (2008) conclude that in-person canvassing is an especially efficacious GOTV treatments. These findings suggest that (a) given the ANES survey content (many questions about the upcoming election, the candidates, and voting intentions) and (b) the relatively intrusive nature of the ANES in-person interview, we should expect to see participation in the ANES boost turnout.

The ANES 2012 in-person survey

Our analysis focuses on the face-to-face component of the 2012 ANES.[13] Completed, in-person interviews were obtained in 2,054 households, or 28.1% of the 7,298 sampled households. All sampled households were mailed an advance letter. An initial screening interview generates a roster of adult, U.S. citizens residing at the household; a respondent is randomly selected for interviewing in households with more than one eligible respondent. The compliance status of sample households varies. Most house-

[13]The 2012 ANES also included a web component, drawn from the (pre-recruited) panel of a well-known, web-based survey research company. The names and addresses of web respondents were not available to us; doing so would constitute a violation of the assurances made to panelists by the survey research company. For this reason ANES opted not to validate the registration and turnodut self-reports of web respondents.

holds generating a refusal refused screening and rostering. In a small fraction of cases the field interviewer recorded that the sampled dwelling was vacant or did not exist. The distinction among these types of refusals is relevant when we estimate the causal effects of study participation, below. Further details on sampling procedures for the in-person arm of the study appear in the Appendix.

Respondents were asked for the name by which they are registered to vote (asked in the pre-election interview and repeated on the post-election interview, if necessary) and critically, the name to whom the respondent incentive payment should be mailed. Virtually all of the ANES face-to-face respondents provided their name or a useable version of it (2,006 out of 2,054). In households where no interview was conducted ANES did not acquire names or registration details. Details on the various forms of non-response and non-contact appear in the Appendix.

Matching Respondents to Voter Files

For compliers (respondents), validating self-reports of registration and turnout is greatly facilitated by having their name, a residential address from the USPS CDS file (the sample address), and responses to questions about the name under which they are registered, where they are registered, when they were born and so on. This information from compliers was sent to voter file vendors for matching. Matches were obtained for 1,717 of the 2,054 ANES face-to-face respondents in the databases of these vendors(83.6%), or 1,693 of the 2,006 respondents providing a useable version of their name(84.4%).

Because our voter file vendors have access to all of the voter file records for the

	Date of Birth Match	Birth Year Match	Mismatch	Missing Data	Total
Matched to Sampled Address	1039	235	7	103	1384
Matched to Alternate Reg Address	70	24	0	2	96
Matched in Same Zip+4	1	0	0	0	1
Matched to Same Street	19	3	0	0	22
Matched in Same City	59	14	0	19	92
Found Elsewhere	74	18	0	30	122
Total	1262	294	7	154	1717

Table 1: Type of match among ANES 2012 respondents. The name of the voter or commercial record always matches the name given to ANES by the respondent.

country (and usually non-voterfile records from commercial data houses as well), we are able to match ANES respondents even when birthdate or address is mismatched or missing. We summarize the ways that we match the ANES respondents in Table 1. Only cases where the name on the record from our list vendor matches the name the respondent gave to ANES are treated as matches, though rough matches (like matching Kim to Kimberly) are allowed.

In some cases, state voter files only report the age of a voter, but not their complete date of birth. In those cases, we classified a match as a "Birth Year Match" if the age from the voter file was within 2 years of the age implied by the respondent's self-reported date of birth. Unfortunately, in some cases no age or date of birth information was available either from the respondent or (more commonly) from the voter files. Those cases are reported as mismatches.

In a surprising number of instances (237 or 11.5% of respondent matches), the respondent was found at an address other than the sampled address or an alternate registration address provided by the respondent. However, we are typically able to verify that the match was correct because of a date of birth match or a birth year match. In cases with missing birth date or year (156 matched cases), we accept the vendor's match, even in the absence of an address match (50 matched cases). Voter file vendors retain individuals' address histories (Ansolabehere and Hersh 2012); this helps them find respondents at addresses other than the sample address, say, in the case of respondents who move between the ANES interview and before we sent our data to the voter file vendors for matching (the first batch of matches was obtained in March 2013). Another explanation may be old registration records: e.g., if the respondent moved to the sample address without updating their voter registration information and forgot to provide the information to ANES, we might be matching them at an old address.

One final possibility is that ANES field interviewers are conducting interviews with people who do not reside at the sampled address. While we cannot categorically rule out this possibility, the small number of people found at another address but in the same zip+4 (1 respondent) or on the same street (22 respondents) suggests that the ANES rarely interviews neighbors when they cannot get an interview at the sampled address. This phenomenon is not geographically concentrated nor more common at the end of the pre-election field period, which suggests that it is not the result of desperattion for completed interviews or a particularly negligent or fraudulent interviewer.

Validating Respondents' Self-Reports of Turnout

Table 2 shows the distribution of validated turnout conditional on reported turnout (top table) and the distribution of reported turnout conditional on validated turnout (lower table). These estimates are generated with ANES post-stratification weights applied, so as to let us interpret these estimates as population estimates and to assess the magnitude of the over-reporting as source of bias in the ANES turnout estimate. Of the 71% of respondents reporting that they turned out, we can validate turnout for 85.2%. Known *over-reporters* constitute 8.7% of the 71% claiming to have turned out, or 6.2% of the (weighted) data.[14] For just 6.1% of the 71% of respondents claiming to have voted can we *not* find a matching record among the voter files.

