Working Paper, nes012554.pdf: The Quality of Government Records and Over-estimation of Registration and Turnout in Surveys: Lessons from the 2008 ANES Panel Study's Registration and Turnout Validation Exercises

The proportion of survey respondents who say they turned out to vote in a particular election is often far greater than the proportion of the population who actually voted. Such "over-reporting" of turnout has lead some analysts to conclude that many survey respondents misreport having voted. To get a better handle on why over reporting occurs, some researchers advocate "turnout validation." Turnout validation compares a survey respondent's report of having turned out to official government records for that individual. In 2009 and 2010, the former ANES PIs Jon Krosnick and Arthur Lupia, and Matt Berent gathered and analyzed official government turnout records from several large states. They then compared these records to ANES survey data. A new working paper based on this work shows that survey respondents' reports of their own turnout behavior is more accurate than previously believed.

The Quality of Government Records and Over-estimation of Registration and Turnout in Surveys:

Lessons from the 2008 ANES Panel Study's Registration and Turnout Validation Exercises

Matthew K. Berent and Jon A. Krosnick

Stanford University

Arthur Lupia

University of Michigan

August, 2011

The Quality of Government Records and Over-estimation of Registration and Turnout in Surveys:

Lessons from the 2008 ANES Panel Study's Registration and Turnout Validation Exercises

Preface

Two central variables measured in all American National Election Studies are whether the respondent was registered to vote and whether the respondent voted. All past ANES surveys have measured these variables via conventional self-reports. Yet over the years, concern has been voiced about whether these self-reports are accurate.

Specifically, much concern has been expressed that respondents may feel pressure to claim that they were registered and voted when they did not, because registering to vote and voting are socially desirable. Consistent with this concern is evidence that ANES surveys (and many other high-quality surveys) yielded rates of respondent registration and turnout that were notably higher than official government reports of those rates.

In the 1980s, the ANES undertook efforts to assess the accuracy of respondent reports by sending interviewers to government offices around the country to examine written records and determine whether each ANES respondent was indeed registered and did in fact vote. This procedure was implemented for a series of surveys. It was abandoned because the directors of the ANES concluded that the government record checks were not more valid than the survey respondents' self-reports.

When we submitted a proposal to the National Science Foundation to co-direct the American National Election Studies from 2005 to 2009, we were aware that this issue was worth revisiting, because commercial organizations, political parties, and academic researchers had

begun to assemble government records of registration and turnout from most states in new ways, thanks to electronic record-keeping. So we told NSF that we would gather and release to users as much data on our respondents' registration and turnout histories as we could obtain from government agencies.

When we began our data gathering effort, we learned of significant challenges. First, all of the commercial companies we spoke with declined to reveal the specifics of how and when they obtained records from each state or locality, how often the records were updated, and how the process of matching survey respondents to record entries was done. Because we believed that the ANES is obligated to fully reveal the processes used to generate the data it releases, we felt we could not purchase and release data from these companies. So we decided to gather the government records data from the states ourselves.

This process of record gathering was surprisingly complex. Some states (e.g., Illinois and Virginia) explicitly bar researchers from obtaining access. While some states (e.g., California) provide records on CD-ROMs in easy-to-use formats, this was not true for all states. And whereas some states make these data available at nominal costs, other states charge large amounts of money. Lastly, we found that different states released different types of information on residents and updated records in different ways, so that we could not implement a single consistent data collection process across all states.

We therefore decided that before we spent large amounts of money collecting and attempting to use records from all states, we would carry out a trial project. The trial project involved gathering information on respondents who participated in the 2008 ANES Panel Study in just a few states. We chose the states out of convenience because we were able to obtain the government records relatively easily and affordably, and the chosen states have reputations for

maintaining relatively well-kept electronic records. Therefore, any conclusions we reach with their data might suggest an upper bound of the quality we could expect if records from all states were obtained and analyzed.

This report describes the findings from our trial project. We learned that record checking of this sort was challenging and time-consuming but doable. However, the results of our efforts led us to conclude that the process of matching survey respondents' names and addresses with official government records is fraught with substantial problems. We concluded that we could not have any more confidence in results based on available records than we had in the original survey self-reports. Indeed, our analyses led us to conclude that the self-reports might be of superior accuracy.

For these reasons, we decided to release this report and the data on which it is based, but not to devote additional project resources to gathering and releasing government records data for respondents from other states.

Of course, the conclusions reached here are specific to the exercise we carried out and may not generalize beyond it. We focused on the 2008 ANES Panel Study respondents and the information collected from them for that study. Perhaps different results would be obtained if data were collected from survey respondents in different ways. Second, we focused on a particular set of states. Perhaps other conclusions would be reached using data from other states. Third, our exercise sought to document behavior in the 2008 election. Perhaps different conclusions would be reached with data on other elections. Lastly, we implemented a series of specific procedures for matching survey respondents with government records. Perhaps other conclusions would be reached using other matching procedures.

We therefore hope that this report serves to inspire other researchers to carry out similar

investigations, make public the programming and other materials that they use available to the public, and to assess the generalizability of the findings reported here.

This work was conducted in the context of our responsibilities as co-principal investigators of the ANES from 2005-2009. Those responsibilities include working with the ANES staff to provide publicly available high quality data for widespread use. The work described in this report uses data that are quite sensitive, including the names and addresses of ANES survey respondents. This information is highly confidential and has never been distributed to anyone other than ANES staff, and has been shared with them only for quality control purposes by the firms that have collected ANES data in the past.

In carrying out the research described in this report, we conformed to long-standing ANES policy regarding access to confidential data. In particular, we assigned tasks to ANES staff members and supervised their conduct of that work. In this case, Dr. Matthew Berent, an ANES staff member, handled the confidential data. We, the Principal Investigators, never had direct access to these data. The research in this report was conducted by working with Dr. Berent, who conducted particular aggregate analyses and described aggregate results to us.

To allow other researchers to evaluate the work reported here, we will make public (on the ANES website) electronic files documenting every step of the empirical work reported in this memo. The only data that cannot be made public are the respondents' names and addresses, which ANES rules prohibit from being distributed. Of course, this information is crucial for any scholar to fully replicate our work and to conduct a reanalysis using a different method. We hope that it will be possible for ANES to find a way to allow scholars to conduct such analyses without compromising the complete confidentiality that was promised to the ANES respondents.

Jon A. Krosnick and Arthur Lupia

The Quality of Government Records and Over-estimation of Registration and Turnout in Surveys:

Lessons from the 2008 ANES Panel Study's Registration and Turnout Validation Exercises

Executive Summary

The proportion of representative sample survey respondents who say they turned out to vote in a particular election is often far greater than the proportion of the population who turned out according to official turnout tallies. Such over-estimation of turnout has lead many analysts to conclude that some survey respondents intentionally lie about having turned out. These conclusions have led broader audiences to question the veracity of survey-based turnout data.

To mitigate problems caused by possible over-reporting, some researchers advocate "turnout validation." Turnout validation is a process whereby a survey respondent's report of having turned out or not is checked against official government records for that individual. A parallel process, registration validation, can be implemented to check whether respondent reports of being registered to vote are consistent with government registration records. Many scholars believe that the data produced by registration validation and turnout validation exercises yield more accurate measurements than do survey respondents' reports of such behaviors.

In this report, we examine key assumptions underlying the claim that registration and turnout validation yield improved turnout measures. We document a number of problems with validation methods that limit their value and their ability to produce improved measures. We also find that survey respondents' reports of their own registration and turnout behavior are more accurate than previously believed.

We base our conclusions on four key findings. Three of these findings pertain to the

quality of the government records that are used in validation exercises.

First, official government records of registration and turnout contain numerous errors that make effective registration and turnout validation more difficult than many scholars might expect.

Second, the character and frequency of such errors varies substantially from state to state. These differences occur because states keep and update registration and turnout records in very different ways.

Third, state-level variations in the number and kinds of government record errors cause another problem. Methods for matching survey respondents to government records that minimize the impact of some states' record errors are quite unsuccessful in other states. Hence, one-size-does-not-fit-all when it comes to the best way to validate registration and turnout across U.S. states.

Our fourth finding pertains to the veracity of survey respondent self-reports of registration and turnout. We found that for respondents whose government records can be identified, the records and self-reports show very high levels of agreement. This finding implies that overestimation of turnout rates by surveys is attributable to factors and processes other than respondent lying.

A further implication of our findings pertains to the seeming success of past turnout validation studies. A typical outcome of a turnout validation exercise is as follows: official vote tallies indicate a turnout rate of 60%, survey respondents report having turned out at a rate of 80%, and turnout validation leads to the conclusion that only 65% of the respondents actually turned out. Such downward revisions of the survey respondents' turnout rates are routinely assumed to be evidence of the success of turnout validation and of lying by respondents. Our

evidence points to a very different explanation for the same outcomes -- competing biases that do not in fact improve accuracy.

The first bias is what we call a "downwards bias" in estimation of registration rates. It is caused by multiple problems inherent in matching survey respondents to government records. Errors in government records and errors in the process of obtaining identifying information from survey respondents inhibit success in matching respondents to their government records. These errors yield implausibly low apparent rates of registration. That is, common means of registration and turnout validation fail to locate records for some respondents who were actually registered and actually turned out. The downwards bias in estimated registration rates occurs because these unmatched respondents are inaccurately classified as not having been registered when in fact they were registered.

The second bias is what we refer to as an "upwards bias" in the rates of reported turnout provided by respondents, but not because of respondent lying. Instead, the actual rate of turnout among people who agree to participate in surveys appears to be higher than the turnout rate in the population in general. That is, people who vote in elections (and thereby choose to express their political preferences) also appear to be unusually likely to participate in political surveys (and thereby choose to express their political preferences). As a result, in an election where government tallies indicate a turnout rate of 60%, survey respondents accurately report having turned out at a considerably higher rate, say 80%, yielding an apparent overstatement of the turnout rate.

Hence, the apparent success of turnout validation exercises in producing aggregate turnout rates that are closer to reported official turnout rates can occur not because the exercises weed out liars, but because the unrecognized "downwards bias" in the matching rate is stronger

than the unrecognized "upwards bias" in the actual rate of turnout among survey participants.

Taken together, this evidence suggests that a major factor driving "turnout over-estimation" in Internet surveys may not be respondent lying. Instead, it may be the fact that people who choose to participate in such surveys also choose to participate in elections at higher rates than do people who do not participate in surveys. Traditional approaches to turnout validation do not solve this problem, they only mask it. Therefore, analysts who want to use surveys to study registration and turnout should seek to (1) improve the quality of records kept and distributed by government agencies, (2) improve the identifying information obtained from survey respondents to implement matching with government records, and (3) seek to increase survey response rates among people who do not typically register to vote and turn out.

The Quality of Government Records and "Over-reporting" of Registration and Voting in Surveys:

Lessons from the 2008 ANES Panel Study's Registration and Turnout Validation Exercises

On August 2, 2008, *The Wall Street Journal* published an article entitled, "When Voters Lie." The article claimed that "it's a given that people fib in surveys" and described ways in which "pollsters are trying to uncover the truth." Articles such as this reflect widely held concerns about the honesty of survey responses.

One measurement about which observers are particularly concerned is turnout (see, e.g., Burden 2000). Several researchers have suggested that some respondents intentionally claim to have voted when they did not, in order to portray themselves as honorable and responsible citizens (Aarts, 2002; Andolina, Keeter, Zukin, & Jenkins, 2003; Blais, Gidengil, Nevitte, & Nadeau, 2004; Brockington & Karp, 2002; Corbett, 1991; Lutz, 2003; Lyons & Scheb, 1999).

Aggregate evidence appears to support this claim. The proportion of survey respondents who claim that they turned out to vote is often far greater than reported by government officials (see, e.g., Bernstein, Chadha, & Montjoy, 2001; Burden, 2000; Clausen, 1968-69; McDonald, 2003; Presser, Traugott, & Traugott, 1990; Traugott & Katosh, 1979). In the U.S. context, it is not uncommon for 70 percent of survey respondents or more to report having voted in an election, when government agencies report turnout rates of 50 percent or lower. Such over-overestimation of turnout leads many analysts to conclude that survey respondents lie about their electoral participation.

But there are other possible explanations for survey over-estimation of turnout rates. For example, it is possible that people who participate in surveys generally, or in surveys about politics in particular, are more likely to turn out than people who do not participate in such

surveys.

To manage problems caused by the potential for respondent lying, some researchers have implemented "turnout validation." Turnout validation is a process in which a survey respondent's self-report is replaced by an official government record of whether that individual turned out in a particular election. Many scholars believe that using official government records yields more accurate assessments of respondent behavior (e.g. Bernstein, Chadha, & Montjoy, 2001; Green & Gerber, 2005).

In a number of past years, the ANES collected government records on its respondents. Analyses of those data repeatedly documented instances in which respondents said they turned out but government records failed to corroborate those claims (Abelson, 1990; 1996; Abramson & Claggett, 1984; 1986; 1989; 1991; 1992; Belli, Traugott, & Beckmann, 2001; Cassell, 2004; Claggett, 1990; Deufel & Kedar, 2010; Duff, Hanmer, Park, & White, 2004; Katosh & Traugott, 1981; Traugott, Traugott, & Presser, 1992). Such findings have reinforced confidence in the idea that government records measure turnout more accurately than do respondent self-reports. As a result, many researchers have used government records of citizens' turnout instead of survey self-reports in their work (Dyck & Gimpel, 2005; Gimpel & Schuknecht, 2003; Gerber & Green, 2000; Haspel & Knotts 2000; Sigelman & Jewell, 1986).

The conclusion that government records of an individual's turnout behavior are more accurate than survey respondents' descriptions of these behaviors rests on two key assumptions. The first is that respondents who are registered and have turned out in an election will have a government record that can be unambiguously located. The second is that inability to locate a government record of registration or turnout for a respondent indicates that the respondent actually is not registered or did not turn out.

There are several reasons to question these assumptions. First, government records undergo frequent updating after elections. If researchers do not obtain government records immediately after the election they are studying (which has often been the case – such records have often been sought weeks or months later), some survey respondents who were earlier listed in government records may no longer be there. Causes of removal include: moving from one address to another, name changes, felony convictions, and dying. As a result, information about people who turned out in an election may have been removed by the time that analysts obtain the records.

Second, local jurisdictions report turnout behavior of individuals to Secretaries of State, who use that information to update their records, but some jurisdictions do not report that behavior until long after an election. For example, McDonald (2007) noted that all but one Florida county reported turnout for the 2004 election within three months of that election. For the remaining county, reported were not provided until 10 months after the election. This means that turnout histories from some areas within a state may not be available until well after an election, by which time other attributes of the voter's government record may have changed.

Moreover, failure to match a survey respondent to a government record may happen because of errors in either the government record or in the survey data. Government records might contain accidental mis-spellings, as might survey records. Or a respondent's name could be represented accurately but differently in the two different databases (e.g., "J. Robert Smith" in one, "John Smith" in the other). Or a respondent might have moved to a new residence, whereas his or her government record might not yet reflect that new address, because he or she has not reported the change. In all such instances, efforts to match a respondent to a government record may fail, which may be interpreted as indicating that he or she was not registered and did not

vote, when in fact, he or she did.

For reasons such as these, it may be difficult or impossible to conclude with certainty that failure to locate a record of turnout for a respondent indicates that the respondent did not turn out. So although researchers who implemented turnout validation have failed to find official government records for a substantial proportion of survey respondents who claim to be registered (Clausen, 1968-69; Fullerton, Dixon, & Borch, 2007; Presser, Traugott, & Traugott, 1990; Traugott & Katosh, 1979), the meaning of such failures is uncertain.

All of the possibilities raised above led us to question whether analysts should use government records instead of respondent reports of registration and turnout. To evaluate the value of such a strategy, we conducted and evaluated registration validation and turnout validation exercises. Specifically, we compared survey reports of registration and turnout from respondents in the 2008 American National Election Studies Panel Study to individual-level government records of registration and turnout for those respondents who lived in several large states. Our analyses were designed to document (1) the accuracy of government records, (2) variation in the accuracy of records across states, (3) the resemblance of the subset of survey respondents for whom government records indicated that they turned out with the voting population, and (4) the differences in conclusions that would be reached using self-reports vs. using government records of turnout.

We begin below by discussing government records and describe some of the challenges involved in locating records for survey respondents. We examine government records from six states for this purpose. We describe how often listings in those records contain several important types of errors. Then, we describe and evaluate different strategies for locating survey respondents' government records in the presence of such errors.

Summary of Findings

From these analyses, we conclude that that the benefits of turnout validation are quite limited and that factors other than respondent lying are a substantial cause of turnout over-estimation. We base these conclusions on four key findings. Three of these findings pertain to the quality of government records:

First, official government records of registration and turnout behavior contain errors that make it difficult to accomplish accurate turnout validation.

Second, the character and frequency of such errors varied substantially across states. These differences occurred, in part, because states kept and updated registration and turnout records in very different ways.

Third, state-level differences in government records, including differences in types of errors, caused procedures that were apparently successful in matching survey respondents to government records in one state to be very unsuccessful in other states. This finding suggests that one-size-does-not-fit-all when it comes to the best way to match registration and turnout records to respondents across the states.

Fourth, for survey respondents whose government records can be located, there are extraordinarily high rates of agreement between their self-reports and what their government records indicate. Indeed, people who participated in the 2008 ANES Panel Study, and whose self-reports could be matched to government records, genuinely turned out to vote at a higher rate than the general vote-eligible population. This finding implies that lying contributed less to turnout over-estimation than is commonly thought. This evidence suggests that surveys yield higher levels of apparent turnout than official vote tallies because people who choose to participate in surveys about elections are also more likely to participate in elections themselves.

Hence, the belief that turnout validation produces more accurate turnout measurements depends on assumptions about government records and turnout validation that analysts should not presume to be true.

A further implication of the findings reported below pertains to the seeming success of past turnout validation studies. Consider, for example, a typical turnout validation exercise that produces the following result: (1) government records indicate a turnout rate of 60%, (2) survey respondents report having turned out at a rate of 80%, and (3) the proportion of survey respondents for whom a government record of turnout can be identified is 65%. Although this 65% figure is often viewed as more plausible than the 80% figure based on self-reports alone, this apparent improvement in accuracy is in fact due to the cumulative effect of two sources of cross-cutting bias in the record-matching and respondent recruiting processes.

The first bias is what we call a "downwards bias" in estimation of registration rates. It is caused by multiple problems inherent in matching survey respondents to government records. Errors in the government records and errors in the process of obtaining identifying information from survey respondents inhibit successful matching of survey respondents to their government records. These errors yield implausibly low apparent rates of registration. In other words, common means of registration and turnout validation fail to locate records for some respondents who were actually registered and actually turned out. The downwards bias in estimated registration rates occurs because such respondents are inaccurately classified as not having been registered when in fact they were registered to vote.