		Validated Turnout		
Self report	All	Not voted	Unknown	Voted
No post-election IV	5.7	31.5	13.5	55.0
Not voted	11.9	72.1	22.9	4.9
Not voted, unregistered	10.9	58.3	40.5	1.2
Unknown	0.6	64.5	7.5	28.0
Voted	70.8	8.1	6.2	85.7

	Validated Turnout		
Self report	Not voted	Unknown	Voted
No post-election IV	7.9	6.3	4.9
Not voted	37.6	22.1	0.9
Not voted, unregistered	27.9	35.8	0.2
Unknown	1.6	0.3	0.2
Voted	25.1	35.5	93.8
All	22.9	12.4	64.7

Table 2: Validated turnout (columns) and reported turnout (rows), 2012 U.S. general election, ANES 2012 face-to-face respondents. Weights applied. The top table shows the distribution of validated turnout conditional on reported turnout (the row entries sum to 100, absent rounding error, with the column labelled "All" showing the marginal distribution of the self-reports of turnout. "Unknown" self-reports arise from refusals and "don't know" responses. The lower table shows the distribution of reported turnout conditional on validated turnout; the columns sum to 100, absent rounding error, with the row labelled "All" showing the marginal distribution of validated turnout.

Most respondents skipping the post-election wave of the study generate missing

[14] If we restrict the analysis to matches to voter files only (putting aside non-voters matched to consumer files), the over-reporting rate drops to 5.8%, similar to the 6.5% overreport rate found by matching ANES 2008-09 panel respondents to voter files in six states (Berent, Krosnick and Lupia 2016, Table 6).

data with respect to a turnout self-report (top row of Table 2). But over half of these respondents (56.1%) are found to have voted.

We also verify that reports of *not* voting are generally correct. Respondents reporting not voting (11.8%) and or those not asked about voting (because they reported not being registered, 10.8%) are generally validated non-voters; we verify this self-report for this group of respondents about two-thirds of the time. In a handful of cases we find a record of the respondent turning out in 2012 when the respondent claims to have not voted: these *under-reporters* constitute 0.7% of the (weighted) data and comprise just 12 respondents.[15] Over-reporting is more than twenty times as prevalent as under-reporting. The difference between weighted, verified over-reporting and under-reporting is 6.2 - 0.7 or 5.5 percentage points, which accounts for a hefty proportion of the ANES over-estimate of voter turnout.

Among respondents who we match to a voter registration records, we can validate the corresponding turnout self-report in 93.2% of cases (a weighted percentage). This is similar to the 93.5% rate of "consistent" self-reports of turnout in the subset of the ANES 2008-09 panel study, matched by Berent, Krosnick and Lupia (2016) to voter files in six states.

Overall, the weighted, validated turnout rate for the 2012 ANES face-to-face study is 64.6% (lower right of Table 2), much closer to the VEP estimate of 58.2% reported by McDonald than the self-reported rate of 71%, and not too far from the 61.8% CPS estimate. But the validated turnout rate of 64.6% assumes that *all* respondents

[15]We carefully scrutinized this small number of respondents who either report not voting or not being registered, contrary to what the voter files suggest. After laborious, manual comparisons of identifying information from the ANES survey and the voter files for these cases we are confident that we have matched the correct individual. We can not rule out the possibility that the respondents actually misreported, or that the official records contain errors. An over-reporting rate of less than 1% in official turnout records does not strike us as implausible, nor especially large.

we were unable to match did not turn out. As Berent, Krosnick and Lupia (2016) highlight, errors in the voter files or in information provided by respondents can cause a failure to match a respondent to their registration record, negatively biasing the validated turnout rate. Since turnout in this group of unmatched respondents is (almost surely) not zero, then the validated turnout rate must be revised upwards. Only 35.5% of this group claims to have turned out and most report either not voting (22.1%) or being unregistered (35.8%), but nonetheless turnout is non-zero for unmatched respondents. The "headline", weighted, validated turnout rate of 64.6% reported in the lower right of Table 2 is a lower bound on the true, weighted turnout rate of ANES face-to-face respondents, which is probably another 4.3 percentage points higher, or 68.9%. [16] Relative to the McDonald estimate of 58.2%, this still leaves almost 11 points of bias to be accounted for.

Validation with Stringent Matching

As discussed in Table 1, some ANES respondents were matched to a voterfile record different from their residential or registration addresses. In table 2, we treated these respondents as having been validly matched, and proceeded to validate their turnout self-reports against their voterfile record. Respondents that couldn't be matched to a commercial voterfile record were treated as unknown.

In this section, we reproduce table 2, but this time, we move the 237 respondents that were matched to a different address to the unknown column. This leaves only

[16] If respondents with an "unknown" validated vote status are being truthful in their survey self-reports, then we obtain an overall, adjusted, validated turnout rate is $64.6 + .355 \times 12.2 = 68.9\%$. Alternatively, if we assume that those with an "unknown" validated turnout status have the same validated turnout rate as other respondents conditional on their self-report, we again obtain the adjusted, weighted, validated turnout rate of 68.9%.

the address-matched respondents, for whom we are most confident in our match, in the two known validation columns. Table 3 shows that the over-reporting rate decreased modestly, by about 1 percentage point, with those over-reporters moving to the unknown column, while the under-reporting rate remained unchanged.

This change might suggest that the non-address-matched respondents have more false positive matches (since a higher proportion of them have a mismatch between their self-reported and validated turnout behavior). But it could also be the case that respondents that have a more complicated registration status (being registered at an address other than the one at which they live) are more likely to over-report voting. In either case, the over-report rate is still substantial, indicating that over-reporting is a substantial contributor to the over-estimate of voter turnout on the 2012 ANES, even among those respondents for whom we have the best voterfile match.