The second bias is what we refer to as an "upwards bias" in the rates of reported turnout provided by respondents, but not because of respondent lying. Instead, the actual rate of turnout among people who agree to participate in surveys appears to be higher than the turnout rate in

the population in general. That is, people who vote in elections also appear to be unusually likely to participate in political surveys. As a result, in an election where government tallies indicate a turnout rate of 60%, surveys can accurately report their respondents as having turned out at a considerably higher rate, say 80%.

Hence, the apparent success of turnout validation exercises in producing aggregate turnout rates that are closer to reported official turnout rates appears to occur not simply because the exercises weed out liars, but because the unrecognized "downwards bias" due to errors in government records is stronger than the unrecognized "upwards bias" in the actual rate of turnout amongst survey participants.

Another implication of this report is particularly relevant for analysts who use "validated registration and turnout" data that they did not assemble on their own. Many analysts give a list of survey respondents to commercial firms, and these firms use a proprietary process to match that list against government records that the firms had previously collected. The commercial firm then tells the analysts the purported registration and turnout behavior of each survey respondent.

In light of our findings, analysts whose credibility depends on the validity and reliability of their empirical claims should consider requiring firms who wish to sell validation services to document how, where, and when they obtained government records and to document the procedure that they used to match survey data to government records. Because of the difficulties we describe in this report, researchers should regard firms' claims about data accuracy as credible only to the extent that they are willing to document their data sources and matching procedures.

More generally, researchers should proceed with caution when using data that purports to use government records to replace or validate survey reports of registration and turnout. Such

data are likely to contain many errors.

Government Records of Registration and Turnout

To begin explaining what we did and what we found, we provide an overview of the processes by which government records of registration and turnout are built and maintained. We then describe the accuracy of such records.

How Government Records Are Assembled and Updated

Each state maintains its own records of registration and turnout. A record for an individual is created when an individual first registers to vote in a state. Individuals supply information about themselves when they register to vote. Federal law requires each individual to provide either the number of his or her state-issued driver's license or the last four digits of his or her social security number when registering. States vary in what other information they solicit. Common types of information include current and former names, current and former residential addresses, mailing addresses, date of birth, sex, party affiliation, race, phone number, place of birth, and the last year in which the individual turned out to vote. States can add other information to a registration record, including an individual's registration number, date and method of registration (e.g. at a state DMV office or by a registration form mailed to a county board of elections), and turnout history.

The Help America Vote Act of 2002 (HAVA) and the National Voter Registration Act of 1993 (NVRA) specify minimum requirements about how state records must be kept, but the states also have some latitude in managing their records. States can decide what information provided by individuals on registration forms will be included in public databases, and whether and how such information is made available to researchers. For example, a state might require an individual to provide a full date of birth on a registration form but only include the year of

birth or age in the released government record. Since states differ in how they manage the

records, information available from some states is not included in the records provided by other

states, including sex, registered party affiliation, race, phone number, place of birth, method of

registration, and previous names and addresses.

States also differ in the kinds of people who are included in their records. In all states,

government records of registration and turnout include information about citizens who have

registered to vote. These records do not include information about citizens who have never

registered. Beyond these similarities, states differ in terms of which people are included in the

government records. For example, in some states, government records include information about

individuals who were previously registered but are no longer eligible to vote (e.g. individuals

who have died, moved, or committed a felony). Some states include information from

individuals who have attempted to register but were denied (e.g., individuals who failed to

provide all required information on a registration form). Other states do not include information

on such people. Hence, the types of people represented in government records vary from state to

state.

State databases are also updated regularly. However, states differ considerably in how

thoroughly and how often they update their records. For example, although all states have access

to frequent changes in the United States Post Office's National Change of Address (NCOA)

directory, states do not reference this list with common frequency or intensity. Some states put

more effort than others into updating their records to reflect changes in the NCOA directory.

States differ in how and when they obtain and use information from federal and state

agencies to identify changes that should be made in individuals' registration records. Such

changes can be caused by lists of individuals who have been convicted of felonies and lists of

citizens who have died. North Carolina obtains information from its Department of Corrections

and Department of Health every month. In contrast, North Carolina obtains the NCOA from the

United States Post Office once every three months. If the NCOA list indicates that a registered

citizen has moved during the previous three months, then that individual's record is updated to

reflect the new information.

In general, the frequency and depth of updating in a state appears to depend on the

training and availability of qualified staff to enter new data and conduct rigorous reviews for

accuracy. Such capacities vary across states. States also vary in the kinds of data that they have

available to them. For example, local jurisdictions within states maintain their own records.

Statewide records can be updated with information from the local jurisdictions only after the

jurisdictions furnish new information to the state. How quickly local jurisdictions provide

updated records to the state can depend on both the jurisdiction and the type of information.

Some information, such as new registrations, tends to be sent to the state shortly after it is

received by a local jurisdiction. Other information, such as updated turnout histories, might not

be sent to the state until well after an election. Hence, a state's records may be more up-to-date

for some jurisdictions than for others, and turnout histories for slower-to-report jurisdictions

within a state may not be updated until well after other changes to the registration database are

made. Hence, there may not be any single point in time when both state registration and turnout

records are simultaneously accurate, nor may there be a point in time when the records of

multiple states are equally accurate.

The Challenge of Locating Government Records for Survey Respondents

In order to validate a survey respondent's reported registration or turnout status, that

individual's government record of registration must be located. A match may be accomplished

by locating a government record for an individual with the same full name as that provided by a

respondent. However, government records from large states may contain many individuals with

the same name. If two individuals in the government records share the same name as a survey

respondent, then attempting to find the survey respondent's government record becomes

complicated.

Researchers can increase the chances of correctly locating a respondent's government

record by using other information in addition to names. If a survey respondent named Pat Jones

has indicated he is male, was born on January 1, 1987, and lives at 123 Elm Street in Anytown,

PA, the likelihood of a successful match can be increased. The more pieces of information from

a survey respondent that match a government record, the greater is the basis for a successful

match. In order for this matching process to be maximally successful, survey respondents must

provide all of the pieces of information (e.g., name, gender, date of birth, residential address) that

the state makes available in its registration and turnout records. Thus, one challenge is that

surveys may sometime fail to obtain all pieces of potentially useful information from

respondents.

Another challenge occurs if mistakes are made in recording information during a survey

interview. For example, a survey respondent might mistype his or her street name or house

number when completing an online survey, or an interviewer might mistype what a respondent

says during an oral interview. Likewise, a researcher might have difficulty deciphering what a

respondent has written on a paper and pencil questionnaire. Such mistakes constitute potential

obstacles to accurately identifying a respondent's government record.

Because people change their names and addresses over time, the information obtained

from survey respondents should ideally refer to the moment when the government records

information is recorded. But this is practically impossible. Imagine, for example, a respondent

interviewed in September before an election, when the researcher will obtain records of

registration and turnout in December, three months later. Ideally, the respondent would be asked

to tell the interviewer what his or her name and address was or will be at the time of the state's

updating of the government record that will be obtained. But of course, it would be impossible

to tell respondents on what date that updating has or will occur, and the question would be

unanswerable if the update will happen weeks or months after the survey interview.

Government records may also be incomplete for some individuals and contain errors for

others. For example, McDonald (2007) reported that individual's sex was missing from 50% of

the 2004 California government records, race was missing from 50% of the 2004 Kentucky

records, and about 86% of the 2004 records from Oklahoma did not include the date on which a

person registered.[1] McDonald (2007) also observed that years of birth for some records were

clearly inaccurate, because they indicated that some individuals were several hundreds of years

old and that others were not yet born. Such errors and omissions may make it impossible, or at

least very difficult, to locate some survey respondents' government records (e.g. Presser,

Traugott, & Traugott, 1990).

The frequencies of inaccurate and missing data in government records are not equivalent

across the states. This inconsistency creates another problem for researchers hoping to use those

records for registration validation or turnout validation. In contrast to California's 2004 records,

none of the 2004 records from Iowa and Kentucky were missing an individual's sex. In contrast

to Kentucky, none of North Carolina's 2004 records were missing race (McDonald, 2007).

---

[1] McDonald (2007) reviewed government records for the 2004 general election that were available from Delaware, California, the District of Columbia (DC), Florida, Iowa, Kentucky, Maryland, North Carolina, Ohio, Oklahoma, and South Carolina. McDonald chose not to review 2004 records from Connecticut, Missouri, and New Jersey "due to a high amount of missingness" in the data (p. 592).

About 10% of the records in the 2010 West Virginia voter list contained addresses that were

identified as undeliverable or probably undeliverable by the Coding Accuracy Support System

(CASS) of the U. S. Post Office, whereas fewer than 2% of the records in Maryland,

Washington, and the District of Columbia displayed such problems (Ansolabehere & Hersh,

2010).[2] Differences in the proportions of inaccurate, problematic, and missing data in the states'

records mean that a method for locating records that works well in one state may not succeed in

other states.

A final problem is that information provided by a survey respondent and information in

their government record may not be identical, even when neither contains errors. For example, a

government record might show an individual's proper first name (e.g., Patrick), while the

individual provided an informal version of that name to a survey researcher (e.g., Pat). This can

be especially problematic for matching when a respondent provides an informal name, or

nickname, that is not commonly associated with a specific proper name (e.g., 'Bud' or 'Butch'

which are sometimes used to distinguish sons from fathers with the same name).

Criteria for Locating Government Records Given Inherent Problems

Researchers using government records for registration validation or turnout validation

must develop criteria by which respondents' government records are to be considered "matched"

or "not matched" to a particular survey respondent. The choice of criteria used to determine

whether a respondent's government record has been matched accurately is likely to affect the

quantity and types of errors the researcher will make. Consider Table 1. The true state of affairs

includes two possibilities for every respondent: the respondent was either registered or not. A

procedure used to locate the record will produce one of two outcomes: the respondent will be

---

[2] Ansolabehere and Hersh (2010) worked with a commercial vendor that has compiled a national voter registration database from state records.

matched or not.

Imagine a respondent who is not registered and who, therefore, does not have a government record. Now suppose a researcher implements a matching procedure that attempts to find the respondent's first and last names in a government record. If no government record exists for those names, the respondent will not be matched to a record. If we use such an outcome to conclude that the respondent was not registered, then our conclusion is *accurate* – a *true negative*. If, however, the procedure locates a government record for a different individual of the same name, and we use such an outcome to conclude that the respondent was registered, then our conclusion is inaccurate -- a *false positive*.

Now imagine an individual who is registered and for whom a record does exist. A procedure might fail to locate the government record for that individual if the respondent provided an informal first name while the government record contains a proper first name. If we use such an outcome to conclude that the individual is not registered, then our conclusion is inaccurate - a *false negative*. Now suppose that the procedure locates a government record for an individual with the same name as the respondent. *The accuracy of this outcome can be uncertain.* If the procedure matches the respondent to the government record of a different individual who happens to have the same name, then the procedure has not accurately matched the respondent to his or her government record. Only when the procedure matches the respondent to her or his own record would we validly conclude that the respondent is actually registered.

The criteria used to determine a "match" will affect the accuracy of turnout validation attempts. A very liberal matching procedure, such as one that requires only that a survey respondent's name be similar to a name in a government record, will generate many matches. Such a procedure can yield many accurate matches of respondents to their actual records (true

positives), this procedure can also match many respondents to the wrong records (yielding uncertain accuracy and false positives).

Strict criteria can reduce the number of false positives. For example, coding as "matched" only respondents for whom a government record shares identical first and last names, address, and date of birth, will decrease the number of "matches." Even a slight difference in the names or an error in the address or date of birth will lead to a failure to match a respondent to her or his record. With fewer matches, there are fewer false positives and/or fewer false negatives.

In sum, decisions about which criteria to use in an attempt to match each survey respondent to a government record will affect the quantity and types of errors present in registration or turnout validation. Less strict procedures can generate more true positives and fewer false negatives (increasing accuracy) but also more false positives and fewer true negatives (decreasing accuracy). More strict procedures can yield more true negatives and fewer false positives (increasing accuracy) but also more false negatives and fewer true positives (decreasing accuracy). Thus, understanding the criteria used to determine whether or not survey respondents are considered "matched" to government records is an essential step in evaluating the value and accuracy of turnout validation exercises.

About Our Data

In this section, we describe the data that we used in our evaluation exercise. We first describe the survey data and then the government records. At the end of this section, we report our first main finding, which pertains to the existence and frequency of missing and incorrect government records.

The 2008-2009 ANES Panel Study

Our survey data comes from the 2008-2009 ANES Panel Study. Respondents in the

2008-2009 ANES Panel Study were recruited in two cohorts by Knowledge Networks via RDD telephone calls to land lines.[3] Recruitment of the first cohort took place between September 26, 2007, and January 27, 2008. The second cohort was recruited between May 28, 2008, and September 9, 2008. A total of 2,367 eligible respondents from the first cohort completed an initial recruitment survey, and 1,839 eligible respondents from the second cohort did so.[4] Thus, the initial sample included 4,206 eligible respondents.

The first cohort was asked to complete monthly surveys online starting in January, 2008. The January, February, June, September, October, and November surveys included questions used in the analyses reported below. Not all respondents from the first cohort completed every survey. The numbers of respondents completing the monthly surveys were: January=1,623; February=1,457; June=1,420; September=1,488; October=1,510; and November=1,507.

The second cohort began completing monthly surveys starting in September, 2008. The numbers of respondents from the second cohort completing the surveys analyzed here were: September=1,098; October=1,118; and November=1,158. Thus, the total numbers of respondents from both cohorts for the September, October, and November surveys were 2,586, 2,628, and 2,665, respectively.

A total of 3,406 people completed at least one survey in January, February, June, September, October, or November. Of these, 994 resided in California, Florida, New York, North Carolina, Ohio, or Pennsylvania. These are the six target states from which we obtained the government records that we use in this report.

Questions used to measure registration were asked during the January, February, June,

---

[3] The recruitment method excluded cell phone only households, which resulted in some coverage error.
[4] A respondent was eligible to participate in the 2008-2008 ANES Panel Study if he or she was a U. S. citizen born on or before November 4, 1990 and living in a household served by a sampled landline telephone number at the time of recruitment.

and September surveys. The registration questions were asked of every respondent in the January and June surveys. Registration was measured in February for respondents who had not answered the question in January. Similarly, registration was measured in September for respondents who had not answered the question in June. Only respondents who were 18 years of age or older on the day they completed a survey were asked to report whether they were registered to vote and, if so, the county in which they were registered to vote. The specific questions, with response options in parentheses, were:

> "Are you registered to vote, or not?" (Yes, registered to vote; No, not registered; Don't know)
>
> If registered: "Your residence is located in [county]. Are you registered to vote in [county] or somewhere else?"[5] (Registered in [county]; Registered somewhere else)
>
> If somewhere else: "In what county and state are you registered?"[6]

This report uses each respondent's most recent answer to these registration questions. That is, self-reported registration was taken from the September survey for respondents who answered the questions during that month. Answers from the June survey were used for respondents who did not answer the questions in September. Similarly, answers from the January or February surveys were used for respondents who had not answered the questions during a later month.

Answers from the September survey were available for 54.28% of the 3,049 respondents. Answers from the June, February, and January surveys were used for 32.04%, .75%, and 6.43% of the respondents respectively. The remaining 6.49% of respondents did not answer the registration questions during any of the surveys.

---

[5] "County" was replaced with "Washington DC", "Parish", or "Borough" if the respondent lived in Washington DC or a state with Parishes or Boroughs instead of counties.

[6] Respondents typed their registration county and selected their state from a pull down menu.

Turnout in the 2008 general election was measured by questions asked during the October and November 2008 surveys. The October survey asked:

"This question is not about the primary elections and caucuses that were held a few months ago. Instead, we'd like to ask you about the election for President to be held on November 4, in which **[BARACK OBAMA / JOHN MCCAIN]** is running against **[JOHN MCCAIN / BARACK OBAMA]**. Have you already voted in that election, or not?" ("Have already voted in that election" and "Have not voted in that election").

The order in which the major party candidates were presented was randomly determined for each respondent.

In the November survey, all respondents were asked:

"The next few questions are about the presidential election that was held on November 4. In asking people about elections, we often find that a lot of people were not able to vote because they weren't registered, they were sick, they didn't have time, or something else happened to prevent them from voting. And sometimes, people who usually vote or who planned to vote forget that something unusual happened on Election Day one year that prevented them from voting that time. So please think carefully for a minute about the election held on November 4, and other past elections in which you may have voted, and answer the following questions about your voting behavior.

Which one of the following best describes what you did in this election?

> Definitely did **not** vote.
>
> Definitely voted in person at a polling place **on** election day.
>
> Definitely voted in person at a polling place **before** election day.
>
> Definitely voted by **mailing** a ballot to elections officials before election day.

Definitely voted in some other way.

Not completely sure whether you voted or not."

Respondents who indicated that they were not completely sure whether they voted were asked a follow-up question:

"If you had to guess, would you say that you probably did vote in the election, or probably did not vote in the election?" ("Probably voted", and "Probably did not vote").

Respondents were coded as having turned out if they selected "Have already voted in that election" during the October survey or selected "Definitely voted in person at a polling place **on** election day", "Definitely voted in person at a polling place **before** election day", "Definitely voted by **mailing** a ballot to elections officials before election day", "Definitely voted in some other way", or "Probably voted" during the November survey. People who selected "Have not voted in that election" during the October survey, and "Definitely did **not** vote" or "Probably did not vote" during the November survey were labeled as having not turned out.

Of the 3,049 respondents, 83.99% answered both the October and November turnout questions. An additional 4.99% answered only the November turnout question, and 4.00% answered only the October question. The remaining 7.02% did not answer the turnout question during either survey.

Distributions of self-reported registration. Table 2 presents statistics that summarize survey respondents' claims about their registration status. 82.52% of the 3,049 respondents reported that they were registered in the state in which they resided. An additional 2.79% reported being registered in a state other than the one in which they lived. 6.95% of the respondents said they were not registered. Another 1.12% did not know if they were registered. The remaining respondents (6.63%) never answered any questions about registration.

Of the 994 respondents residing in one of the target states, 81.89% said they were

registered in the state in which they lived (see the second row of Table 2), 3.12% reported

registration in a different state, 7.44% said they were not registered, and .80% were unsure of

their registration. The remaining 6.74% never answered any registration questions.

Distributions of self-reported turnout. Table 3 presents statistics that summarize survey

respondents' claims about their turnout behavior. 80.06% of the 3,049 respondents reported

turning out in the 2008 general election. 9.25% of this sample said that they did not turn out.

The remaining 10.69% did not answer any turnout questions.

Of the respondents living in one of the target states, 80.58% reported turning out in the

2008 general election. 8.55% claimed not to have turned out. The remaining 10.87% did not

answer turnout questions.[7]

---

[7] Some respondents answered registration and turnout questions during more than one survey. Three different types
of inconsistencies could occur among registration and turnout answers given by a person in different surveys. First,
a respondent might say that she or he was registered during one survey and say that she or he was not registered in a
different survey. Second, a respondent might say that he or she was not registered and then later report that he or
she turned out in the election. Finally, a respondent might indicate during the October survey that he or she had
already voted and then in the November survey, say that he or she did not vote at all or voted on election day.

Inconsistencies in reports of registration, and inconsistencies between reports of registration and turnout,
are not necessarily errors. There are many reasons why a respondent's registration status may have changed
between surveys, and why respondents who said they were not registered may have actually turned out in the 2008
general election. Some individuals may have registered, or requested to become unregistered, between surveys.
Others may have registered after the September survey was completed and were eligible to vote on Election Day.
Still others may have moved between surveys and not registered at their new residences. Finally, states can change
registration records for individuals who have moved out of state, who have not turned out in two successive federal
elections, or who have committed a felony.

In 2,291 instances, respondents answered registration questions in multiple surveys. In most cases
(96.68%), people provided identical registration reports. In 1.53% of the instances of multiple registration reports, a
person changed from being registered to being not registered, and in .96% of the instances, a person changed from
being not registered to being registered.

Inconsistencies between reported registration and reported turnout were more common. Of the 188
respondents who said they were not registered as late as September of 2008, 28.72% said they turned out in the 2008
general election. If these self-reports of non-registration and turnout are accurate, then a sizable portion of
respondents who were not registered as of September 2008 later registered early enough to be eligible to vote in the
2008 general election.

Inconsistencies in the October and November reports of early turnout were also in evidence. Of the 130
respondents who reported early turnout during the October survey, 21.54% later reported that they "Definitely voted
in person at a polling place on election day", and an additional 3.08% later said that they "Definitely did not vote".
Only 66.15% of the respondents who reported early turnout during the October survey also reported the same type
of turnout during the November survey (i.e., "Definitely voted in person at a polling place before election day",

Government Records

In the sections that follow, we compare respondent records from the 2008-2009 ANES Panel Study to government records from the six large states. In this section, we describe the government records. A key finding in this section is that government records contain a remarkable number of errors. Consequently, when characterizing government records, we are forced to make several assumptions about the meaning of certain data in the records. Our practice is to make assumptions in ways that are biased towards concluding that the records are accurate. This practice makes this section's negative findings all the more worthy of attention.

We began by obtaining government records of registration and turnout from California (N=17,094,209), Florida (N=12,570,868), New York (N=12,718,771), North Carolina (N=6,230,749), Ohio (N=8,246,881), and Pennsylvania (N=8,444,317) between February and June of 2010 (see Appendix A for details on the government records we obtained). Each file contained the names, registration, and turnout information for every individual registered in those states at the time the records were obtained, plus information for some people who used to be registered but were no longer registered at the time the list was obtained.

We coded every first and last name, residential and mailing address, and date of birth or age in the government records as "potentially accurate", "probably inaccurate", or "missing".

"Definitely voted by mailing a ballot to elections officials before election day", or "Definitely voted some other way"). The remaining 9.23% did not answer the turnout question during the November survey. This indicates that at least one quarter of the respondents who said they voted early misreported their early turnout in either the October or November survey.

Respondents who said they were not registered during one survey and that they were registered during another survey may have misreported their actual registration at least once. Similarly, people who said they were not registered but also said they voted may have misreported either their registration or turnout. In contrast, people who reported not turning out early after having previously reported early turnout clearly made at least one error. However, self-reports alone cannot be used to determine whether the misreporting is due to respondents' desires to portray themselves in a favorable light or respondents' memory errors. It is not possible to determine whether reports of registration or early turnout are accurate without accurate matches to accurate government records.

Given the huge size of the government records, visually inspecting every entry was not practical.

We therefore first sorted the government records by a variable to be coded. This resulted in

records with similar characteristics being grouped together. For example, all records that were

missing an entry for a variable were automatically grouped together. Similarly, entries

beginning with a character of the English alphabet were grouped together, while entries

beginning with any other character were grouped separately. Also, sorting by numeric values or

dates resulted in a list of records in order of the values or dates. These sorting procedures

facilitated a relatively efficient method for identifying and coding names, addresses, and dates of

birth as "potentially accurate", "probably inaccurate", or "missing".

Names, addresses, and dates of birth were coded as "potentially accurate" unless they

appeared to contain one or more errors, in which case they were coded "probably inaccurate."

The criteria used to determine whether a name, address, or date of birth was "potentially

accurate" were fairly lenient, producing a bias in favor of the records' accuracy. Thus, our

estimates of the frequencies with which government records contain information that is

"probably inaccurate" probably underestimate the actual frequencies of errors in those records.

Every first and last name that began with a character of the English alphabet was coded

as "potentially accurate". Names beginning with a space followed by an English alphabet

character were also coded as "potentially accurate," as were strings of English alphabet

characters appearing between starting and ending single or double quotations. Names beginning

with any other character were coded as "probably inaccurate". Some of the initial characters for

first and last names that led to those entries being coded as "probably inaccurate" included

numerals, `, -, ?, ., (, *, and +. If such characters appeared in locations other than the first, the

name was not coded as "probably inaccurate". Finally, first and last names were coded as

"missing" if no characters were entered for a record.

All records with some characters entered for a residential or mailing address were coded as "potentially accurate" for those variables. Records with no characters entered for a residential street or mailing address were coded as "missing" that information. Individuals living in rural areas of some states may not have traditional residential addresses. Such individuals can describe their residential locations with words or pictures in lieu of a traditional address, and missing residential addresses would be expected in their records. All address records containing characters were coded as "potentially accurate", all address records containing no characters were coded as "missing", and no addresses were coded as "probably inaccurate".

A birth date or age was coded as "potentially accurate" if it indicated the individual would have been less than 120 years old on November 4, 2008. Records that indicated the individual was 120 years or older on the date of the 2008 general election were coded as "probably inaccurate". Date of birth or age was coded as "missing" if no valid date or age appeared in the record. Coding date of birth in this manner probably resulted in some records for which date of birth should have been coded as "missing" but were coded as "potentially accurate". Like McDonald (2007), we found that records with a year of birth in 1900 disproportionately had January 1 recorded as the birth month and day. For example, about 63% of the Pennsylvania records with a birth date in 1900 had January 1 as the birth month and day, and no other month and day appeared for more than 1% of those records. This suggests that 01/01/00 may be used in some cases as a placeholder when an individual's date of birth is not known. However, 01/01/00 could be the actual date of birth for some individuals with a government record, and there is no way to distinguish the valid entries from the placeholder entries. Given the dilemma, government records were given the benefit of the doubt, and records

with 01/01/00 as the date of birth were coded as "potentially accurate".

Basic attributes of the registration records. Table 4 contains summary statistics of the government registration records in the six target states. These records identify 98.26% of the 65,305,796 people included in the records as being registered. New York's 12,718,771 listings included the highest proportion of individuals who were no longer eligible to vote (8.32%). In North Carolina, 1.22% of the listings were for individuals who attempted to register and were denied. Records from California, Florida, Ohio, and Pennsylvania only included information for individuals who were registered and eligible to vote.[8]

We compared the total number of registration records in each target state to the state's official statistics about the number of people who were registered to vote in the 2008 general election (California Secretary of State, 2009; Florida Department of State, 2010; New York State Board of Elections, 2009; North Carolina State Board of Elections, 2009; Ohio Secretary of State, 2009; Pennsylvania Department of State, 2010). However, the numbers of registrants in the records we obtained as well as the statistics published by the states are probably inflated due to "deadwood" (Ansolabehere & Hersh, 2010; McDonald, 2007). The term "deadwood" refers to records for individuals who no longer reside at the address on file, have died, have committed a felony and are no longer eligible to vote, or who have become ineligible to vote for some other reason. Although HAVA requires the states to make reasonable efforts to update registration records, individual states vary in how and how often they make such efforts. This means that the amount of deadwood in registration records can vary substantially across the states.

Some research supports this possibility. Ansolabehere and Hersh (2010), working with a commercial vendor, estimated the amount of deadwood separately for each state and the District

---

[8] The Florida and Pennsylvania listings included records for "inactive" individuals. Inactive individuals are still eligible to vote, but have not turned out during at least two previous election cycles.

of Columbia (see Appendix B for information about deadwood estimation). The vendor used

registrants' ages, registrants that could be identified as deceased through other sources,

registrants with an "inactive" registration status, registrants who have moved and registered at a

different address, and recent voting history for their deadwood estimates. Deadwood estimates

ranged from .003% (Wyoming) to 15.68% (Arkansas), and 3.63% nationwide. Apparently,

deadwood in registration records presents a much larger problem in some states than others.

If the deadwood estimates reported by Ansolabehere and Hersh (2010) are accurate,

adjusting the numbers in the records we obtained and the official statistics should yield more

accurate numbers of the people who are actually registered. That is, the number of registration

records that are not obsolete can be estimated by subtracting the estimated number of deadwood

records from the total number in the government files.[9]

Table 5 reports both the original numbers in the government records we obtained and the

official government statistics, as well as the adjusted numbers after correcting for deadwood.

Overall, the number of registered individuals in the government records contained .50% more

registered individuals than the number of individuals who official statistics reported were

registered to vote in the 2008 general election. The differences between the numbers after

correcting for deadwood are similar. Eliminating estimated deadwood resulted in 4.05% fewer

people registered than the government records or official statistics suggest.[10] This indicates that

the records we obtained between February and June of 2010 were different from the records used

by the states to produce their published election statistics.

Basic attributes of the government records. Table 6 reports summary statistics computed

using the government records in the six target states as well as statistics using numbers that gave

---

[9] Deadwood estimates in the states we examined were: California .33%, Florida 7.89%, New York 5.17%, North Carolina 4.94%, Ohio .85%, and Pennsylvania 6.42%.

[10] Ansolabehere and Hersh's (2010) estimated nationwide deadwood rate is 3.63%.

been adjusted for deadwood. The unadjusted records indicate that 68.30% of registered

individuals turned out in the 2008 general election. This is smaller than the 72.23% turnout rate

based on statistics published by the six states (California Secretary of State, 2009; Florida

Department of State, 2010; New York State Board of Elections, 2009; North Carolina State

Board of Elections, 2009; Ohio Secretary of State, 2009; Pennsylvania Department of State,

2010). The adjusted records indicate that 71.15% of registered individuals turned out, while the

adjusted state statistics indicate that 75.18% turned out. Both of these differences indicate that

the government records we obtained after the election were different from the records at the time

of the election.

Some of the discrepancy between the turnout rate in the government records and the

states' published statistics might be due to missing turnout histories in the records we obtained.

Turnout histories for the 2008 general election were not included for four of the 442 counties in

the government records that we received from the six target states.[11] As of June, 2009, the

California government records did not include 2008 general election turnout data for three

counties; and as of February, 2009, the North Carolina records were missing these data for one

county. The four counties had relatively small populations, so the missing data affected only

.17% of registered individuals in California and .13% in North Carolina.

Errors and missing information in the government records. Table 7 contains summary

statistics on errors and omissions in the target states' government records. The government

records included names, addresses, and dates of birth that were clearly erroneous. About .02%

of the listings in the government records did not contain last names or had listings that were

probably inaccurate. Another .02% did not contain first names or had first names that were

---

[11] None of the 2008 ANES Panel Study respondents reported living in one of the four counties without turnout histories for the 2008 general election.

probably inaccurate. Records with addresses that were missing or probably inaccurate made up

.14% of the government records, while a surprising .46% had no date of birth or age or had a

probably inaccurate date of birth or age. That is, *almost 1 out of every 200 government records*

had either no information about an individual's date of birth or indicated that an individual was

120 years of age or older during the 2008 general election.

Table 8 breaks down these problems down further. Government records with missing

birth date or age made up .008% of the cases in the six states we examined, while probably

inaccurate birth dates or ages made up the other .46%.

Tables 7 and 8 illustrate that government records are sometimes incomplete or contain

information that is unlikely to be accurate. Hence, researchers may fail to match a survey

respondent to a government record because the government records are missing information or

contain inaccurate information.

Lessons from Our Attempt to Match Survey Responses to Government Records

Our attempt to match respondents in the 2008 ANES Panel Study to government records

led us to learn a series of lessons. Our main focus was to investigate the correspondence between

survey respondents' reports of their registration and turnout behavior and government records of

the same phenomena for the same people. An extremely high correspondence between self-

reports and government records would suggest that self-reports are accurate. If we find

differences between the self-reports and government records for respondents, then questions

arise as to whether there are (1) errors in the self-reports, (2) errors in the government records,

and/or (3) errors in the process of matching a government record to a survey respondent. Our

goal in this section is to shed light on the extent to which each of these possible explanations is

consistent with the evidence we have accumulated.

This section proceeds as follows. First, we review methods that others used to match survey responses to government records and describe why they are difficult to evaluate. Second, we describe the set of matching methods that we evaluate. Then, we present findings evaluating key assumptions underlying the claim that registration validations yield improved individual-level measures of voter registration. Finally, we report parallel analyses to clarify the accuracy and value of attempts to use government records to validate turnout reports.

Matching Methods Have Common Elements But Are Difficult to Evaluate

Many turnout validation attempts have used names, addresses, and respondent birth dates to match government records to survey respondents. Following general elections from 1976 to 1990, for example, the ANES worked with local officials responsible for managing government records to obtain registration and turnout data. These officials attempted to locate records given ANES respondents' names, addresses, and birth dates. Descriptions of those attempts suggest that not all of the "matches" between the ANES data and the government records were exact. In some cases, people involved in the turnout validation process made judgments about when a respondent's first and last names, address, and date of birth were "close enough" to a government record to be considered a "match."

In the years since the ANES discontinued regular turnout validation exercises, new technologies and laws make such endeavors less onerous to conduct. Computers hasten and reduce the cost of attempts to match records from various sources. And federal requirements that every state maintain a statewide electronic registration database facilitate more automated matching methods.

Although technologies have made data easier to assemble, matching strategies continue to resemble those used in earlier decades. Names, addresses, and dates of birth remain the

primary variables with which scholars have attempted to match survey respondents to government records. For example, Ansolabehere and Hersh (forthcoming) used names, addresses, and ages to electronically match respondents in the 2006 Cooperative Congressional Election Survey (CCES) to registration and turnout records. The electronic matching method used "fuzzy logic" to determine the probability of a true match given variations and discrepancies between the survey data and the government records. This practice is in a sense similar to the judgments that the ANES presumably used to determine that a respondent's first and last names, address, and date of birth were "close enough" to a government record to be considered a "match."

Several commercial vendors also use names, addresses, and dates of birth to electronically match individuals to government records. We attempted to contact 22 commercial vendors who maintain databases of such records to see if we could match the ANES 2008 Panel Study respondents to their records. Appendix C contains a more comprehensive description of these interactions.

All of the vendors who maintained national databases and said they could match respondents to government records indicated they would need at least names and addresses to conduct the matching. Most vendors indicated that the accuracy of the matches could be improved with dates of birth. A few others indicated the accuracy could be further improved with additional information such as sex or telephone numbers. However, among the commercial vendors we contacted, the use of names, addresses, and dates of birth is the norm.

Little has changed since ANES first attempted turnout validation in 1976 in terms of the information that researchers use to match respondents to government records, and detailed descriptions of exactly how a "match" was determined has also been consistently scarce. For the

1984 and 1988 general elections, ANES noted whether or not survey respondents' names, addresses, and dates of birth were exactly the same as the government records to which they were matched. However, details about the principles implemented in making decisions about when inexact matches were considered "close enough" are not available. Similarly, the "fuzzy logic" used by the commercial vendor employed by Ansolabehere and Hersh (forthcoming) is not described in enough detail to enable independent replication.   Here too, then, the details of what constitutes "close enough" are not publicly available.

Most commercial vendors with whom we spoke were unwilling to provide the detail necessary to replicate and validate their methods and results. Many vendors simply asserted that their results would be "reasonably accurate" or "as accurate as the data allow." Only one vendor was willing to provide full information on the names and addresses that were determined to be matches.

For researchers who wish to replicate existing turnout validation methods, or for researchers whose concerns about omissions and errors in government records lead to questions about the accuracy and value of registration and turnout validation attempts, there is little available evidence with which to evaluate the quality of the vendors' offerings. Beyond knowledge of the lists of the variables most often used to match respondents to records and the assertion that some latitude is needed when deciding when a match is "close enough", little or no information is provided about specific procedures used. We therefore set out to compare the performance of a set of different plausible matching methods.

Matching Methods We Evaluated

We first matched respondents to government records using first and last names, addresses, and dates of birth. The California, Florida, New York, and Pennsylvania records

included month, day, and year of birth. The Ohio records included only year of birth. The North

Carolina records included only current age.

The matching process was conducted in three phases. First, potential matches between

individuals in the government records and all respondents were identified by Link Plus.[12] Link

Plus identified potential matches by linking records with the most similarities across names,

address, and birth date (or year or age). Similarities between last names were determined by

shared sequences of characters (e.g. "Robertson" and "Roberson" share the character sequences

"Rober" and "son", but not "t"). Shared character sequences were also used to identify similar

addresses. First name similarity was determined initially by exact matches, then by matching

formal names to common informal names, and last by the number of shared character sequences.

Finally, birth date similarities were determined by how much of the month, day, and year

information was the same. For states with only birth year or age, similarity was determined by

the numerical proximity of the two numbers (e.g. 1975 is numerically more proximal to 1976

than is 1972). The potential matches for a survey respondent included listings with identical

name, address, and birth date (or year or age) in both the ANES and government records, as well

as those with an exact match on only one of those fields.

The second phase involved scoring potential matches based on increasingly strict sets of

criteria. A potential match was scored as meeting the *least strict* criteria (LEAST) if the

information provided by a survey respondent and the information in a government record listing

had the same or similar first and last names and one of the following; (1) the same or similar

residential address with the same or similar birth date information, or (2) different addresses with

---

[12] Link Plus is a probabilistic record linkage software program developed at the Center for Disease Control's Division of Cancer Prevention and Control in support of the center's National Program of Cancer Registries (NPCR).

identical birth information.[13] Additional potential matches were also scored as meeting the least

strict criteria if they had identical last names, addresses, and dates of birth. Potential matches

were scored as meeting the *moderately strict* criteria (MOD) if information provided by a survey

respondent and a government record listing had the same or similar first and last names, had the

same or similar addresses, and had the identical birth date information. Matches scored as

meeting the *most strict* criteria (STRICT) had identical first and last names, identical addresses,

and identical birth date information.[14]

Results: Evaluating Registration Measures

    Comparing our matching rates to rates in prior studies. We began by comparing the

matching methods in terms of their matching rates. If two matching methods are equally

accurate, then the two methods should yield equal match rates. However, the converse need not

be true. Suppose that two matching methods match 65% of survey respondents to a government

record. This outcome can occur because both methods correctly matched 65% of the sample, or

because one or both of the methods matched 50% of the sample correctly and matched another

15% to the wrong records (i.e., survey respondents were matched to different people who had

similar attributes). In what follows, we offer comparative match rate information as an initial

way of clarifying the properties of the LEAST, MOD, and STRICT methods.