Self report	All	Validated Turnout Not voted	Unknown	Voted
No post-election IV	5.7	24.6	20.4	55.0
Not voted	11.9	63.2	31.8	4.9
Not voted, unregistered	10.9	58.3	40.5	1.2
Unknown	0.6	64.5	7.5	28.0
Voted	70.8	7.1	8.6	84.2

Self report	Validated Turnout Not voted	Unknown	Voted
No post-election IV	6.8	7.5	5.0
Not voted	36.4	24.4	0.9
Not voted, unregistered	30.8	28.5	0.2
Unknown	1.7	0.3	0.2
Voted	24.3	39.3	93.7
All	22.9	12.4	64.7

Table 3: Validated turnout (columns) and reported turnout (rows), 2012 U.S. general election, ANES 2012 face-to-face respondents. Weights applied. Respondents that were matched to an address other than their residential or registration address are assigned to the unknown column. The top table shows the distribution of validated turnout conditional on reported turnout (the row entries sum to 100, absent rounding error, with the column labelled "All" showing the marginal distribution of the self-reports of turnout. "Unknown" self-reports arise from refusals and "don't know" responses. The lower table shows the distribution of reported turnout conditional on validated turnout; the columns sum to 100, absent rounding error, with the row labelled "All" showing the marginal distribution of validated turnout.

Non-response bias: differences between compliers and non-compliers

We now consider non-response bias, the bias in the ANES turnout estimated induced by higher rates of non-response amongst non-voters. Like voting, participating in the ANES is costly, involving time and disruption of daily routines. ANES tries to mitigate this problem by repeated attempts to make contact with sampled households, offering flexibility with scheduling interviews and offering monetary incentives. Nonetheless, there remain differences in turnout between ANES respondents and non-compliers even on pre-survey indicators, such as turnout in previous elections.

For this analysis, we employ records from voterfile vendors spanning compliers, people in complier households and people in non-complying households.[17]

One difficulty in taking records from administrative files is the false listing of individuals no longer in residence at an address ("deadwood"). We address this problem by performing our analysis at the level of the address, so that any deadwood records appear in both complier and non-complier households. Further, we note that there's a slight imbalance in the number of people listed at each address, with an average of 2.2 people per address in non-complier households and 2.5 people per address in complier households. To correct for this, we run our comparisons on a matched dataset, matching complier households to non-complier households of the same size, so that the amount of deadwood is roughly the same on both sides of the comparison. We also limit our analysis to households where 3 or fewer people

[17] We supplied the 7,298 ANES sampled addresses to two voterfile vendors, asking them to return all records listed as living or registered at the sampled address. They returned at least one matching record for 92.3% of non-complying households and 96.7% of complying households.

are listed, so that large households, which are very likely to contain deadwood, are excluded from the analysis. Finally, we exclude cases where an ANES interviewer identified the household as ineligible for an interview.[18]

We compute average rates of voter registration and turnout for even-year, general elections from 2006 through 2012 by household. We limit our analysis to main sample cases (households that weren't part of the minority over-samples) because in the minority oversample tracts, non-complying households may not be black or hispanic, making them ineligible for the survey.

One concern in this analysis is that differences between complier and non-compliers may be driven by differences in the quality of their administrative records, rather than actual differences in behavior. Ideally, we would limit our analysis to people in non-complier households who might actually take the survey. But since it is impossible to know who actually lives in a non-complier household (because in most cases they were not rostered), we settle for using data about the household as it exists on commercial voterfiles. So that we are balanced about our comparisons, we use the commercial voterfile data for both complier and non-complier households. Further, all differences are computed on a subsample that has been balanced on household size, so that the complier and non-complier groups both have the same distribution of the number of voterfile records across households. This balance should reduce or eliminate differences in the number of deadwood records included in the analysis. To account for uncertainty induced by our use of a matching algorithm, we report

[18]Specifically, we exclude sampled addresses assigned the following dispositions: "Address does not exist", "Out of sample area", "Address does not have a permanently occupied household", "No adult citizen", "OS only, no Hispanic/AA." These dispositions indicate that there is no one residing at the sampled household eligible to take the survey, and as such, the household has neither complied nor failed to do so. We also exclude 270 "Other non-contact cases" that were sent an advance letter, but were never assigned to a field interviewer, as these households similarly never had the chance to be interviewed, nor refuse to do so.

adjusted standard errors (Abadie and Imbens 2006).

Figure 1 shows that in the two later pre-survey elections, 2008 and 2010, complier and non-complier households show a roughly 3.5 percentage point difference in turnout (distinguishable from zero at conventional levels of statistical significance). Since Presidential election turnout patterns will be more similar to each other than to turnout patterns in midterm elections, the 2008 election is our best guide to non-response bias in 2012. From this, we infer that the differences in turnout rates absent any 2012-election specific effects would be about 3.5 percentage points (± 3.5 p.p.).

Moving to the 2012 election, the difference in turnout rates are about 2 percentage points larger in 2012 than in 2008. This difference suggests that additional forces may be driving higher rates of turnout for compliers relative to non-compliers in the 2012 election. One possibility is that compliers are especially interested in the 2012 election, relative to earlier elections. That is, there may be some portion of the complier population that would have voted at non-complier rates in 2012, were it not for the fact that they were especially stimulated by the 2012 election. Given that both the 2008 and 2012 general elections featured contests between Barack Obama and Republican nominees associated with the establishment wing of the party, it seems unlikely that nearly 2% of compliers could be stimulated relative to 2008 levels of turnout. Instead, the stimulus of the survey itself seems the most likely explanation for the growth in complier turnout.