    In Table 9, we report match rates for LEAST, MOD, and STRICT and match rates

---

[13] Similar last names included minor differences that could easily be explained by a typographical error (e.g., Brown and Brow) and one name appearing as a portion of a hyphenated name in the other (e.g. Brown and Brown-Smith). Similar first names included minor differences that could be explained by a typographical error (e.g., Christopher and Christophere) and one formal and a comparable informal name (e.g., Christopher and Chris).   Similar addresses included minor differences that could be explained by typographical error (e.g., 123 Elm St and 123 Elm Ct). Similar birth information included minor differences that could be explained by a typographical error (e.g., 01/01/1990 and 01/11/1990). Date of birth was determined to be the same for North Carolina records (which did not contain any date of birth information) if the age listed in the government record was equal to, or one year greater than, the age of the ANES respondent.

[14] All of the potential matches that met one set of criteria also met the criteria that were less strict. Respondents matched to a government record listing that did not meet the least strict criteria were categorized as not matched.

produced in earlier ANES turnout validation studies and the match rate reported by

Ansolabehere and Hersh (2008) in their 2006 Cooperative Congressional Election Study's

(CCES) turnout validation effort.[15] Of our three methods, LEAST produced the highest match

rate (77.46%), and STRICT produced the lowest (45.57%). The differences between LEAST's

match rate and those of the ANES from 1976-1990 ranged from a low of 4.75% for the 1984

ANES study (72.71%) to a high of 16.65% for the 1990 ANES study (60.81%). LEAST's match

rate was 14.96 percentage points higher than that of the 2006 CCES study (62.50%). Overall,

LEAST matched a higher proportion of respondents to government records than did the other

methods described in Table 9.

　　　There are at least two possible reasons why prior ANES turnout validation studies

generated lower match rates than LEAST. First, ANES studies from 1984 to 1990 only

attempted to find government records for respondents who said, or implied, that they were

registered. In our 2008 validation exercise, we attempted to match all respondents. Second,

respondents in the 2008 study were asked explicitly if they were registered to vote. In most prior

ANES studies, respondents were asked this question only if they said they did not turn out.

Respondents who said they turned out were coded as if they were also registered.

　　　To clarify how such differences can affect various methods' match rates, the bottom of

Table 9 offers match rates for only respondents in each of the surveys who said, or implied, they

were registered.[16] The match rates for the 1976-1990 ANES studies averaged 84.98%, and the

---

[15] The CCES is a 30,000 person survey that measures attitudes within congressional districts in the United States. CCES respondents are drawn from a sample of people who volunteer to participate in surveys in exchange for payment. Ansolabehere and Hersh (forthcoming) limited matching to respondents living in one of the 26 states with the highest quality government records. A description of the variables used in CCES' validation attempt is available at http://web.mit.edu/polisci/portl/cces/commoncontent/Voter%20Validation%20Variables.pdf.

[16] The proportions of respondents who indicated they were registered and lived in a location for which government records were available were: 2008 ANES Panel Study=83.0%, ANES 1976=80.3%, ANES 1978=71.8%, ANES 1980=72.2%, ANES 1984=71.9%, ANES 1986=73.5%, ANES 1988=79.9%, ANES 1990=69.8%, CCES 2006=95.1%.

match rates for ANES studies conducted after 1980 were consistently higher than LEAST. Among the prior ANES studies, the 1984 study had the highest match rate (90.81%), and the 1976 study had the lowest rate (77.57%) for respondents who indicated they were registered. The match rate among respondents who said they were registered in the 2006 CCES study (64.90%) was more than 20 percentage points lower than the average of the prior ANES studies, and 15.02% higher than STRICT (49.88%). These results indicate that the 1974-1990 ANES turnout validation methods (which relied heavily on interviews with local officials and some judgment calls) often yielded the highest match rates among respondents who said they were registered. The CCES validation attempt (which relied on a commercial vendor) yielded the lowest match rate amongst such respondents.

With these initial figures in mind, it is important to note that larger match rates do not necessarily imply greater accuracy. A method may match a large proportion of respondents to records, but many of those respondents may be matched to the wrong records. Thus, the preceding results do not support any conclusions about which method described in Table 9 was most accurate. Rather, the preceding results indicate that LEAST produced match rates that were comparable to, or within the range of, rates found in previous validation studies, whereas MOD and STRICT produced lower match rates than previous studies. And Table 9 shows that different methods can yield very different results, yet it is not obvious *a priori* that one of those methods is necessarily more accurate than another.

Comparing the percent of survey respondents who were registered to the percent in the population who were registered. Next, we attempted to assess the accuracy of various matching methods by comparing official government registration rates with (1) the proportion of survey respondents who said they were registered, and (2) the proportion of survey respondents who

could be matched to a government record indicating that they were registered. A closer match of the survey results with the official government rates may be taken as a signal of greater accuracy.

However, such comparisons are again imperfect measures of accuracy. The proportion of "matched" and "registered" respondents could be the same as the government's reported registration rate because: (a) survey respondents accurately reported their registration status, (b) the sample of survey respondents was representative of the total population from which they were drawn, (c) selection for, or participation in, the survey did not influence a person's decision to register, and (d) the method of matching responses to government records was completely accurate. However, other circumstances can produce the same results. If, for example, a matching method was biased against finding government records for some respondents who were actually registered at the same time that the sample of survey respondents over-represented citizens who participated in an election, then the aggregate consequence could be that the proportion of "matched" and "registered" respondents is the same as the government's reported registration rate. Consequently, although the comparisons we are about to describe can be viewed as providing evidence that is useful for understanding the properties of different matching methods, factors (a) through (d) above can interact in ways that reduce the informative value of such aggregate comparisons.

The ANES 2008 sample was drawn from the population of U.S. citizens, aged 18 years or older, who lived in a household served by a landline. Aggregate data from different sources were available to estimate the registration rate in this population. Information on the number of U.S. citizens aged 18 years or older living in each state as of November, 2008 has been published by McDonald (2009). Data on the number of individuals living in households served by landline telephones has been collected during the National Health Interview Survey (Center for Disease

Control and Prevention, 2009), and each target state has published the number of people

registered to vote as of November, 2008 (California Secretary of State, 2009; Florida Department

of State, 2010; New York State Board of Elections, 2009; North Carolina State Board of

Elections, 2009; Ohio Secretary of State, 2009; Pennsylvania Department of State, 2010).   We

used these data to estimate the proportions of U.S. citizens, aged 18 years or older, living in

households served by a landline telephone as of November, 2008 who were registered to vote in

each target state.[17] We then compared these estimates to the proportions of survey respondents

who said they were registered and to the proportions of such individuals who could be matched

to a government record using LEAST, MOD, or STRICT.

Table 10 shows that respondent self-reports came closer than any matching method to the

registration rate that we derived from aggregate government statistics (see Table 10). The

aggregate government statistics imply that 87.82% of U.S. citizens, aged 18 years or older, living

in any of the target states and in households served by a landline telephone as of November,

2008 were registered to vote. The proportion of survey respondents living in any of those states

who said they were registered was 85.01%. The proportion of respondents living in any of those

states who said they were registered *in the state in which they lived* was somewhat smaller

(81.89%), and all of the registration rates based on attempts to match responses to government

records were even smaller, sometimes dramatically so (STRICT=45.57%; MOD=65.39%;

LEAST=77.46%). Even after adjusting the registration rates by eliminating estimated

deadwood, the self-reported registration rates across the target states were closer to the

---

[17] The number of U.S. citizens, aged 18 years or older, living in households served by a landline telephone was computed by subtracting the percentage of non-U.S. citizens in a state from the number of people aged 18 years or older in a state and multiplying that value by the proportion of individuals living in households served by a landline (.7978). The number of people registered to vote living in households served by a landline telephone was computed by multiplying the number of people registered by the proportion of individuals living in households served by a landline (.7978).

registration rate based on the government records (84.31%) than any of the rates based on matching.

If the ANES 2008 Panel Study respondents were representative of the target population, and if participating in the study did not influence registration, then these results are inconsistent with the claim that the matching methods produced more accurate registration measures than did the self-reports. If such assumptions were true, then the results are more consistent with the conclusion that self-reports of registration are more accurate than registration measures based on matches with government records.

Problems with using the government records to measure registration became even more apparent when comparing registration rates from state to state. In California, the estimated registration rate using LEAST was the most similar to the true rate (unadjusted=76.33%, adjusted=76.08%, LEAST=76.54%). In North Carolina, by contrast, LEAST produced a registration rate that was nearly more than 30 percentage points below both the unadjusted rate (unadjusted=93.05%, LEAST=58.43%) and the adjusted rate (adjusted=88.45%, LEAST=58.43%). Government records from Florida, New York, Ohio, and Pennsylvania also produced registration rates that fell below the true rate even when LEAST was used.

Table 10 has an important implication. A matching method that is effective for locating respondents' records in one state may be quite ineffective in other states. As Table 10 shows, no single method was most effective across all states, and no methods were equally effective across all states.

Differences in the performance of the three matching methods across states may be due to differences in the states' government records. In a recent analysis of government records, Ansolabehere and Hersh (2010) reported substantial differences across states in the proportions

of records with "undeliverable" or "probably undeliverable" addresses.[18] Those proportions

ranged from less than 2% in Maryland, Washington, and the District of Columbia to a high of

around 10% for West Virginia. Furthermore, McDonald (2007) documented substantial time

lags between the 2004 general election and when counties in some states reported turnout data to

the state. It is possible that between-state differences in the timeliness and accuracy of

government records, as well as differences in the types of information contained in those records,

contributes to the state-specific performance variations observed in Table 10.

However, actual differences among the government records challenge this assertion.

First, the proportions of "undeliverable" addresses reported by Ansolabehere and Hersh

(forthcoming) were unrelated to the match rates observed for different states. The proportion of

"undeliverable" addresses in the New York records was quite low (around 2%) and the

proportion in the Pennsylvania was above-average, yet both produced similar match rates. North

Carolina records were close to the national average (less than 4%), yet produced consistently low

lowest match rates for all three matching methods. California had the fifth lowest proportion of

such addresses and yet generated match rates that were close to the average of the six target

states. Second, McDonald (2007) reported no substantial reporting delays or high proportions of

missing data for North Carolina, which produced some of the lowest match rates of the six target

states.[19] Finally, the frequencies of missing, implausible, and suppressed data in the government

records from the different states reported in Table 7 are unrelated to the match rates. New York

and North Carolina had few missing, implausible, or suppressed names, addresses, and dates of

birth compared to the other states, whereas Florida had higher proportions for those variables,

---

[18] A commercial vendor used the Coding Accuracy Support System (CASS) process of the U. S. Post Office to
predict whether addresses were "undeliverable" or "probably undeliverable".
[19] The government records examined by McDonald (2007) were limited to 11 states and the District of Columbia
due to the availability and costs of those records. In addition to North Carolina, McDonald (2007) analyzed records
from Delaware, California, Florida, Iowa, Kentucky, Maryland, Ohio, Oklahoma, and South Carolina.

and California and Ohio both had high proportions of missing dates of birth. This evidence

therefore challenges the notion that problems in the government records were major contributors

to the variation in match rates across states.

Another possible explanation for the differences in Table 10 is differences between the

2008 Panel Study respondents living in various states in terms of attributes related to the

phenomena of interest here. First, respondents living in some states may have had characteristics

that made locating their government records more difficult than respondents living in other

states. For example, locating records for respondents who did not own a home may be more

difficult than locating records for homeowners, and home ownership may have been more

prevalent in some states than others. Second, respondents in some states may have been more

prone to omit information, or to provide inaccurate or problematic information, for the variables

used in matching. For example, respondents in some states may have been more prone to

provide a post office box when asked for their residential address. Both types of differences

could cause difficulties when attempting to locate records for respondents from some states

compared to others.

Some evidence suggests that such differences could explain variation in performance

shown in Table 10. Consider, for example, the low match rates in North Carolina. Table 11

shows that a higher proportion of survey respondents living in North Carolina failed to provide a

date of birth (12.36%) compared to the average across all target states (7.95%). North Carolina

respondents were also more likely to provide a post office box in lieu of a residential address

(7.87%) compared to respondents from the other five target states (3.62%). However, other

factors in Table 11 are not clearly associated with the Table 10 performance differences. Thus,

differences in the characteristics of, and information provided by, the ANES 2008 Panel Study

respondents across states is a more plausible explanation for the different match rates across states than are differences in the accuracy of different states' records.

    Inconsistencies between self-reports and government records of registration. If self-reports of registration are generally accurate and a matching method accurately locates government records, then there should be few inconsistencies between self-reports and government records. A government record of registration should be located for nearly every respondent who said he or she was registered, and no record should be found for nearly every respondent who said she or he was not registered.[20]

    As Table 12 shows, we found a very different pattern of results. Government records agreed with self-reports for only 52.88% of the 919 panel respondents who reported living, and being registered, in one of the six states using STRICT.[21] Amongst these matched records, self-reports and government records agreed that 45.38% of respondents were registered and 7.51% were not registered. Most inconsistencies occurred among respondents who said they were registered while the government records suggested they were not (46.57%). There were few instances in which a government record suggested a respondent who reported being not registered was actually registered (.54%).

    MOD resulted in fewer inconsistencies between self-reports and government registration records (27.31%). Compared to STRICT, there were fewer cases in which respondents said they were registered while government records suggested they were not (26.55%), and about the same number of cases in which government records suggested a respondent who reported being not

---

[20] Some government records may have been updated to reflect information that changed between the 2008 general election and the time when government records were obtained.

[21] The percentages reported in Table 10 do not reflect three of the 208 respondents who moved at least once between October of 2007 and December of 2009 and were matched to government records with a previous address using the least strict criteria. Including these three respondents would have increased the proportion of respondents who said they were registered and for whom government records suggested they were registered using the least strict criteria for all target states by .53% (from 76.22% to 76.75%). These three respondents would not affect the proportions for matches based on the moderately strict or most strict criteria.

registered actually was (.76%). Thus, relaxing the matching criteria appeared to successfully match more respondents a government record, but some of these methods may have matched more respondents who were not registered to someone else's record and thereby misleadingly suggested that they were registered.

Relaxing the matching criteria even more resulted in even closer apparent correspondence of self-reports with government records. Using LEAST, only 16.65% of the cases resulted in inconsistencies between self-reports and government records. The proportion of respondents who said they were registered while the government records suggested they were not dropped to 15.13%, and the proportion who reported they were not registered while the government records suggested they were increased very slightly to 1.52%. Allowing for typos, hyphenated last names, informal first names, and a major difference in one matching variable resulted in locating records for more respondents who may have been registered, but also matched more respondents who were probably not registered to someone else's record. Overall, no single matching method minimized both types of inconsistencies.

The effect of adding telephone numbers to the original criteria. In some states, government records include phone numbers. In this section, and in Tables 13 and 14, we show how adding the requirement of telephone number matches affects the performance of the matching methods. Government records from California, Florida, North Carolina, and Pennsylvania included telephone numbers which, when included among the different match criteria, could potentially reduce the number of respondents matched to the wrong records.

As Table 13 shows, adding telephone numbers to STRICT (i.e., a "match" now required identical first and last names, identical addresses, identical birth dates, and identical phone numbers) resulted in only one respondent who reported being not registered but was matched to

a government record suggesting she or he was registered (.15% of the 650 respondents from the four states). However, this small percentage was offset by large numbers of respondents who reported being registered but were not matched to a government record when telephone numbers were added to STRICT (77.38%), which resulted in a minority of cases for which self-reports and government records of registration were consistent (22.46%). Even when telephone numbers were added to LEAST, only one-third of self-reports and government records of registration were consistent (33.69%). Requiring identical matches between telephone numbers in addition to the original matching criteria nominally reduced the number of respondents who probably were not registered who were matched to a record suggesting they were. However, the additional criterion substantially increased the number of respondents who may have been registered and for whom no match was made.

As Table 14 shows, adding matches based on identical phone numbers to matches based on the original criteria somewhat improved the consistency between self-reports and government records of registration. Self-reports and government records agreed on a respondent's registration status for more than half of the cases in which the additional matches were added to those based on STRICT (54.00%). The increased consistency was due primarily to fewer cases in which government records were not located for respondents who reported being registered (45.38%). However, the number of cases in which records were matched to respondents who said they were not registered did increase somewhat (.62%). And these proportions were even higher when the additional matches were added to those based on LEAST (1.38%). Thus, adding matches based on same or similar names and identical telephone numbers to those based on the original increased the overall consistency between self-reports and government records of registration, but may have done so by matching a few more respondents who were probably not

registered to other people's records.

If self-reports of registration are generally accurate, the inconsistencies between self-reports and government records indicate why using government records to measure registration is problematic. Matching respondents to government records based on identical first and last names, address, date of birth, and phone number resulted in locating fewer government records. In the extreme, records were located for only 10 of the 129 respondents living in Pennsylvania who said they were registered in that state when using STRICT plus telephone numbers. The overall consistency between self-reports and government records can be increased by relaxing the matching criteria, but this occurs at the expense of matching records to respondents who probably were not registered. And even with the more relaxed criteria, the consistency between self-reports and government records varied considerably across the different states. Thus, no single matching method was equally effective in every state.

The effect of various validation methods on attempts to explain registration choices. Analysts often use validated information on respondent registration and turnout to study why some people decide to register and/or turnout while others do not. Using multiple regression, these analyses attempt to identify independent effects of a number of factors that are plausibly associated with registration and turnout. We have compared the three matching methods in the context of such regressions.

To do so, we began with the regression equation proposed by Rosenstone and Hansen (1993). We chose this starting point because their model has been referenced frequently in the academic literature on political participation. In their influential examination of participation in general elections from 1956 to 1988, Rosenstone and Hansen (1993) used 31 variables to explain turnout as well as other forms of political participation. Their regressions included many

explanatory variables, including variables measuring respondents' evaluations of parties and

candidates, party and campaign mobilization efforts, and select demographic variables. They

found that many of the explanatory variables had statistically significant relationships with

turnout.[22]

We used as many of their explanatory variables as were available in the 2008 ANES

Panel Study to attempt to explain individual registration decisions. Appendix D contains a

complete listing of these explanatory variables.[23]

Table 15 shows parameter estimates for four regressions that differed only in the measure

of registration that we used. The results in the first column use respondents' self-reports of their

registration status. The results in the second, third and fourth columns use the results of the

STRICT, MOD, and LEAST matching methods, respectively.

Table 15 shows that varying effects of many explanatory variables depending on which

measure of registration that we used. When we used self-reports of registration status,

registration had a statistically significant positive association with education (b=1.31, p<.001),

identifying strongly with a political party (b=.99, p<.0011), and caring about who won the 2008

general election (b=1.76, p<.001). Self-reported turnout was significantly lower for people with

a greater preference for one major party presidential candidate over the other (b=-.70, p<.05) or

those living in a southern state (b=-.39, p<.05).[24]

---

[22] Other researchers have attempted to explain registration (see, e.g., Erikson, 1981; Timpone, 1998; Uhlaner, 1989). Several of the explanatory variables used in such work are the same as, or similar to, those examined by Rosenstone and Hansen (1983), others are not.

[23] Some explanatory variables used by Rosenstone and Hansen (1993) were not available in the 2008 ANES Panel Study, and others were aggregate measures that would be constants among respondents living in the same region. These differences prevent meaningful comparisons between the results reported by Rosenstone and Hansen (1993) and those found with the 2008 ANES Panel Study, and have therefore been excluded from the present investigation.

[24] Rosenstone and Hansen (1993) computed Affect for a presidential candidate as the absolute value of the difference between the number of Democrat candidate "likes" plus the number of Republican candidate "dislikes" and the number of Republican candidate "likes" plus the number of Democrat candidate "dislikes". Affect for a presidential candidate in the current study was computed as the absolute value of the difference between how much

When we used results from the STRICT matching method, registration had significant positive associations with caring about who won the 2008 general election (b=.67, p<.01), owning a home (b=.34, p<.01), and perceived closeness of the election (b=.23, p<.05). When we used MOD, registration had significant positive associations with caring about who won the election (MOD: b=1.11, p<.001) and perceived closeness of the election (MOD: b=.31, p<.01). When we used LEAST, registration had significant positive associations with education (LEAST: b=.63, p<.05) and caring about who won the election (LEAST: b=1.08, p<.001). Hence, conclusions about which factors influence registration will depend on whether one predicts self-reports or government records and the particular matching method used.

Our analyses indicate that the results of attempts to explain registration statistically are not independent of decisions about which methods a registration validator uses to match respondent information to government records. For example, effects for some of the explanatory variables were stronger when we used self-reports than when we used registration variables derived from the government matching methods. Education had a stronger effect on self-reported registration than it did on STRICT and MOD measure of registration (difference$_{\text{self-report vs. STRICT}}$= 1.49, p<.001; difference$_{\text{self-report vs. MOD}}$= .89, p<.05), while strength of party identification had a stronger relation with self-reports of registration than with all other measures (difference$_{\text{self-report vs. STRICT}}$= .89, p<.01; difference$_{\text{self-report vs. MOD}}$= .90, p<.01; difference$_{\text{self-report vs. LEAST}}$= .67, p<.05). Caring about who won the presidential election had a stronger effect on self-reported registration than it did on registration measures produced by STRICT (difference$_{\text{self-report vs. STRICT}}$= 1.09, p<.01). Moreover, results also varied across the three matching methods. Education had a

---

respondents liked or disliked the Democrat candidate and how much they liked or disliked the Republican candidate (see Appendix D).

stronger effect when we used LEAST than when we used STRICT (difference=.81, p<.01).[25]

Caring about who won the presidential election had a stronger effect on MOD than on STRICT (difference=1.11, p<.01). The effects of education were different for MOD and STRICT (difference=.60, p<.01), although neither of the effects were statistically significant. This was also the case for the effects of unemployment on MOD and STRICT (difference=.58, p<.01).

Thus, Table 15 provides evidence that when attempting to explain registration as a function of explanatory variables commonly identified in the academic literature, the results will depend not just on the explanatory variables included in the equation but also on the assumptions of people who attempt to validate self-reports of registration. For analysts working in this area, questions about the validity of assumptions underlying registration validation exercises are not merely an academic curiosity, the validity of these assumptions may well determine the accuracy and reliability of their own claims.

Comparing predicted registration based on self-reports and government records.
Regressions using different measures of registration may lead to different explanations of which types of people were most likely to be registered in the population. That is, an equation using self-reported registration might find that some types of people are most likely to be registered, whereas an equation using government records might suggest that these same people are unlikely to be registered, and *vice versa*. We explored the similarities of regression estimates generated using self-reports, STRICT, MOD, and LEAST. For each respondent, we computed four different registration estimates, with each estimate generated using the parameter estimates from one of the four equations in Table 15. We then computed the proportion of respondents for whom a given pair of models made the same prediction about whether or not a particular

---

[25] The effect for MOD was also significantly different than the effect for STRICT, although neither coefficient was statistically significant.

respondent was registered. Predicted registration using the parameters of each equation was

coded as "registered" if the equation indicated the probability of being registered was equal to or

greater than .5, and "not registered" if the probability was less than .5.

The lowest agreements in predicted registration status occurred between the models

predicting LEAST and STRICT (44.50%) and the models predicting self-reports and STRICT

(47.07%). By contrast, the models predicting self-reports and LEAST produced identical

registration predictions 93.52% of the time (both predicted "registered" for 91.56% of

respondents and "not registered" for 1.96%). Agreement between predictions based on MOD

and the other registration measures fell between the two extremes (MOD and self-report=90.34%

agreement, MOD and LEAST=92.91% agreement, MOD and STRICT=51.59% agreement).

Despite some differences in the effects of explanatory variables in the self-report and

LEAST regressions, these two registration measures produced very similar predictions about

which respondents were registered. However, models based on STRICT or MOD were in less

agreement. Again, the lesson here is that how registration is measured can have important effects

on the conclusions we would reach about registration.

Results: Evaluating Turnout Measures

Comparing the estimated percent of respondents who turned out to the percent of the

population who turned out. In this section, we attempt to assess the accuracy of various

matching methods by comparing official government turnout rates with the proportion of survey

respondents who said they turned out to vote in the 2008 general election and the proportion of

survey respondents who could be matched to a government record that indicated that they turned

out to vote.

We again used data from McDonald (2009), the National Health Interview Survey

(Center for Disease Control and Prevention, 2009), and each target state (California Secretary of State, 2009; Florida Department of State, 2010; New York State Board of Elections, 2009; North Carolina State Board of Elections, 2009; Ohio Secretary of State, 2009; Pennsylvania Department of State, 2010) to estimate the turnout rates for U.S. citizens, aged 18 years or older, who lived in any of the target states and who lived in households served by a landline telephone as of November, 2008. We then compared these estimates to the proportions of survey respondents who said they turned out and to the proportions of such individuals who could be matched to a government record using LEAST, MOD, or STRICT.

As Table 16 shows, self-reported turnout was consistently higher than reported in official government tallies. LEAST, MOD, and STRICT, by contrast, sometimes overestimated and at other times underestimated the actual turnout rate. Across all target states, the turnout rate based on respondent self-reports was 15 percentage points higher than the actual rate (actual=63.01%, self-reported=78.47%). The opposite was true for the turnout rate produced by STRICT (39.64%), which was 23 percentage points lower than the actual rate. MOD produced a higher rate than STRICT (56.34%), but even this was six points lower than actual rate. Of the four turnout measures, LEAST (65.90%) yielded estimated rates closest to the actual rate across all states.

However, no single matching method produced rates that were closest to actual rates in every state. LEAST's turnout measure was closest to the government figure for New York (actual=58.58%, LEAST=60.13%), and North Carolina (actual=65.82%, LEAST=37.08%). However, MOD's turnout measure was closest to the government rates in California (actual=61.36%, MOD=59.82%), Florida (actual=66.15%, LEAST=62.32%), Ohio (actual=67.07%, MOD=62.96%), and Pennsylvania (actual=63.19%, MOD =61.59%).

Variability across states in the effectiveness of a matching method in locating government records may contribute to the variability observed in Table 16. Inabilities to locate records for respondents who did turn out would lead those respondents to be coded as having not turned out, and this problem could be more pronounced for some states than others. To evaluate this possibility, we compared official turnout rates to the turnout rates produced by looking only at respondents who were matched to a government record. In this case, the appropriate actual turnout rate to use is the proportion of registered voters who turned out in the 2008 election. Those rates were available through data published by McDonald (2009), the National Health Interview Survey (Center for Disease Control and Prevention, 2009), and each target state (California Secretary of State, 2009; Florida Department of State, 2010; New York State Board of Elections, 2009; North Carolina State Board of Elections, 2009; Ohio Secretary of State, 2009; Pennsylvania Department of State, 2010). We also adjusted the turnout rates using the estimated deadwood numbers available from Ansolabehere and Hersh (2010) to account for obsolete records. For self-reports, the appropriate comparison rate would be based on turnout among only people who reported being registered.

Table 17 shows that turnout rates based on self-reports or any matching method were considerably higher than either of the rates based on official statistics. The official turnout rate among registered individuals across the target states was 71.75%. This rate increased to 74.74% after eliminating estimated deadwood in the government records. All turnout rates based on self-reports and government records were at least 13 points higher than either of the official rates (self-reports=94.71%, STRICT =89.85%, MOD=89.23%, LEAST=88.44%). In this case, both self-reports and government records indicate that turnout was higher among registered respondents in the sample than among registered individuals in the population.

With the exceptions of North Carolina, the turnout measures culled from self-reports and from STRICT, MOD, and LEAST were very similar. In California, Florida, New York, Ohio, and Pennsylvania, no two rates differed by more than eight points. The difference between self-report and government records turnout rates differed by as much as 28 points for North Carolina, and that difference can be tied to the matching challenges described above. Overall, these analyses provide evidence that registered individuals who participated in the Panel Study were more likely to turn out to vote than registered individuals who did not participate in the survey.

Inconsistencies between self-reports and government records of turnout. The similar turnout rates among respondents who said they were registered and respondents matched to a government record suggest that self-reports of turnout were generally accurate. If this was so, and if a method of locating government records accurately matched respondents to records, then there should be few inconsistencies between what an individual said they did on Election Day and government records of their behavior. But as Table 18 shows, this was clearly not the case.

Self-reports and government records of turnout agreed for only 51.24% of the respondents when STRICT was used. Most of the inconsistencies were due to the inability to match a government record to a respondent who reported turning out for the election (48.76% of all respondents). LEAST increased the overall consistency between self-reports and government records (79.46%). Interestingly, none of the matching criteria resulted in any cases in which a respondent who reported not turning out was matched to a record suggesting he or she did turn out. Thus, using more relaxed matching criteria to locate records for respondents who may have turned out did not result in government records suggesting turnout among respondents who may not have turned out.

The pattern of inconsistencies between self-reports and government records of turnout

reveal important variations in the relative performance of LEAST, MOD, and STRICT. Whereas LEAST produced turnout inconsistencies for less than one quarter of respondents with an address in California (19.81%), Florida (8.94%), New York (21.37%), Ohio (17.21%), or Pennsylvania (21.14%), it resulted in inconsistencies for 44.30% of respondents in North Carolina. As reported earlier, LEAST resulted in few registration inconsistencies for California (16.08%), Florida (10.24%), New York (17.02%), Ohio (13.28%), and Pennsylvania (16.67%), but many more for North Carolina (33.75%). No single method was equally accurate in every state.

In turnout validation, if no government record was located for a respondent, then that respondent has usually been coded as not registered and consequently as having not turned out. This practice can result in many respondents who actually did turn out being coded as having not turned out because a matching method could not locate their records. To evaluate this possibility, we now compare self-reports and government records of turnout for only respondents for whom a government record was located.

As Table 19 shows, there were many fewer inconsistencies between self-reports and government records among respondents matched to a government record. Across all states, STRICT resulted in inconsistencies for only 7.28% of the cases. This proportion was similar to that of MOD (7.35%) and LEAST (7.54%). Similar results were also obtained when looking at individual states. The proportion of respondents within a state whose self-reports were inconsistent with their government records of turnout never differed by more than four points across the different matching methods.[26] Regardless of the matching method used, once a record was located for a respondent, that record tended to be consistent with the respondent's self-

[26] The largest difference occurred among respondents with an address in Pennsylvania. The least strict criteria resulted in 8.49% of those respondents with inconsistencies between self-reports and government records, while the proportion for the moderately strict criteria was 4.82% (difference=3.67%).

reported turnout.

Using respondent attributes to explain turnout. Next, we compared results obtained

using a Rosenstone-Hansen-like regression to explain turnout. As was the case when we used the

same technique to estimate registration, our findings depended on which turnout measure we

used.

In Table 20, we can see that self-reported turnout had a positive and significant

association with education (b=.92, p<.01), identifying strongly with a political party (b=1.02,

p<.001), and caring about who won the presidential election (b=1.85, p<.001). Turnout based on

matching a self-report to a government record had a positive and significant association with

education (b_{LEAST}=.51, p<.05), caring about who won the presidential election (b_{STRICT}=.74,

p<.001; b_{MOD}=1.01, p<.001; b_{LEAST}=.89, p<.001), and homeownership (b_{STRICT}=.33, p<.05;

b_{LEAST}=.32, p<.01), while having a significant negative association with living in a Southern

state (b_{MOD}=-.23 p<.05). Only caring about who won the presidential election had significant

relations with turnout regardless of which turnout measure was used, and these relations varied

depending on which turnout measure we used.

As Table 20 indicates, several of the significant relations were stronger when we used

self-reported turnout than when we used the turnout measures based on STRICT, MOD, or

LEAST. The converse never occurred. Moreover, as was the case for government records of

registration, none of the methods for matching respondents to government records produced a

measure of turnout that led to consistently larger and statistically significant coefficients for the

explanatory variables.

Differences in predicted turnout. As was the case for registration, we used the four

equations reported in Table 20 to generate predictions about turnout in the 2008 general election.

Predicted turnout from each model was coded as "turned out" if the equation indicated the

probability of turning out was equal to or greater than .5, and "did not turn out" if the probability

was less than .5.

The largest difference in predicted turnout occurred between predictions generated by

self-reported turnout and STRICT. Those two equations produced identical turnout predictions

for only 30.32% of the respondents. Agreement between predictions based on LEAST and

STRICT was almost as low (32.64%). By contrast, predictions derived from self-reports and

from LEAST yielded identical predictions 88.88% of the time (both models predicted "turned

out" for 84.60% of respondents and "did not turn out" for 4.28%). The agreement rates of other

comparisons fell between these extremes (MOD and self-reports=81.05%, MOD and

STRICT=43.89%, MOD and LEAST=88.26%). As was the case for predictions of registration

status, predictions about who turned out varied considerably depending on which turnout

measure was used.

Discussion

This report has identified several problems with using government records of registration

and turnout instead of survey self-reports to generate measures of registration and turnout. We

found that some listings in government records did not contain names, addresses, or birth date

information while other listings contain information that is probably inaccurate. These

inaccuracies and omissions, along with errors that undoubtedly exist in the information provided

by respondents, the use of formal and informal names, and name and address changes all

introduce difficulties into attempts to conduct registration validation or turnout validation by

matching respondents to government records.

Problems in locating respondents' government records resulted in estimated registration

rates for the survey sample that were implausibly low when compared to the registration rates reported in states' official statistics. In contrast, self-reports yielded registration rates that were quite similar to the government-reported registration rates for the various states. This similarity suggests that either self-reports of registration are generally accurate, or that the sample was biased in favor of respondents who were not registered. Given that the sample recruitment did not focus on politics, it seems unlikely that the recruitment method resulted in a disproportionate number of unregistered individuals. The more likely explanation is that the sample generally resembled the population, and that self-reports of registration were generally accurate.

Additional evidence that self-reported turnout may be accurate came from analyses using only respondents who said they were registered or for whom a government record was located. These analyses found considerable similarity between self-reports and government records of turnout. This consistency suggests that even the least strict matching criteria may have failed to locate government records for some respondents who were registered and did turn out in the 2008 general election. Failures to locate records for these respondents probably resulted in them being inaccurately coded as not registered and not turning out, which in turn produced inaccurately low turnout rates in the entire sample. Thus, even the least strict matching criteria may have generated registration and turnout measures with more error than self-reports.

The similarity between registration rates based on self-reports and true registration rates emerged despite some logical inconsistencies in the self-reports. Some respondents who said they were registered during one interview then said they were not registered in a later interview. Although some respondents may have cancelled their registration before the election or learned at some point during the Panel Study that they were not registered, the inconsistencies suggest that self-reports are not always accurate. But even with some inaccuracies in self-reports of

registration, those reports reproduced registration rates closer to official government figures than did attempts to match survey respondents to government records.

Difficulties in locating government records caused problems for attempts to measure and explain turnout. Turnout rates based on strict matching criteria substantially underestimated turnout rates published by the states. Turnout rates based on more relaxed matching criteria came closer to reproducing the true rate, while the rate based on self-reports overestimated the true rates.

When turnout rates were limited to respondents who said they were registered or were matched to a government record, both self-reports and government records resulted in turnout rates that were larger than the overall turnout rate among registered individuals. The similarity among the turnout rates for these subsamples offer converging evidence that turnout among the ANES 2008 Panel Study respondents was truly higher than turnout in the general population. If this is true, then the higher turnout rate reported by Panel Study respondents compared to the true turnout rate need not be the result of respondent lying and may accurately reflect the behavior of those respondents.

It may be serendipitous that registration-validated and turnout-validated data can occasionally produce turnout rates that come close to the population rates. Turnout among the panel respondents appeared to be higher than turnout in the general population, but failure to locate records for respondents who were registered and did turn out could lead to many respondents being inaccurately coded as having not turned out. The higher turnout among the respondents in combination with inaccurate coding based on government records could cancel each other out. As a result, the turnout rate from registration-validated and turnout-validated data may happen to fall close to the actual rate.

Overall, these findings cast doubt on the notion that validated data based on government records can and should be used instead of self-reports. Turnout does appear to be higher among respondents who participated in the Panel Study compared to those who did not. This could be a result of respondents completing several surveys about the election prior to Election Day, and such experiences leading respondents to register and turn out. Higher turnout among panel respondents could also be the result of a sample recruitment process that was more effective in attracting people who were especially likely to turn out. In either case, the evidence suggests that self-reports of registration and turnout were more accurate descriptions of respondents' behavior than has been previously believed.

One Size Does Not Fit All

We also found large differences between the states in the proportions of respondents matched to government records.   Some of these differences appeared to result from different characteristics of the populations of different states. For example, matching rates for states with proportionately fewer homeowners were lower than rates for states in which more residents owned their home. Respondents from one of the states were less likely to provide a date of birth, and more likely to provide a post office box in lieu of a residential address than respondents from the other states; and that state generated comparatively low match rates.