How does this non-response bias arise? An interesting hint is the 3.5 point difference in registration rates between complying and non-complying households. This differences suggest that unregistered people are harder to contact and less likely to assent to participate.

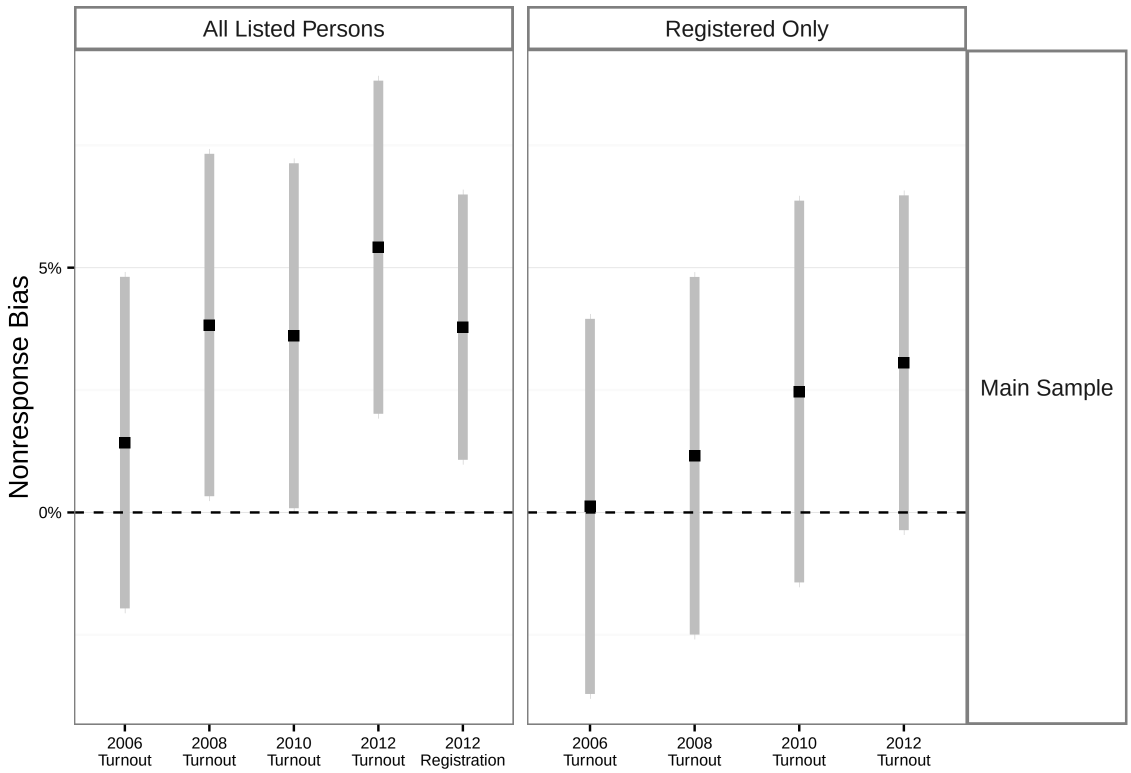

Figure 1: Estimates of the diffence in turnout and registration between people in complier households and non-complier households for main sample cases only and the full sample. Estimates are also broken out looking at registered voters only. Among registered voters, no statistically significant differences exist in the pre-treatment elections. Shaded vertical bars indicate 95% confidence intervals.

The panel on the right of Figure 1 show differences in turnout rates in our matched sample conditional on being registered. For pre-2012 elections, there are much smaller differences between the complier and non-complier groups, with statistically significant differences only manifesting in the 2012 election, where turnout occurred after recruitment by ANES. This suggests that the differences in turnout for earlier elections are largely attributable to a failure to recruit survey-takers from the unregistered population.

To summarize, non-response bias is responsible for roughly 3.5 percentage points of the bias observed in the ANES estimate of 2012 turnout. A key driver of this bias lies in failing to get enough unregistered people to take the ANES survey. Solutions would include checking the registration status of individuals thought to reside at sampled households ahead of field operations, and tailoring recruitment strategies and field efforts accordingly or stratifying the ANES sample on the registration status of sampled households. As is so often the case with surveys, it is far easier and more elegant to deal with likely sources of bias at the design and sampling stages of the survey, versus engaging in the laborious forensic exercise reported here.

The causal effect of survey participation

We estimate the effect of survey participation on turnout by exploiting the random selection of individuals within sampled households, allowing us to estimate an intent-to-treat effect of taking the ANES survey. Individuals are chosen for an ANES interview in a multi-stage process, culminating in random, within-household selection in households that are successfully rostered and have more than one, eligible individual residing at the household (American National Election Studies 2014, 29). No within-household selection is necessary in households with just one eligible respondent. Interviews were not conducted with any other member of the household. Once a person was randomly selected from within the household, field interviewers were usually successful in completing an interview. Of 2,608 individuals selected for interviewing, 2,054 completed the ANES pre-election interview.

This procedure lets us identify the causal effect of being selected for an ANES interview on turnout, conditional on (a) being in a household with more than one eligible adult citizen and (b) multiple members of the household being found on a commercial voter file. Since the causal effect is estimated by comparing the turnout of a selected member of a household to another member of their same household, the estimated effect is net of any within-household spillovers.

In rostering sampled households, field interviewers attempt to obtain the names, ages and gender of the household's adult residents. Matching respondents (who supply a full name and date of birth) is much more straightforward than matching individuals with the extremely limited set of personally identifying information obtained from household rostering. Usually only first names were recorded, but sometimes initials or signifiers like "husband" and "adult male" were reported. Using the list of people known to live at that address from a commercial voter file (as it existed at the start of the survey field period, in early September 2012), we selected the best match for each rostered person. In cases where there was no good match or the match was thought to be unreliable, no match was recorded, and the case was discarded. Of the 1,484 enumerated households that contained more than one eligible person, matches to commercial voter files for two or more persons were made in 687 households. 1,510 people were matched, of which 1,448 matched on name, age and address, 31 matched on their initials and address and the remaining 31 cases were matched on their gender and/or age and address.