If a researcher assumes that using government records produces more accurate turnout measurement than self-reports and uses a single method for matching respondents in different states to government records, problems are likely to arise. Had we adopted such a perspective, the unusually low match rates we found for respondents living in North Carolina and New York using even the least strict criteria would suggest that those respondents are much more likely to over-report registration and turnout than respondents from other states. There is no apparent

reason why people living in some states would be so much more likely to misreport registration and turnout than people from other states. However, reliance on a single method for matching respondents to records and an assumption that reliance on government records produces accurate measures would require a researcher to reach such conclusions about respondents from different states.

This report suggests an alternative possibility. Namely, any method that effectively and accurately matches respondents from one state to government records will not necessarily be as accurate in all other states. Both the ANES 2008 Panel Study and the CPS data for November 2008 indicated that there are real demographic differences among people living in different states, and these demographic differences may impact the effectiveness of various matching methods. For example, proportionately fewer people living in California and New York own their homes compared to people living in other states, and matching homeowners to records may be easier than matching people who do not own their homes. Thus, in order to have fairly equivalent measures of registration and turnout among different states, researchers may need to validate the relative accuracy of different matching methods within each state.

The Measure Matters

Self-reports and government records can lead to very different conclusions about who in a population is likely to be registered and turn out in an election. The demographic and attitudinal variables that explain self-reported registration and turnout can be quite different from those that explain such variables when they are created by attempts to match survey respondents to government records. We found such differences to be especially large when we compared self-reports of registration and turnout to estimates of such behaviors based on MOD and STRICT. Most of the discrepancies were due to models based on self-reports predicting a

respondent would be registered or turn out, while models based on government records predicted that the respondent would not be registered or would not turn out. This means that explanations of turnout based on most strict and moderately strict matching criteria lead to much lower predicted rates of registration and turnout than predictions based on self-reports.

The Impact of Survey Mode

Our conclusion that turnout reports in the 2008-009 ANES Panel Study were highly accurate reports of those respondents' experiences might seem to contradict evidence reported by Holbrook and Krosnick (2010). Those investigators used the Item Count Technique (ICT) and found that turnout reports were indeed intentionally inflated due to social desirability response bias. However, that evidence came from a telephone survey and therefore converges with other evidence suggesting that social desirability pressures can indeed distort reports collected by telephone (e.g., Chang & Krosnick, 2009, 2010; Holbrook, Green, & Krosnick, 2003). The 2008-2009 ANES Panel Study involved data collected exclusively via self-completion on computers and submitted via the Internet, a mode that has been shown to be largely immune to social desirability pressures in reports of turnout (Holbrook & Krosnick, 2010). Therefore, the evidence reported here of high accuracy is in harmony with Holbrook and Krosnick's (2010) conclusion about computer administration. We look forward to future research exploring these same issues with self-reports collected via face-to-face interviewing, a mode in which data collection is also relatively resistant to social desirability pressures (Holbrook, Green, & Krosnick, 2003).

Advice for Analysts Who Are Considering Using Commercial Vendors to Obtain Government Registration and Turnout Records

Researchers interested in using government records of registration and turnout have the

options of relying on commercial vendors to match their respondents to the records, or purchasing government records and conducting the matching themselves. Several commercial vendors maintain databases that combine government registration and turnout records with data obtained from other sources (e.g. driver's license and product information). Some vendors' combine registration and turnout records with information from additional government (e.g. vehicle registration) and commercial (e.g. product registration) sources. Several vendors limit their records to individuals living in particular states or geographic regions, while others obtain records from as many states as is legally possible.[27] Some vendors deal exclusively with political campaigns and political groups, while others target marketing departments of businesses. There are many options available for researchers interested in obtaining additional information about survey respondents. Appendix C lists information about different commercial vendors that were contacted about the current project.

Although relying on commercial vendors might seem to be an attractive and efficient alternative to conducting one's own registration validation or turnout validation exercise, we have identified several problems with such a strategy. In what follows, we review benefits and problems associated with the use of commercial vendors.

Benefits of using a commercial vendor. Using a commercial vendor to locate government registration and turnout records for survey respondents has three main advantages over purchasing the records and conducting the matching themselves. One advantage of using a commercial vendor is that many vendors have developed some expertise in matching records to individuals based on names, addresses, dates of birth, and other identifying information. Vendors

---

[27] Some states limit distribution of their voter lists to very select groups for very specific purposes. For example, Illinois will only release their records to political candidates or political officials for bona fide political purposes. Similarly, Virginia will only release their records to state courts, political candidates, political incumbents, political parties, political action committees, and non-profit groups; and uses are limited to jury selection, political campaigning, reporting to constituents, promoting voter registration, or other political purposes.

who have such expertise may be able to quickly locate records for respondents. Researchers

attempting to match their respondents to government records must develop their own matching

criteria, as well as develop or learn and customize a computer application capable of locating

records using those criteria. Relying on the experience and resources of a group that has a

history of locating records for individuals can simplify this task.

      Another advantage is that several commercial vendors are aware of differences with

which different jurisdictions report turnout histories to the state. As such, they are able to update

only those jurisdictions within a state when those jurisdictions reported new information.

Vendors may also be aware of which jurisdictions do not contain accurate information for an

election of interest, and can therefore exclude the relevant records from analyses. While

researchers could certainly familiarize themselves with, and account for, differences among

jurisdictions, such activities would require additional time and expenses for researchers.

      Relying on a commercial vendor may also yield considerable monetary savings. The

states vary widely in the price they charge for their government records. Some states make their

records available to qualified people or groups free of charge (e.g., Wyoming), while others

charge in excess of $10,000 (e.g., Wisconsin). The cost to obtain records from every state that

makes its records available to researchers following the 2008 general election would have run in

excess of $90,000, and even at that expense a researcher would be unable to purchase records

from every state (see Table A2 for information about the costs and restrictions for government

records from the different states). Most commercial vendors charge based on the number of

records to be matched, which could have totaled less than $500 for the 2008 ANES Panel Study.

      Problems with relying on commercial vendors. Despite the benefits that relying on a

commercial vendor can produce, there are also serious risks associated with this route. We

convey some of these results in the form of lessons learned from our own attempts to work with a commercial vendor to conduct the kinds of analyses described in this report.

The first problem we encountered stems from the fact that we needed to know that any government reports that a commercial vendor would use were accurate as of the 2008 general election. Such knowledge can come from vendors providing detailed information about how and when they obtained records from each state. While some vendors were willing to disclose the date on which they obtained government records, many others were not. Moreover, some vendors were unwilling to disclose the sources from which they obtained their records. Indeed, several less expensive vendors were only willing to give "verbal assurances" that their records are "up to date."

Given that methodological transparency and accuracy of the records were of prime importance to us in our evaluation, vendors who were unwilling to commit in writing as to the date on which they obtained their government records were immediately eliminated from our list of firms with whom we could work.

A second problem in relying on a commercial vendor is the transparency with which government records were matched to respondents. All vendors indicated what identifying information they would need for each respondent in order to locate her or his records. This information typically included name, address, and date of birth, but occasionally included sex or race. However, no vendor was willing to disclose the method or procedure they would use to locate respondents' records. Vendors viewed such methods and procedures as proprietary. They would not provide the level of methodological detail that responsible scholars require as a way of protecting their own credibility and the reliability of their estimates. Given that the process by which records were located for respondents required enough detail so as to enable future

replications, vendors who were unwilling to fully disclose their processes were eliminated from

our list of firms with whom we could work.

These requirements, full disclosure regarding the dates on which, and sources from

which, government records were obtained and full disclosure regarding the method used to

match respondents to records, prevented us from working with any of the commercial vendors

that we interviewed. Consequently, we obtained all government records directly from the state

or through a known third party who obtained them directly from the state on a known date. All

matching was conducted in-house.

Despite our inability to locate a commercial vendor that met all our requirements, the

benefits of having a vendor match respondents to government records might remain an attractive

option. The findings described in this report suggest that researchers interested in pursuing this

option should consider the following questions when choosing a vendor for such matching.

- Is the vendor willing to fully disclose all information about the government records they use, including when and from whom they obtained those records?

- Does the vendor monitor when different jurisdictions within a state report turnout histories, and update only those jurisdictions when new information becomes available?

- Will the vendor describe the method they use to match respondents to government records with enough detail so as to enable conceptual replications of that method in future research by other individuals?

- Does the vendor use different strategies for matching respondents from different states to government records that control for differences in the structure and content of the states' records?

- Does the vendor use different strategies for matching respondents from different states to government records to control for differences in the characteristics of respondents living in different states?

The more of these questions that are answered in the negative for a vendor, the less

confidence a researcher should have in using that vendor to match respondents to government

records.

Even after contemplating these questions, it is reasonable to ask whether a vendor's product would yield results that better reflect actual registration and turnout behavior than the results described in this memo. One possible reason why this could be true involves the mistakes in government records that we have documented: about .57% of the government records had missing or implausible names, addresses, or dates of birth (see Table 7). Some commercial vendors use records from non-government sources in an attempt to fix these mistakes. Perhaps such attempts would increase the proportion of respondents who can be correctly matched to government records.

Judging from one recent instance, however, such attempts do not yield results that better reflect actual registration and turnout behavior than the results described in this memo. The CCES, for its 2006 study, employed a commercial vendor whose procedures were designed to fix mistakes in government records. This commercial vendor matched 62.5% of the 2006 CCES respondents to government records (see Table 9). This is about 15 points less than the officially reported 77.4% registration rate for the voting eligible population.[28] The least strict matching criteria we used matched 73.4% of the respondents to government records. This was about 13 points less than the officially reported 86.1% registration rate in the target states.[29] In other words, procedures designed to fix mistakes in government records did not produce a match rate that came closer to the official registration rate than the least strict matching method we used. Therefore, in this instance, fixing mistakes in government records did not improve upon the

---

[28] The registration rate for the voting age population is based on the total U.S. voting age population in 2006 published by McDonald (2011a) and the total number of registered individuals in 2006 (McDonald, 2011b).
[29] According to McDonald (2011b), the registration rate increased only 6.9% from 2006 to 2008. The 8.7% difference between the 2006 and 2008 registration rates reported above is due to that increase plus: (1) the exclusion of non-U.S. citizens from the 2008 ANES Panel Study sampling population, and (2) limiting the rate to the six target states.

results reported in this memo.

Conclusion

Survey self-reports of registration and turnout clearly contain errors. So do government records of registration and turnout. These latter errors create problems in locating government records for some survey respondents. Such problems also reduce the ability of record matching to generate improved measures of registration and turnout. Indeed, given these problems, it is inappropriate to simply conclude that any failure to locate a respondent's government record is due to misreporting information during a survey. Likewise, it seems inappropriate to conclude that any government record located for a survey respondent is actually that respondent's record.

What, then, should researchers do? The most economical course of action would be to accept what survey respondents say about registration and turnout as generally accurate. The proportion of survey respondents who say they are registered appears to accurately reproduce actual registration rates. And whereas self-reports of turnout produce rates that are higher than turnout in the population, this appears to occur because the turnout rate among respondents is actually higher than the rate in the population. Despite the occasional inaccuracies in self-reports, our results suggest that self-reports of registration are at least as accurate as those produced by a variety of turnout validation methods. Using government records in lieu of self-reports, which can be both time-consuming and expensive, appears to inject more error than accuracy into measurements of registration and turnout.

References

Aarts, K. (2002). *Electoral turnout in West-European democracies*. Paper presented at the annual meeting of the American Political Science Association, Boston, MA.

Abelson, R. P. (1990). *Results of vote validation experiment*, NES Pilot Study Report Series (Document. nes002276).

Abelson, R. P. (1996). *Results of experiment on improving the accuracy of self-reported turnout*, NES Pilot Study Report Series (Document nes002262).

Abramson, P. R., & Claggett, W. (1984). Race-related differences in self-reported and validated turnout. *Journal of Politics, 46*, 719-738.

Abramson, P. R., & Claggett, W. (1986). Race-related differences in self-reported and validated turnout in 1984. *Journal of Politics, 48*, 412-422.

Abramson, P. R., & Claggett, W. (1989). Race-related differences in self-reported and validated turnout in 1986. *Journal of Politics, 51*, 397-408.

Abramson, P. R., & Claggett, W. (1991). Racial differences in self-reported and validated turnout in the 1988 presidential election. *Journal of Politics, 53*, 186-197.

Abramson, P. R., & Claggett, W. (1992). The quality of record keeping and racial differences in validated turnout. *Journal of Politics, 54*, 871-880.

Andolina, M., Keeter, S., Zukin, C., & Jenkins, K. (2003). *A guide to the index of civic and political engagement*. College Park, MD: The Center for Information and Research on Civic Learning and Engagement.

Ansolabehere, S., & Hersh, E. (2008). Vote validation in the 2006 CCES. Unpublished Manuscript. Retrieved from

http://web.mit.edu/polisci/portl/cces/material/website%20paper%20changes%2010_3_08 .pdf.

Ansolabehere, S., & Hersh, E. (2010). *The quality of voter registration records: A state-by-state analysis*. (Voting Technology Project Report). Retrieved from http://www.vote.caltech.edu/drupal/files/report/quality_of_voter_report_pdf_4c45d05624 .pdf.

Ansolabehere, S., & Hersh, E. (Forthcoming). Who really votes? In P. M. Sniderman and B. Highton (Eds.) *Facing the Challenge of Democracy: Explorations in the Analysis of Public Opinion and Political Participation*. Princeton, NJ: Princeton University Press.

Belli, R. F., Traugott, M. W., & Beckmann, M. N. (2001). What leads to voting overreports? Contrasts of overreporters to validated voters and admitted nonvoters in the American National Election Studies. *Journal of Official Statisticsm, 17*, 479-498.

Bernstein, R., Chadha, A., & Montjoy, R. (2001). Over-reporting voting: Why it happens and why it matters. *Public Opinion Quarterly, 65*, 1–22.

Blais, A., Gidengil, E., Nevitte, N., & Nadeau, R. (2004). Where does turnout decline come from? *European Journal of Political Research, 43*, 221–236.

Brockington, D., & Karp, J. (2002). *Social desirability and response validity: A comparative analysis of over-reporting turnout in five countries*. Paper presented at the annual meeting of the American Political Science Association, Boston, MA.

Burden, B. C. (2000). Voter turnout and the National Election Studies. *Political Analysis*, *8*, 389–398.

California Secretary of State. (2009). *Voter participation statistics by county*. Retrieved from http://www.sos.ca.gov/elections/sov/2008_general/

Cassel, C. A. (2004). Voting records and validated voting studies. *Public Opinion Quarterly, 68*, 102-108.

Center for Disease Control and Prevention. (2009). National Health Interview Survey [2008 Data Release]. Retrieved from http://www.cdc.gov/nchs/

Chang, L., & Krosnick, J.A. (2009). National surveys via RDD telephone interviewing vs. the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly, 73*, 641-678.

Chang, L. & Krosnick, J. A. (2010). Comparing oral interviewing with self-administered computerized questionnaires: An experiment. *Public Opinion Quarterly*, *74*, 154-167.

Claggett, W. (1990). Reported and validated voter registration. *American Politics Quarterly, 18*, 197-207.

Clausen, A. R. (1968-69). Response validity: Vote report. *Public Opinion Quarterly*, *32*, 588–606.

Corbett, M. (1991). *American Public Opinion*. White Plains, NY: Longman.

Deufel, B. J., & Kedar, O. (2010). Race and turnout in U. S. elections: Exposing hidden effects. *Public Opinion Quarterly*, *74*, 286-318.

Duff, B., Hanmer, M. J., Park, W., & White, I. K. (2004). *How good is this excuse?: Correction of overreporting of voter turnout bias in the 2002 American National Election Study*, NES Technical Report Series (Document nes010872).

Dyck, J. J. & Gimpel, J. G. (2005). Distance, turnout, and the convenience of voting. *Social Science Quarterly*, *86*, 531-548.

Erikson, R. S. (1981). Why do people vote? Because they are registered. *American Politics Research*, *9*, 259-276.

Florida Department of State. (2010). *Voter turnout statistics*. Retrieved from

    http://election.dos.state.fl.us/statistics.shtml.

Fullerton, A. S., Dixon, J. C. & Borch, C. (2007). Bringing registration into models of vote

    overreporting. *Public Opinion Quarterly*, 71(4), 649-660.

Gerber, A. S., & Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail

    on voter turnout: A field experiment. *The American Political Science Review*, *94*, 653-

    663.

Gimpel, J. G. & Schuknecht, J. E. (2003). Political participation and the accessibility of the

    ballot box. *Political Geography*, 22(6), 471-488.

Green, D. P., & Gerber, A. S. (2005). Recent advances in the science of voter mobilization.

    The Annals of the American Academy of Political and Social Science, *601*, 6–9.

Haspel, M., & Knotts, H. G. (2000). Location, location, location: Precinct placement and the

    costs of voting. *The Journal of Politics*, *67*, 560–573.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone vs. face-to-face

    interviewing of national probability samples with long questionnaires: Comparisons of

    respondent satisficing and social desirability response bias. *Public Opinion Quarterly, 67*,

    79-125.

Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports:

    Tests using the item count technique. *Public Opinion Quarterly, 74*, 37-67.

Katosh, J. P., & Traugott, M. W. (1981). The consequences of validated and self-reported

    voting measures in studies of electoral participation. *Public Opinion Quarterly, 45*, 519-

    535.

Lutz, G. (2003). *Participation, cognitive involvement, and democracy: When do low turnout*

and low cognitive involvement make a difference, and why? Paper presented at the

European Consortium for Political Research Joint Sessions of Workshops, Edinburgh,

UK.

Lyons, W., & Scheb, J. M., II (1999). Early voting and the timing of the vote: Unanticipated

consequences of electoral reform. *State and Local Government Review,* 31, 147–152.

McDonald, M. P. (2003). On the over-report bias of the National Election Study. *Political*

Analysis, *11*, 180–186.

McDonald, M. P. (2007). The true electorate: A cross-validation of voter registration files and

election survey demographics. *Public Opinion Quarterly*, *71*, 588-602.

McDonald, M. P. (2009). United States Elections Project [2008 general election turnout rates].

Retrieved from http://elections.gmu.edu/

McDonald, M. P. (2011a). United States Elections Project [2006 general election turnout rates].

Retrieved from http://elections.gmu.edu/Turnout_2006G.html.

McDonald, M. P. (2011b). United States Elections Project [2008 general election Voter

Registration Statistics]. Retrieved from

http://elections.gmu.edu/Registration_2008G.html

New York State Board of Elections. (2009). 2008 election results. Retrieved from

http://www.elections.state.ny.us/

North Carolina State Board of Elections. (2009). 2008 general election. Retrieved from

http://results.enr.clarityelections.com/

Ohio Secretary of State. (2009). Voter turnout: November 4, 2008. Retrieved from

http://www.sos.state.oh.us/SOS/elections/electResultsMain/

Pennsylvania Department of State. (2010). Voter registration statistics. Retrieved from

http://www.dos.state.pa.us/portal/server.pt/community/voter_registration_statistics/

12725.

Presser, S., Traugott, M. W., & Traugott, S. (1990). *Vote 'over' reporting in surveys: The records or the respondents?* (ANES Technical Report Series. No. nes010157). Retrieved from http://www.umich.edu/~nes/.

Rosenstone, S. J. & Hansen, J. M. (1993). *Mobilization, participation, and democracy in America.* New York: MacMillan Publishing Company.

Sigelman, L. & Jewell, M. E. (1986). From core to periphery: A note on the imagery of concentric electorates. *The Journal of Politics*, *48*, 440-449.

Timpone, R. J. (1998). Structure, behavior, and vote turnout in the United States. *American Political Science Review*, *92*, 145-158.

Traugott, M. W. & Katosh, J. P. (1979). Response validity in surveys of voting behavior. *Public Opinion Quarterly, 43*, 359–377.

Traugott, M. W., Traugott, S., & Presser, S. (1992). *Revalidation of self-reported vote.* ANES Technical Report Series (Document nes010160).

Uhlaner, C. J. (1989). Turnout in recent American presidential elections. *Political Behavior*, *11*, 57-79.

Appendix A

State Voter Files

California state voter file. The California state government file, obtained June 19, 2009,

contains records for N=17,094,209 individuals.[30] The file contains each individual's full name,

current residential address, previous residential address, mailing address, date of birth, sex,

telephone number, language, registration date, registration status, party registration, and turnout

for the previous eight elections.

Each county maintains its own records and receives some information from the California

Secretary of State about listings that need to be updated or changed. The California Secretary of

State receives information from the California Department of Motor Vehicles (CDMV), the

California Department of Public Health (CDPH), the California Department of Corrections

(CDCR), and the United States Postal Service (USPS). Change of address information is

received from the CDMV daily, and from the USPS monthly. Death records are obtained from

the CDPH on an intermittent basis, and state felony conviction information is received from the

CDCR monthly. Any information requiring a change in an individual's registration listing is

transmitted electronically to the appropriate county.

Florida voter files. The Florida records are maintained separately for each of the 67

county, and registration records and turnout histories are maintained in separate files. The

combined government files, obtained February 3, 2009, contain records for a total of

N=12,570,868 individuals.[31] The registration records included listings for N=12,558,413

individuals, and the turnout records contained listings for an additional N=12,544 records for

---

[30] The California Secretary of State website reports 17,304,091 individuals registered to vote in California for the 2008 general election (available at http://www.sos.ca.gov/elections/sov/2008_general/2_county_reg_totals.pdf).
[31] The Florida Division of Elections website reports 11,247,634 individuals registered to vote in Florida as of October 6, 2008 (available at http://election.dos.state.fl.us/statistics.shtml).

people who voted in the 2008 general election but were no longer registered. Each county's

registration records contain each individual's full name, residential and mailing addresses, date

of birth, sex, race, date of registration, registration status, registered party affiliation, and

telephone number. Each county's turnout history records included turnout for every election

since at least 1998.[32]

Registration records are updated based on information from a variety of sources. New

listings are added as the county receives registration requests from individuals. The counties

also update information (e.g. new addresses or name changes) in existing registration listings

when that information is provided by registered individuals. Each month the state receives the

names of individuals whose eligibility to vote has changed due to death, felony convictions,

adjudication of mental capacity and relocation to another state from the Department of Health,

clerks of the circuit court, United State Attorney, Department of Law Enforcement, Board of

Executive Clemency, Department of Corrections, and Department of Highway Safety and Motor

Vehicles. The state distributes this information to the county supervisors of elections, who

update the listings in their records. County supervisors are also responsible for updating records

based on change of address information supplied by the United States Postal Service.

New York state voter file. The New York state government file, obtained June 11, 2009,

contains records for N=12,718,771 individuals.[33] The records contain each individual's full

name, current residential address, mailing address, date of birth, sex, date of registration, method

of registration, registration status, registered party affiliation, and turnout history since 2002.

New York counties maintain individual registration records, and obtain information about

---

[32] Several counties included vote histories for elections prior to 1990, and one county included vote histories for an election apparently conducted in 1900.

[33] The New York State Board of Election website reports 12,031,312 individuals registered to vote in New York as of November 1, 2008 (available at http://www.elections.state.ny.us/NYSBOE/download/AnnualReport2008.pdf).

records that need to be changed from the State Board of Elections. The State Board of Elections

receives death notifications from the New York State Department of Health and the New York

City Department of Health and Mental Health, and notifications of felony convictions and

persons determined mentally incompetent from the New York State Office of Court

Administration. The State Board of Elections forwards this information to the counties, which

have 25 days to update their registration records. The state also compares its statewide list of

registered individuals to the National Change of Address registry managed by the United States

Postal Service at least once a year and forwards information about address changes to the

counties.

North Carolina voter files. The North Carolina government records, obtained February

12, 2009, are maintained separately for each of the 100 counties in that state, and each county

also maintains registration and turnout history information in separate records. The 100

counties' records contain listings for a total of N=6,230,749 individuals.[34] The registration

records included listings for N=6,223,974 individuals, and the turnout records contained listings

for an additional N=6,776 records for people who voted in the 2008 general election but were no

longer registered. Although the counties differ in the types of information maintained, all county

records contain each individual's full name, residential and mailing addresses, current age, sex,

date of registration, registration status, registered party affiliation, and race. Each county's

records contain turnout histories for elections during the previous 14 years.

County records are updated based on information from a variety of sources. New listings

are added as the county receives registration requests from individuals. The counties also update

information (e.g. new addresses or name changes) in existing registration listings when that

---

[34] The North Carolina State Board of Elections website reports 6,277,007 individuals registered to vote in North Carolina as of November 29, 2008 (available at
http://www.app.sboe.state.nc.us/NCSBE/VR/VR%20Stats/vr_stats_results.asp?EC=11-29-2008).

information is provided by registered individuals. The state receives the names of deceased

individuals who resided in the state from the Department of Health and Human Services once a

month. The state distributes this information to the counties, which then remove the listings

from their records. Each month the state also receives the names of individuals who have

committed felonies from the United States Attorney, and distributes that information to the

counties so that registration listings may be updated.   The counties also receive information

monthly from the State Board of Elections regarding individuals who have been convicted of a

felony within the state. Change of address information is acquired from the United States Postal

Service and distributed to the counties every three months.

   Ohio state voter file. The Ohio state government file, obtained April 25, 2009, contains

records for N=8,246,881 individuals.[35] The records contain each individual's full name,

residential and mailing addresses, year of birth, registration date, turnout history for elections

during the previous 7 years, and the party's primary in which registered individuals turned out.

   Ohio government records of registration are managed almost entirely by the county

boards. Each county board obtains information monthly about deaths from the Board of Health.

Counties also receive information from the courts about residents who have been convicted of

felonies or deemed incompetent to vote. The counties update registration records and send them

to the Ohio Secretary of State, which serves predominantly as the repository of those records.

However, the state does check information about registered individuals once every two years

against the United States Postal Service's National Change of Address registry.

   Pennsylvania state voter file. The Pennsylvania state voter file, obtained June 1, 2010,

---

[35] The Ohio Secretary of State website reports 8,287,665 individuals registered to vote in Ohio for the 2008 general
election (available at
http://www.sos.state.oh.us/SOS/elections/electResultsMain/2008ElectionResults/turnout110408.aspx).

contains records for N=9,444,317 people in the state. Of these, N=9,246,579 were registered

prior to the 2008 general election.[36] Government records are maintained separately for each of

the 67 counties in that state, and each county maintains registration and turnout history

information in separate records. Each county's registration records contain each individual's full

name, residential and mailing addresses, date of birth, sex, date of registration, registration

status, registered party affiliation, and telephone number. Each county's turnout history records

include turnout and voting method for the last 40 elections in the county.

---

[36] The Pennsylvania Department of State website reports 8,755,588 individuals were registered to vote in
Pennsylvania as of November 4, 2008 (available at
(http://www.dos.state.pa.us/portal/server.pt/community/voter_registration_statistics/12725).

Table A1

Government Files Matching Variables by State

State	Last Name	First Name	Address	Age	Birth Month	Day	Year
California	X	X	X		X	X	X
Florida	X	X	X		X	X	X
New York	X	X	X		X	X	X
North Carolina	X	X	X	X			
Ohio	X	X	X				X
Pennsylvania	X	X	X		X	X	X

Appendix B

Information Regarding Deadwood Estimation

Ansolabehere and Hersh's (2010) deadwood estimates came from Catalist, LLC.[37]

Complete information about the method used by Catalist to estimate deadwood is not available.

Part of the method involves matching government registration and turnout records to records

from other sources, and the procedures they uses to match records are proprietary. However,

Catalist did disclose the following information about the method they used to generate the

deadwood estimates reported in Ansolabehere and Hersh (2010).

Catalist uses data from several sources to categorize individual registration records as

"Likely Deadwood", "Probably Deadwood", "Possibly Deadwood", or "Not Deadwood".[38]

Sources of data included registration records, turnout histories, the Social Security Death Index

file (SSDI), the National Change of Address file (NCOA), and information from another

commercial vendor that maintains a database of consumer data.[39] Data from these sources are

use to categorize records along several factors that contribute to deadwood estimation. Those

factors are:

1. *The chances that the record is for someone who is deceased.* The chances that a record is

    for someone who is deceased is based mostly on the person's age and the SSDI file. The

    additional commercial vendor also flags records as belonging to deceased individual.[40]

    Individuals listed in the SSDI file or who are flagged as deceased by the additional

---

[37] Information about Catalist, LLC is available in Appendix C.

[38] Deadwood estimates Ansolabehere and Hersh (2010) are based on records categorized as Likely or Probably Deadwood.

[39] The additional commercial vendor creates records for consumers based on data provided by businesses. Those data come from product registrations and consumer account information.

[40] The additional commercial vendor flags individuals as deceased when the companies from whom they obtain information report that correspondence mailed to individuals are returned as undeliverable.

commercial vendor, and who are more than 65 years of age, are categorized as likely

deceased. Those listed in the SSDI file or flagged as deceased, but are less than 65 years

of age, are categorized as possibly deceased.

2. *The state lists the individual in the registration record as inactive*. According to HAVA,

    registrants may be classified as inactive if they have not voted in two consecutive election

    cycles. Registration records from some states include listings for individuals identified as

    inactive (others only release information for active registrants). However, states differ in

    the strategies used to identify registrants as inactive. Some states adopt and apply very

    strict procedures, while others are more relaxed in identifying registrants as inactive.

    Records identified as inactive from states that follow strict procedures are categorized as

    Strong Inactive while those from states with more relaxed procedures are categorized as

    weak inactive.

3. *The state has dropped a registrant's record*. The commercial vendor obtains government

    records from the states on a periodic basis. A listing that appears in records from one

    year but does not appear in later records is categorized as dropped.

4. *The record is for an individual who has moved*. A registrant may be categorized as

    having moved for two different reasons. First, a record exists for a registrant in one state

    that is better, or more recent, than the registrant's record in a different state. Second, the

    NCOA indicates a registrant has moved, and the registrant has a mailing address in a

    different county than the county in which the record appears.

5. *The turnout history indicates a registrant has not voted recently.* Records indicating the

    registrant has not voted since 2000 are categorized as not having voted recently.

The five factors described above used to determine if a record is Likely, Probably, Possibly,

or Not Deadwood. The specific combinations of factors associated with the different levels of

deadwood are:

1. *Likely Deadwood*. A record is categorized as Likely Deadwood if: (1) the registrant is

    possibly or likely deceased, and has been dropped or classified as inactive by the state; or

    (2) the registrant has moved or has a better record in a different state, and has been

    dropped by a state or been categorized as Strong Inactive.

2. *Probably Deadwood*. A record is categorized as Probably Deadwood if: (1) the registrant

    has moved or has a better record in a different state, and has been categorized as Weak

    Inactive; (2) the registrant has not voted recently, and has moved or has a better record in

    a different state; or (3) the registrant has not voted recently, and has been dropped or

    classified as inactive by the state.

3. *Possibly Deadwood*. A record is categorized as Possibly Deadwood if: (1) the registrant

    is likely deceased; (2) the registrant has moved or has a better record in a different state;

    or (3) the registrant has not voted recently.

4. *Not Deadwood*. All records that have not been categorized as Likely, Probably, or

    Possibly Deadwood are categorized as Not Deadwood.

Appendix C

Information About Commercial Vendors and Government Records

Table C1

Information About Selected Commercial Vendors of Government Records

Vendor[b]	Costs[a]		Records coverage	Comments
	Outsourcing matching	Purchasing records		
AmeriCan GOTV Voter Lists	$9,000 to $12,000	NA	50 states	• 2008 voter data would not be updated until 2010. • Cost is determined at a per match rate.
Aristotle International	$4,000	$250,000 to $500,000	49 states	• Records are obtained directly from government sources for 41 states (IL, IN, MN, VA, WV, and WY are obtained through an agent; and ND does make a voter file available. • Records from AZ and MA are preprocessed via proxies. • Documented assurance that the records would be current for November 2008. • Could not provide assurance that the public records are accurate. • Records are matched based on first and last names, up to 8 address characters, and zip code. • Non-partisan
Catalist, LLC	$1500	$50,000 to $90,000	50 states	• All records are acquired from governmental sources. • An online tool is available that allows

- organizations to match records themselves.
- Services are made available only to progressive organizations

ElectionMall.com			48 states	
E-merges	NA	$12,000	50 states	
First Tuesday in November	NR	NR	1.5 million records	
Labels and Lists	$9,350	$13,000	11 states	• Records are obtained directly from government sources.
Perry Political Software	$20,000+	NR	Limited	• Vote histories are not in the database. • Vote histories would need to be purchased from the states at the customer's expense.
Political Media Inc.				
Political Data Inc.	NR	NR	California only	• Records are available for California only.
SMA Communications	$500	NR	50 states	• The vendor refused to provide any information concerning the sources of their records, the accuracy of their records as of November 2008, or the partisanship of their organization.
Voter Contact Services	NR	NR	50 states	• Records from several states had not been updated to include 2008 general election results
YouGov/Polimetrix	$15,000	NA	50 states	• Records are obtained directly from government sources. • Matching uses a proprietary algorithm that the vendor will not reveal.

[a] Inquiries were made with each vendor regarding two options. The first option was outsourcing the matching process, which would have involved supplying vendors with personally identifying information of the respondent's and having them locate registration and turnout records. The second option was purchasing records from vendors and conducting the matching process in-house. Thus, working with a commercial vendor would have incurred either the cost of outsourcing or the cost of purchasing the records. For the outsourcing option, "NA" indicates the vendor does not match records to those from external sources. For the purchasing records option, "NA" indicates the vendor does not sell its records. For both options, "NR" indicates a cost was not requested for reasons

listed under the "comments" column.

[b]Additional commercial vendors who failed to respond to inquiries about their services with respect to this project included: American Information Marketing, Best ROI Lists, Bose Public Affairs Group, Cornerstone Solutions, Campaign Products of the Rockies, Fieldworks, The Monaco Group, S D Associates, and Target Marketing USA Inc.

Table C2

Information about Government Records by State

State	Cost	Purchase restrictions	Permissible uses
Alabama	$31,000	None	Political purposes
Alaska	$178	None	All
Arizona	$149,373	Political parties Political campaigns	Political purposes
Arkansas	$5	None	All
California	$30	None	Political purposes Voter surveys
Colorado	$750	None	All
Connecticut	$300	None	All
Delaware	$10	None	All
District of Columbia	$10	None	All
Florida	$10	None	All
Georgia	$500	None	Non-commercial
Hawaii	$450	Candidates Political parties Political committees Governmental agencies	Political purposes

State	Fee	Restricted to	Permitted use
Idaho	$20	None	Non-commercial
Illinois	$500	State registered political committees State political office holders State political candidates	Political purposes
Indiana	$5,000	None	Non-commercial
Iowa	$1,060	None	Political purposes Non-commercial
Kansas	$200	None	Non-commercial
Kentucky	$450	None	Non-commercial
Louisiana	$5,000	None	All
Maine	$2,000	None	Non-commercial
Maryland	$125	State registered voters	Non-commercial
Massachusetts	$0	State party committees Statewide candidate committees State ballot question committees State jury commissioner	Political purposes
Michigan	$22.08	None	All
Minnesota	$51	State residents	Political purposes Law enforcement
Mississippi	$2,100	None	All
Missouri	$127	None	Non-commercial
Montana	$1,000	None	Non-commercial

Nebraska	$500	None		Non-commercial
Nevada	$0	None		All
New Hampshire	$600	None		All
New Jersey	$53	None		Non-commercial
New Mexico	$4,000	None		Non-commercial
New York	$0	None		Election related
North Carolina	$0	None		All
North Dakota	$3,850	None		All
Ohio	$0	None		All
Oklahoma	$150	None		All
Oregon	$500	None		Non-commercial
Pennsylvania	$20	None		Election related Political purposes Law enforcement
Rhode Island	$25	None		Political purposes
South Carolina	$160	State voter		Non-commercial
South Dakota	$2,500	None		Non-commercial
Tennessee	$2,500	None		Political purposes

Texas	$5,953	None	Non-commercial
Utah	$1,050	None	All
Vermont	$0	None	Non-commercial
Virginia	$4,791	State courts State political candidates State political parties State political action committees State political office holders Non-profit organizations promoting voter participation	Jury selection Political purposes Promoting voter participation
Washington	$30	None	Non-commercial
West Virginia	$6,000	None	Non-commercial Non-charitable solicitation
Wisconsin	$12,500	None	All
Wyoming	$20	State political candidates State political office holders State political action committees Organizations promoting voter participation	Non-commercial

Appendix D

Explanatory Variables for Registration and Turnout Regressions

The following list of explanatory variables follows from Rosenstone and Hansen (1993). The ANES 2008 Panel Study variable names that provide measures of these variables are in parentheses.

Income (der06): coded 0 if $\leq$ "$24,999", .25 if $\geq$ "$25,000" and $\leq$ "$39,999", .5 if $\geq$ "$40,000" and $\leq$ "$84,999", .75 if $\geq$ "$85,000" and $\leq$ "$174,999", and 1 if $\geq$ "$175,000". Missing values (n=5) coded to .5 (modal response).

Education (w9zw5 or cpq15): coded 0 if $\leq$ "7th or 8th grade", .25 if $\geq$ "9th grade" and $\leq$ "12th grade no diploma", .5 if = "high school graduate high school diploma or the equivalent (GED)", .75 if = "some college no degree", and 1 if $\geq$ "Associates degree". Missing value (n=1) coded to 1 (modal response).