Not all persons randomly selected for an interview comply. Of the 687 matched, multiple-person households in this analysis, the randomly chosen household member took the survey (the treatment) in 531 households, yielding a compliance rate of 77.3%. In this context, non-compliance (not taking the survey after successful contact, household rostering and within-household random selection) is typically a "hard refusal" (76 households) or the randomly selected household member being "not available" (63 households).

We compare the observed difference in turnout of the sampled individual to that of a different randomly selected person in their household. We compute this difference over 1,000 randomizations of the control householder and compute an average intent-to-treat (ITT)effect as the average difference in turnout rates between the selected person and the randomly selected control case. Note that in households with only two rostered adults, the same control case is chosen in each iteration. To characterize the null distribution of this procedure, we used the same approach, but this time completely randomizing selection to the survey over 1,000 iterations. In this randmomization inference procedure, sometimes the selected person will be chosen as the control case and non-selected individuals will be selected to be the treatment case. We use this null distribution to compute a p-value for the intent-to-treat effect. A full technical treatment of this procedure appears in the appendix.

To make inferences about the complier-average causal effect (CACE), we take each bootstrap estimate of $\widehat{ITT}$ and divide it by a simulated compliance rate. The compliance rate is a generated draws from a Beta($\alpha = 531, \beta = 156$) density for the full sample comparisons and Beta($\alpha = 410, \beta = 121$) for the main sample comparisons.[19] We report the mean and 2.5th and 97.5th percentiles of the bootstrap draws for the CACE.

Election	Sample	$\widehat{ITT}$	p	CACE	2.5%ile	97.5%ile
2012	Full Sample	2.6	.11	3.3	-1.0	7.5
2012	Main Sample	3.5	.04	4.5	0.0	9.1
2010	Full Sample	0.4	.75	0.5	-4.2	5.4
2010	Main Sample	0.9	.63	1.1	-4.4	6.5
2008	Full Sample	0.5	.73	0.6	-4.2	5.7
2008	Main Sample	1.0	.61	1.3	-4.0	6.8
2006	Full Sample	0.8	.64	1.1	-4.0	5.7
2006	Main Sample	0.9	.63	1.2	-4.5	6.4

Table 4: Statistics for each election on both the full sample and main sample only. p-values refer to the estimated intent-to-treat effects. The 2.5%ile and 97.5%ile columns refer to those quantiles of the CATE.

The estimated treatment effects with respect to 2012 turnout are large. The average treatment effect for compliers is 3.3% for the full set of cases, rising to 4.5% among main sample cases only. $\widehat{ITT}$ is distinguishable from zero at the .05 level for main sample cases; for the full sample, the intent-to-treat effect is significant at the .05 level for a one-tailed test only, placing zero weight on the possibility that the survey would have a demobilizing effect. The placebo effects are all close to 0; we fail to reject the null of zero effects in the three placebo elections ($p = .84$ for the full sample cases). The observed pattern of treatment effects fits with the assumption that turnout in the placebo/pre-2012 elections is unrelated to selection for an interview. While we are primarily concerned with the treatment effect of taking the survey on all ANES households, it might be more appropriate to analyze only the main sample households, since in the minority oversample tracts, white residents

[19] i.e., α is chosen to be equal to the number of completed interviews among matched respondents in the sample and β equal to the number of matched and selected respondents that did not finish the interview. This is equivalent to drawing from the posterior probability density for r under an improper Beta$(0, 0)$ prior.

would be ineligible to take the survey while minority residents of the same house would be eligible, confounding our estimate of the treatment effect.

These estimates comport with turnout effects observed in field experiments designed to stimulate mobilization (Arceneaux and Nickerson 2009). But, again, what is the nature of the effect? One possibility is that we are observing a classic Hawthorne effect: respondents see that their political behavior is being studied by ANES, and alter their behavior to comply with social norms favoring voting. Another possibility is that respondents feel social pressure from the particular interviewer rather than the study per se. Others might incorrectly believe that turning out is a prerequisite for being eligible for the post-election survey and the accompanying incentive payment. Disambiguating these possibilities will require a design and instrumentation not presently utilized by ANES.

We stress that caution is warranted in accepting the estimates from this within-household analysis. We did not match all enumerated individuals and unmatched individuals are less likely to have voted, because they may not be registered at their current address. However, the direction of the bias that excluding these unmatched individuals induce is ambiguous. Recall that the ITT effect is measured as a difference between the rates of voting for those selected for an interview and those not selected. Excluding unmatched individuals raises our estimate of both voting rates, since the unmatched are less likely to have voted. Since the bias for both quantities is positive and difficult to quantify, the bias of their difference is unknown.

Conclusion

Three distinct mechanisms account for the bias in the ANES estimate of turnout: over-reporting, non-response bias and treatment effects. Over-reporting is the single largest source of bias in the 2012 ANES face-to-face survey, responsible for 6.2 percentage points of bias in the ANES turnout estimate. Non-response bias and the treatment effect of taking the survey have smaller but still considerable effects of roughly 4 and 3 percentage points each. Together, our estimates account for 13 percentage points of bias in the turnout estimate from the 2012 face-to-face components of ANES.

By estimating these quantities in the same survey, we are able to characterize for the first time how much each source of bias impacts the overall estimate. Happily, the largest source of bias, over-reporting, is also the easiest source of bias to eliminate through better measurement. By matching to voter files, respondents' self-reports of their turnout behavior can be compared against official records. This comparison makes clear that while the vast majority of respondents correctly report their own behavior, a few report voting when they did not.

Eradicating treatment effects and over-reporting seems hard to do, at least in the face-to-face mode. It is plausible that a self-complete survey mode — taking the interviewer out of the picture — might reduce social desirability, and thus reduce over-reporting of turnout. Likewise, it is plausible that there are smaller treatment effects when the survey is self-administered rather than administered in-person.

The are several, important consequences of the over-estimate phenomenon that we do not explore here. For instance, if voter turnout is either over-reported, over-estimated due to non-response bias, or stimulated by survey participation, then it is quite likely that other political behaviors and attitudes are similarly affected. Examples include voter registration, engagement with candidates and campaigns, political interest and political efficacy. Note that for some of the variables we lack corresponding population benchmarks or micro-data that could be used for validating or adjusting respondent self-reports.

While producing decompositions of biases in those measures will be difficult, this paper lays out a guide by which bias decompositions could be computed. Validating self-reports of campaign contact seems particularly promising. Information about campaign contacts has become increasingly centralized by the parties and commercial data warehouses like Catalist (Issenberg 2012). Thus in principle, validation data should be available. From there, the same set of matching procedures and comparisons could be run, and a decomposition of the bias of self-reported campaign contact could be produced.

Finally, our analysis has implications for the conduct of election surveys. The kind of data collection we have performed, taking sample addresses and respondent information to voter file vendors and using multiple matching methods to identify respondents, should be standard practice and integrated into the design and implementation of surveys, especially public-use, canonical election surveys like ANES. Second, we ought to consider sampling designs that use the information on voter files, say stratifying the ANES sample by whether sampled households contain registered or active voters or not, or are of indeterminate status; this is part of the sampling plan for the self-complete/Internet arm of the 2016 ANES. In face-to-face interviewing, field operations and interviewer effort could be more efficiently deployed to ensure that these households yield complete interviews, helping to reduce non-response bias (e.g., Czajka 2013).

References

Abadie, Alberto and Guido W Imbens. 2006. "Large sample properties of matching estimators for average treatment effects." *econometrica* 74(1):235–267.

Abelson, Robert P., Elizabeth F. Loftus and Anthony G. Greenwald. 1992. Attempts to Improve the Accuracy of Self-Reports of Voting. In *Questions about Questions*, ed. Judith M. Tanur. New York: Russell Sage Foundation pp. 138–53.

Aldrich, John H. and Kathleen M. McGraw. 2011. *Improving Public Opinion Surveys: Interdisciplinary Innovation and the American National Election Studies*. Princeton, New Jersey: Princeton University Press.

American National Election Studies. 2014. *User's Guide and Codebook for the ANES 2012 Time Series Study*. Ann Arbor, Michigan and Stanford, California: University of Michigan and Stanford University.

URL: http://electionstudies.org/studypages/anes_timeseries_2012/anes_timeseries_2012_userguidecodebook.pdf

Anderson, Barbara A. and Brian D. Silver. 1986. "Measurement and Mismeasurement of the Validity of the Self-Reported Vote." *American Journal of Political Science* 30(4):771–785.

URL: http://www.jstor.org/stable/2111272

Ansolabehere, Stephen and Eitan Hersh. 2010. "The quality of voter registration records: A state-by-state analysis." *Report, Caltech/MIT Voting Technology Project* .

URL: http://v-ads-web5.ads.caltech.edu/content/quality-voter-registration-records-state-state-analysis

Ansolabehere, Stephen and Eitan Hersh. 2012. "Validation: What big data reveal about survey misreporting and the real electorate." *Political Analysis* 20(4):437–459.

Arceneaux, Kevin and David W Nickerson. 2009. "Who Is Mobilized to Vote? A Re-Analysis of 11 Field Experiments." *American Journal of Political Science* 53(1):1–16.

Belli, Robert F., Michael W. Traugott, Margaret Young and Katherine A. McGonagle. 1999. "Reducing Vote Overreporting in Surveys: Social Desirablity, Memory Failure, and Source Monitoring." *Public Opinion Quarterly* 63(1):90–108.
URL: http://poq.oxfordjournals.org/content/63/1/90.short

Belli, Robert F., Michael W. Traugott and Matthew N. Beckmann. 2001. "What leads to voting overreports? Contrasts of overreporters to validated voters and admitted nonvoters in the American National Election Studies." *Journal of Official Statistics* 17(4):479.

Berent, Matthew K, Jon A Krosnick and Arthur Lupia. 2016. "Measuring Voter Registration and Turnout in Surveys Do Official Government Records Yield More Accurate Assessments?" *Public Opinion Quarterly* p. nfw021.

Bernstein, Robert, Anita Chadha and Robert Montjoy. 2001. "Overreporting Voting: Why It Happens and Why It Matters." *Public Opinion Quarterly* 65(1):22–44.
URL: http://poq.oxfordjournals.org/content/65/1/22.abstract

Blais, André. 2000. *To vote or not to vote?: The merits and limits of rational choice theory.* University of Pittsburgh Press.

Brehm, John. 1993. *The Phantom Respondents: Opinion Surveys and Political Representation*. Ann Arbor, Michigan: University of Michigan Press.

Burden, Barry C. 2000. "Voter turnout and the national election studies." *Political Analysis* 8(4):389–398.

Campbell, Angus, Warren E Miller and Philip E Converse. 1960. *The American voter*. Wiley.

Clausen, Aage R. 1968. "Response validity: Vote report." *Public Opinion Quarterly* pp. 588–606.

Czajka, John L. 2013. "Can Administrative Records Be Used to Reduce Nonresponse Bias?" *The Annals of the American Academy of Political and Social Science* 645(1):171–184.
URL: http://ann.sagepub.com/content/645/1/171.abstract

Dillman, Don A., John L. Eltinge, Robert M. Groves and Roderick J.A. Little. 2002. Survey nonresponse in design, data collection, and analysis. In *Survey Nonresponse*, ed. Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J. A. Little. New York: John Wiley & Sons New York pp. 3–26.

Green, Donald P and Alan S Gerber. 2008. *Get out the vote: How to increase voter turnout*. Washington DC: Brookings Institution Press.

Greenwald, Anthony G., Catherine G. Carnot, Rebecca Beach and Barbara Young. 1987. "Increasing Voting Behavior by Asking People if They Expect to Vote." *Journal of Applied Psychology* 72(315-18).

Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Household

Surveys." *Public Opinion Quarterly* 70(5):646–675.
URL: http://poq.oxfordjournals.org/content/70/5/646.abstract

Groves, Robert M. and M. P. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: Wiley.

Hill, Kim Quaile and Patricia A. Hurley. 1984. "Nonvoters in Voters' clothing: the impact of voting behavior misreporting on voting behavior research." *Social Science Quarterly* 65(1):199–206.
URL: http://www.jstor.org/stable/42861691

Issenberg, Sasha. 2012. *The victory lab: The secret science of winning campaigns*. Random House Digital, Inc.

Kanazawa, Satoshi. 2005. "Who Lies on Surveys, and What Can We Do About It?" *The Journal of Social, Political and Economic Studies* 30(3):361–370.

Kraut, Robert E. and John B. McConahay. 1973. "How being interviewed affects voting: an experiment." *Public Opinion Quarterly* 37(3):398–406.
URL: http://poq.oxfordjournals.org/content/37/3/398.short

Krumpal, Ivar. 2013. "Determinants of social desirability bias in sensitive surveys: a literature review." *Quality and Quantity* 47(4):2025–2047.
URL: http://dx.doi.org/10.1007/s11135-011-9640-9

Little, Roderick J. A. and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data*. Second ed. New York: Wiley.

Mann, Christopher B. 2005. "Unintentional voter mobilization: Does participation in preelection surveys increase voter turnout?" *The ANNALS of the American Academy of Political and Social Science* 601(1):155–168.

McDonald, Michael. 2013. 2012 General Election Turnout Rates. Technical report United States Elections Project, George Mason University.
URL: http://elections.gmu.edu/Turnout_2012G.html

National Research Council. 2013. *Nonresponse in Social Science Surveys: A Research Agenda*. Washington, DC: The National Academies Press. Roger Tourangeau and Thomas J. Plewes (eds.). Panel on a Research Agenda for the Future of Social Science Data Collection, Committee on National Statistics. Division of Behavioral and Social Sciences and Education.

Nickerson, David W. and Todd Rogers. 2010. "Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making." *Psychological Science* 21(2):194–199.

Olson, Kristen and Lindsey Witt. 2011. "Are we keeping the people who used to stay? Changes in correlates of panel survey attrition over time." *Social Science Research* 40(4):1037–1050.

Parry, Hugh J. and Helen M. Crossley. 1950. "Validity of Responses to Survey Questions." *Public Opinion Quarterly* 14(1):61–80.
URL: http://poq.oxfordjournals.org/content/14/1/61.abstract

Persson, Mikael. 2014. "Does Survey Participation Increase Voter Turnout?: Re-examining the Hawthorne Effect in the Swedish National Election Studies." *Political Science Research and Method* pp. 1–11.
URL: doi://10.1017/psrm.2014.8

Silver, Brian D., Barbara A. Anderson and Paul R. Abramson. 1986. "Who Overreports Voting?" *American Political Science Review* 80(2):613–624.
URL: http://www.jstor.org/stable/1958277

Smith, Jennifer K., Alan S. Gerber and Anton Orlich. 2003. "Self-Prophecy Effects and Voter Turnout: An Experimental Replication." *Political Psychology* 24(3):593–604.

URL: http://dx.doi.org/10.1111/0162-895X.00342

Tourangeau, Roger and Ting Yan. 2007. "Sensitive questions in surveys." *Psychological Bulletin* 133(5):859–883.

Tourangeau, Roger and Tom W. Smith. 1996. "Asking sensitive questions: the impact of data collection mode, question format, and question context." *Public Opinion Quarterly* 60(2):275–304.

URL: http://poq.oxfordjournals.org/content/60/2/275.abstract

Traugott, Michael W. and John P. Katosh. 1979. "Response Validity in Surveys of Voting Behavior." *Public Opinion Quarterly* 43(3):359–377.

URL: http://poq.oxfordjournals.org/content/43/3/359.abstract

Traugott, Santa. 1989. *Validating self-reported vote: 1964-1988.* Ann Arbor, Michigan: Institute for Social Research, University of Michigan.

Verba, Sidney, Kay Lehman Schlozman and Henry E. Brady. 1995. *Voice and equality: Civic voluntarism in American politics.* Harvard University Press.

Weisberg, Herbert F. 2009. *The total survey error approach: A guide to the new science of survey research.* University of Chicago Press.

Wolfinger, Raymond E and Steven J Rosenstone. 1980. *Who votes?* Vol. 22 Yale University Press.

Yalch, Richard F. 1976. "Pre-election Interview Effects on Voter Turnout." *Public*

Opinion Quarterly 40(3):331–336.

URL: http://poq.oxfordjournals.org/content/40/3/331.abstract

A The ANES 2012 in-person survey: additional details

Face-to-face respondents to the 2012 ANES were sampled by a multi-stage cluster design. The design sought a minimum of 2,000 completed cases in 125 primary sampling units. A sampled tract in Native American-governed territory was dropped when it became apparent that a request to that territory's IRB for permission to conduct interviews there would not be considered ahead of the field period. Census tracts served as primary sampling units and were selected with probability proportional to population (as estimated by the 2012 Census), within nine strata (Census divisions). Over-samples of Latino and African-American respondents — 300 completed cases each — were also part of the design; census tracts known to have relatively high proportions of Latinos and/or African-Americans were selected for over-sampling. Within census tracts, households were randomly selected from the USPS computerized delivery sequence (CDS) file. In a small number of rural tracts, field enumeration was used to check for CDS under-coverage, resulting in 81 households being added to the sample. Given expectations about the target number of completed interviews and contact, eligibility and response rates, a total of 7,298 addresses were sampled. Field interviewers were trained in two batches, each batch spanning two days, in-person, in Fort Lauderdale, Florida, conducted in the 2nd half of August 2012.

Sampled households were mailed an advance letter with a cover letter introducing the study (using the "cover name" discussed above) with a $2 cash goodwill payment. 71% of the sample households were mailed an advance letter on August 29, 2012, with interviews commencing on September 9. Small releases of additional sample occurred throughout October 2012, with a relatively large release in late October comprising

14% of the sample.

If contact could be made with a sampled household, the interviewer first administered a short "screener" interview, obtaining a listing eligible potential respondents residing at the sampled household. The ANES target frame is adult, U.S. citizens residing in households. Seventeen year-olds who will turn eighteen before the election are part of the ANES frame. The minority over-samples involved screening for at least one adult citizen with the race/ethnicity appropriate to the particular over-sample. If there was more than one eligible potential respondent residing at the dwelling then one was selected randomly. A small number of selected individuals were not interviewed because they were mentally or physically incapable ($n = 32$) or spoke a language other than English or Spanish ($n = 17$).

In a large number of cases the sampled household is never contacted. Field interviewers reported that 6.1% of sampled addresses are unoccupied or vacation houses ($n = 448$). Some addresses are reported to be empty lots, vacant dwellings or new construction, or simply can't be located by the interviewer. Some rural addresses or densely urban addresses are ambiguous or unreliable. Apartment buildings are sometimes well-protected by doormen, or the address is in a gated community. 336 sampled addresses (4.6%) were located but unable to be accessed by the interviewer. Another 84 (1.2%) sampled dwellings were coded as "address does not exist". Other forms of non-contact account for another 1,125 sampled households (16.8% of the sample). All forms of non-contact amount to 32.6% of sampled households.

In some cases a screening interview was successfully conducted, only to discover that there are no eligible potential respondents in residence. No adult citizens were found at 259 (3.5%) of the sampled households. The more stringent eligibility criteria used for the minority oversamples resulted in 1,083 or 14.8% of our sampled households being deemed ineligible (accounting for 35.4% of the households in the minority oversamples). Refusals before screening (12.9%), after screening and within-household random selection (3.3%) was also another common occurrence. In another 3.6% of cases the eligible person selected was never available to be interviewed.

B Randomization Inference for Measuring the Causal Effect of ANES Selection

Over M=1,000 iterations of the following process, we compare the sampled individual's turnout to a randomly selected control case from the respondent's household; let $\mathbf{W}_{\text{obs}}$ denote this observed set of random assignments of household residents to treatment. Households with one person were excluded, and in households of two people, the same control case was selected on each iteration. At each iteration $p = 1, \dots, M$ we compute the difference in turnout rates for the treatment and control groups. This difference, $\widehat{ITT}_p(\mathbf{W}_{\text{obs}})$ is estimate of the intent-to-treat effect of taking the ANES pre-election interview, recalling that 22.7% of the households in this analysis are non-compliers.

A similar procedure is used to characterize the distribution of the intent-to-treat effect under the sharp null hypothesis of no treatment effects for any household. We randomly choose two people from each household, randomly assigning one to be the sampled individual (the treatment condition) and one to control, and then compute their difference in turnout. Each iteration $q = 1, \dots, M$ yields a set of random assignments $\tilde{\mathbf{W}}_q$. We average the within-household differences in turnout under permutation

$\tilde{\mathbf{W}}_q$ to produce a sample from the null distribution of intent-to-treat effects, $ITT(\tilde{\mathbf{W}}_q)$.

We used these two sets of intent-to-treat estimates (one based on comparing the household member randomly selected for treatment to one other control cases in the household, the other based on random within-households pairings) to create 1,000 x 1,000 = 1 million pairs, $\{(\widehat{ITT}_p(\mathbf{W}_{\text{obs}}), \widehat{ITT}_q(\tilde{\mathbf{W}})), \forall\, p, q \in 1, \ldots, M\}$. We estimate (1) the intent-to-treat effect with $\widehat{ITT} = M^{-1} \sum_{p=1}^{M} \widehat{ITT}_p(\mathbf{W}_{\text{obs}})$; and (2) the two-tailed p-value as the proportion of the M^2 pairings in which $|\widehat{ITT}_p(\mathbf{W}_{\text{obs}})| < |\widehat{ITT}_q(\tilde{\mathbf{W}})|$. We implement this procedure for 2012 general election turnout, as well as for three placebo elections: 2010, 2008 and 2006. We compute estimates for main sample cases only and for the full sample.