Unemployed (cpq17 or w11zg1): coded 0 if = "Working-as a paid employee", "Working-self-employed", "Not working-on temporary layoff from a job", "Not working-retired", "Not working-disabled", or "Not working-other" and 1 if = "Not working-looking for work". Missing values (n=5) coded to 0 (modal response).

Age (der02): coded on 0 to 1 scale in which 0=minimum observed age (18 years) and 1=maximum observed age (89 years). No missing values.

Age2 (der02): coded on 0 to 1 scale in which 0=minimum observed squared age (324 squared years) and 1=maximum observed squared age (7921 squared years). No missing values.

External political efficacy (w10j1 and w10j2): the average of the recoded responses. Both coded 0 if = "Not at all", .25 if = "A little", .5 if = "A moderate amount", .75 if = "A lot", and

1 if = "A great deal". Missing values (n=39) coded to .5 (scale midpoint).

Party identification strength (der08w10 or der08w9 or der08w1): coded 0 if = "Independent Independent", .33 if = "Independent Democrat" or "Independent Republican", .5 if = "Missing", .67 if = "Weak Democrat" or "Weak Republican", and 1 if = "Strong Democrat" or "Strong Republican". No missing values.

Preference for a party (w10e2 and w10e3 and w10e4 and w10e5 and w10e6 and w10e7, or w9e2 and w9e3 and w9e4 and w9e5 and w9e6 and w9e7, or w6e2 and w6e3 and w6e4 and w6e5 and w6e6 and w6e7, or w2e2 and w2e3 and w2e4 and w2e5 and w2e6 and w2e7, or w1e2 and w1e3 and w1e4 and w1e5 and w1e6 and w1e7): coded as the absolute value of the difference between "like" for the two major parties.[41] "Like" for a party coded as 0 if = "dislike a great deal", .17 if = "dislike a moderate amount", .33 if = "dislike a little", .5 if = "neither like nor dislike", .67 if = "like a little", .83 if = "like a moderate amount", and 1 if = like a great deal". No missing values.

Care who wins the election (w10v1 or w9u1): coded 0 if = "Not at all", .25 if = "A little", .5 if = "A moderate amount", .75 if = "A lot", and 1 if = "A great deal".[42] Missing values (n=9) coded to .5 (scale midpoint).

Preference for a candidate (w10e14 and w10e15 and w10e16 and w10e38 and w10e39 and w10e40): coded as the absolute value of the difference between "like" for the two major party candidates.[43] "Like" for a candidate coded as 0 if = "dislike a great deal", .17 if = "dislike

---

[41] Rosenstone and Hansen (1993) computed Affect for a party as the absolute value of the difference between the number of Democrat Party "likes" plus Republican Party "Dislikes" and the number of Democrat Party "dislikes" plus the number of Republican Party "likes", recoded to a 0 to 1 scale.

[42] Rosenstone and Hansen (1993) computed a similar estimate that was based on responses to the question, "Generally speaking, would you say that you personally care a good deal which party wins the presidential; election this fall, or don't you care very much which party wins?" Caring which party wins was not asked during the Panel Study, so Caring which candidate wins was used in its place.

[43] Rosenstone and Hansen (1993) computed Affect for a candidate using the same procedure as they did for Affect for a party.

a moderate amount", .33 if = "dislike a little", .5 if = "neither like nor dislike", .67 if = "like a

little", .83 if = "like a moderate amount", and 1 if = like a great deal". Missing values (n=32)

coded to .5 (scale midpoint).

Church attendance (w1j1a_1 and w1j1a_2 and w1j1a_3 and w1j1c_1 and w1j1c_2 and

w1j1c_3, or w9zg1a_1 and w9zg1a_2 and w9zg1a_3 and w9zg1c_1 and w9zg1c_2 and

w9zg1c_3): coded 0 if no religious service attendance during the past 12 months, .25 if

attendance at 1 to 10 religious services during the past 12 months, .5 if attendance at 11 to 24

religious services during the past 12 months, .75 if attendance at 25 to 51 religious services

during the past 12 months, and 1 if attendance at 52 or more religious services during the past 12

months.[44] Missing values (n=38) coded to .5 (scale midpoint).

Home owner (w10zw4 or w9zw4 or w2w4 or w1y4 or cpq2 or w11zf1): coded as 0 if

="R rents home" or "Other arrangement" and 1 if ="R owns home". Missing value (n=1) coded

to 1 (modal response).

Working (cpq17 or w11zg1): coded 0 if ="Not working-on temporary layoff from a

job", "Not working-looking for work", "Not working-retired", "Not working-disabled", or "Not

working-other" and 1 if ="Working-as a paid employee" or "Working-self-employed". Missing

values (n=5) coded to 1 (modal response).

Perceived closeness of the election (w10v3 or w9u3): coded 0 if ="Candidate will win

by quite a bit" and 1 if ="Will be close". Missing values (n=10) coded to 1 (modal value).

Southern (w10xhome or w9xhomes or w6xhomes or w1state): coded 1 if = "FL" or

"NC" and 0 otherwise. No missing values

---

[44] Rosenstone and Hansen (1993) computed Church attendance from responses to two different questions (asked in different years). The question asked from 1952 through 1968 was worded, "Would you say that you go to church regularly, often, seldom, or never?" The question asked from 1970 through 1988 was worded, "Would you say that you go to (church/synagogue) every week, almost every week, once or twice a month, a few times a year, or never?"

Black (der04): coded as 0 if ="White, non-Hispanic", "Hispanic", or "Other, non-Hispanic" and 1 if ="Black, non-Hispanic". No missing values.

Hispanic (der04): coded as 0 if ="White, non-Hispanic", "Black, non-Hispanic", or "Other, non-Hispanic", and 1 if ="Hispanic".   No missing values

Table 1

Possible Outcomes of Matching Study Respondents to Government Records of Registration

Matching outcome	True state of affairs	
	Respondent was registered	Respondent was not registered
Matched	accuracy uncertain – true positive or false negative	inaccurate – false positive
Not matched	inaccurate – false negative	accurate – true negative

Table 2

Percents of ANES Respondents Who Said They were Registered and Not Registered by State

| Sample | State | Self-reported registration | | | | | | Study respondents whose Knowledge Networks address was in the state |
		Registered in the state where their Knowledge Networks address was located	Registered in another state	Not registered	Don't know	Not answered	Total	
Total sample	All states	82.52%	2.79%	6.95%	1.12%	6.63%	100%	3049
Target states	All target states	81.89%	3.12%	7.44%	.80%	6.74%	100%	994
	California	80.65%	3.23%	7.33%	1.47%	7.33%	100%	341
	Florida	80.43%	1.45%	10.14%	.00%	7.97%	100%	138
	New York	78.43%	4.58%	9.15%	.65%	7.19%	100%	153
	North Carolina	79.78%	4.49%	5.62%	1.12%	8.99%	100%	89
	Ohio	87.41%	2.96%	4.44%	.00%	5.19%	100%	135
	Pennsylvania	86.23%	2.17%	7.25%	.72%	3.62%	100%	138

Table 3

Percents of ANES Respondents Who Said They Turned Out and Did Not Turn Out by State

Sample	State	Self-reported turnout		Not answered	Total	Study respondents Whose Knowledge Networks address was in the state
		Voted	Did not vote			
Total sample	All states	80.06%	9.25%	10.69%	100%	3049
Target states	All target states	80.58%	8.55%	10.87%	100%	994
	California	82.99%	7.33%	9.68%	100%	341
	Florida	81.88%	7.25%	10.87%	100%	138
	New York	73.86%	11.76%	14.38%	100%	153
	North Carolina	77.53%	11.24%	11.24%	100%	89
	Ohio	82.96%	7.41%	9.63%	100%	135
	Pennsylvania	80.43%	8.70%	10.87%	100%	138

Table 4

Registration According to the Government Files by State

| | Registration status | | | Number of |
	Registered[46]	Not registered	Total	listings[45]
All target states	98.26%	1.74%	100%	65,305,796
California	100.00%	.00%	100%	17,094,209
Florida	100.00%	.00%	100%	12,570,869
New York	91.68%	8.32%	100%	12,718,771
North Carolina	98.78%	1.22%	100%	6,230,749
Ohio	100.00%	.00%	100%	8,246,881
Pennsylvania	100.00%	.00%	100%	8,444,317

---

[45] The number of listings only includes records in the registration file. Florida and North Carolina maintain separate registration and turnout records, and the number of listings in the turnout records is different than the number of listings in the registration records.

[46] The Florida files included n=12,544 records for people who turned out for the 2008 general election, but who were no longer registered when we obtained the files in February of 2009. The North Carolina files included n=6,776 such records. These people are counted among those registered in Florida and North Carolina.

Table 5

Comparison of Registration Statistics Published by States to Registration Information in

the Government Records

State	Number of individuals registered in the government records we obtained[47]	Number of individuals registered according to statistics published by states	Percent differences
All target states	64,171,163	63,852,494	.50%
	[61,605,200]	[61,347,173]	.42%
California[a]	17,094,209	17,304,091	-1.23%
	[17,038,415]	[17,247,612]	-1.23%
Florida	12,570,869	11,247,634	10.53%
	[11,579,406]	[10,360,534]	10.53%
New York	11,660,114	12,031,312	-3.18%
	[11,057,847]	[11,409,872]	-3.18%
North Carolina	6,154,773	6,226,204	-1.16%
	[5,850,547]	[5,918,447]	-1.16%
Ohio[a]	8,246,881	8,287,665	-.49%
	[8,176,892]	[8,217,330]	-.49%
Pennsylvania	8,444,317	8,755,588	-3.69%
	[7,902,094]	[8,193,377]	-3.69%

Notes: Percent difference is computed as 1-(# of registered individuals according to

published statistics/# of registered individuals in the government files we obtained). The

numbers after eliminating estimated deadwood are in brackets.

[a]The government records we obtained only included "active" registrants.

---

[47] The numbers include the n=12,544 records in Florida and n=6,776 records with turnout histories for the 2008 general election who were no longer registered in February of 2009.

Table 6

Comparison of Turnout Statistics Published by the States to Turnout Statistics from the Government Records

State	Statistics from the government records we obtained				Statistics published by the states			
	Total number of individuals registered in the government file	Total number of individuals who voted in the 2008 general election	Percent of registered individuals who voted	Percent of the residents over 18 years of age who voted[a]	Total number of individuals registered for the 2008 general election	Total number of individuals who voted in the 2008 general election	Percent of registered individuals who voted	Percent of the residents over 18 years of age who voted[a]
All target states	64,171,163 [61,605,200]	43,831,383	68.30% [71.15%]	53.23%	63,852,494 [61,347,173]	46,120,934	72.23% [75.18%]	56.01%
California	17,094,209 [17,038,415]	13,216,967	77.32% [77.57%]	48.45%	17,304,091 [17,247,612]	13,743,177	79.42% [79.68%]	50.38%
Florida	12,570,869 [11,579,406]	8,442,112	67.16% [72.91%]	58.64%	11,247,634 [10,360,534]	8,456,329	75.18% [81.62%]	58.74%
New York	11,660,114 [11,057,847]	7,549,393	64.75% [68.27%]	50.17%	12,031,312 [11,409,872]	7,722,019	64.18% [67.68%]	51.31%
North Carolina	6,154,773 [5,850,547]	3,414,253	55.47% [58.36%]	48.57%	6,226,204 [5,918,447]	4,353,739	69.93% [73.56%]	61.93%
Ohio	8,246,881 [8,176,892]	5,319,382	64.50% [65.05%]	60.43%	8,287,665 [8,217,330]	5,773,777	69.67% [70.26%]	65.59%
Pennsylvania	8,444,317 [7,902,094]	5,889,276	69.74% [74.53%]	60.15%	8,755,588 [8,193,377]	6,071,893	69.35% [74.11%]	62.02%

[a]The numbers of residents over 18 years of age in the different states is from McDonald (2009).

Note: The numbers after eliminating estimated deadwood are in brackets.

Table 7

Frequencies of Missing or Implausible Information in the Government Records

| State | Percent missing, implausible, or suppressed | | | | Total number of registration records |
	Last name	First Name	Address	Birth	
Total	.02%	.02%	.14%	.46%	65,305,796
California	.0005%	.006%	.07%	1.08%	17,094,209
Florida	.10%	.10%	.39%	.10%	12,570,869
New York	.00006%	.0005%	.24%	.13%	12,718,771
North Carolina[48]	.0002%	.0008%	.005%	.0004%	6,161,400
Ohio	.00006%	.0006%	.00%	1.00%	8,246,881
Pennsylvania	.00005%	.001%	.02%	.06%	8,444,317

---

[48] The North Carolina records included turnout histories for n=6776 people who turned out in the 2008 general election, but whose registration records had been purged when we obtained the registration data in February of 2009. Because records for those people were not included in the registration data, they are not included in the total number of registration records here.

Table 8

Implausible, Plausible, and Missing Birth Information in Government Records for Registrations Added or Updated Since the Help America Vote Act of 2002 (HAVA), the National Voter Registration Act of 1993 (NVRA), and Total Registrations

| State | Act | Birth information | | | | Registrations added or updated since the Act |
		Birth year in or before 1888	Birth year after 1888	Missing	Total	
All target states	HAVA	.007%	99.99%	.002%	100%	27,407,947
	NVRA	.01%	99.99%	.002%	100%	45,244,522
	Total	.46%	99.53%	.008%	100%	65,236,438
California	HAVA	.0009%	99.99%	.006%	100%	9,564,736
	NVRA	.001%	99.99%	.007%	100%	13,630,811
	Total	1.05%	98.92%	.03%	100%	17,094,209
Florida	HAVA	.0004%	100.00%	.0000%	100%	5,083,475
	NVRA	.0002%	100.00%	.0000%	100%	9,118,993
	Total	.10%	99.90%	.0004%	100%	12,570,868
New York	HAVA	.006%	99.99%	.00003%	100%	3,771,696
	NVRA	.007%	99.99%	.00001%	100%	7,511,476
	Total	.13%	99.87%	.000008%	100%	12,718,771
North Carolina	HAVA	.0008%	100.00%	.00%	100%	2,589,419
	NVRA	.0005%	100.00%	.00%	100%	4,206,711
	Total	.0004%	100.00%	.00%	100%	6,161,400
Ohio	HAVA	.02%	99.98%	.00%	100%	3,512,211
	NVRA	.03%	99.97%	.00%	100%	5,157,117
	Total	1.00%	99.00%	.00%	100%	8,246,881
Pennsylvania	HAVA	.03%	99.97%	.002%	100%	2,886,410
	NVRA	.04%	99.96%	.001%	100%	5,619,414
	Total	.10%	99.90%	.0008%	100%	8,444,309

Table 9

Match rates of different criteria and previous vote validation studies

| Sample group | 2008 ANES Panel Study | | | ANES Time Series | | | | | | | 2006 CCES study |
	STRICT	MOD	LEAST	1976[a]	1978[b]	1980[a]	1984[b]	1986[b]	1988[b]	1990[b]	
Respondents living in a location for which government records were available	45.57%	65.39%	77.46%	65.17%	63.59%	64.32%	72.71%	63.36%	68.53%	60.81%	62.50%
N	*994*	*994*	*994*	*2,908*	*2,304*	*1,522*	*2,257*	*2,140*	*2,005*	*1,980*	*22,930*
Respondents who said (or implied) they were registered in their state of residence or an area for which government records were available	49.88%	71.62%	84.03%	77.57%	83.64%	81.00%	90.81%	86.26%	88.50%	87.06%	64.90%
N	*814*	*814*	*814*	*2,256*	*1,650*	*1,095*	*1,622*	*1,572*	*1,383*	*1,383*	*21,806*

[a]Self-reported registration is based on responses to a single registration status question asked during pre-election interviews (V763032 in 1976, and V800134 in 1980).

[b]Self-reported registration is based on responses to turnout and registration questions (the latter of which was only asked to respondents who reported not turning out in a general election) asked during post-election interviews. Respondents were coded as reporting registration if they reported turning out in the general election or being registered (V780470 and V780484 in 1978, V840783 and V840802 in 1984, V860261 and V860276 in 1986, V880756 and V880757 in 1988, and V900279 and V900280 in 1990).

Table 10

Comparison of the Percent of People Who were Registered to the Percent of Study Respondents Who Said They Were Registered and

What the Government Records Suggest about the Percent of Study Respondents Who Were Registered

State	Percent of citizens over 18 years of age who lived in households served by a landline telephone as of November, 2008 who were registered	Percent of survey respondents whose Knowledge Networks address was in the state who said they were registered in that state	Percent of survey respondents whose Knowledge Networks address was in the state who said they were registered in any state	What government records suggest about the percent of survey respondents whose Knowledge Networks address was in the state who were registered in that state using different matching criteria			Study respondents whose Knowledge Networks address was in the state
				STRICT	MOD	LEAST	
All target states	87.82% [84.31%]	81.89%	85.01%	45.57%	65.39%	77.46%	994
California	76.33% [76.08%]	80.65%	83.87%	43.11%	65.98%	77.13%	341
Florida	98.34% [90.58%]	80.43%	81.88%	48.55%	68.12%	84.06%	138
New York	88.45% [83.88%]	78.43%	83.01%	35.29%	59.48%	77.12%	153
North Carolina	93.05% [88.45%]	79.78%	84.27%	38.20%	52.81%	58.43%	89
Ohio	95.80% [94.98%]	87.41%	90.37%	59.26%	72.59%	85.19%	135
Pennsylvania	89.01% [87.82%]	86.23%	88.41%	51.45%	68.84%	76.81%	138

Note: Estimated percentages after eliminating estimated deadwood are in brackets.

Table 11

Matching Criteria Variable Characteristics among the 2008 Panel Study Respondents

	Characteristics of Panel Study respondents with a residential address in one of the target states				
	Missing date of birth	Age 55 years or more[a]	Non-homeowners	Gave P.O. box for residential address	N
All target states	7.95%	38.66%	23.87%	3.62%	994
California	9.38%	37.09%	31.47%	3.52%	341
Florida	7.25%	47.83%	9.42%	2.90%	138
New York	6.54%	37.75%	32.03%	3.27%	153
North Carolina	12.36%	36.36%	14.61%	7.87%	89
Ohio	8.89%	39.10%	22.22%	2.22%	135
Pennsylvania	2.90%	35.29%	18.12%	3.62%	138

[a]Excludes respondents with a missing date of birth.

Table 12

Consistencies and inconsistencies between self-reports and government records of registration

| State | Criteria | Consistencies between self-reports and government records | | | Inconsistencies between self-reports and government records | | | | |
		Self-reports and government records agree that the respondent is registered	Self-reports and government records agree that the respondent is not registered	Total	Respondent reports registered but government records suggest not registered	Respondent reports not registered but government records suggest registered	Total	Total	N
All target states	STRICT	45.38%	7.51%	52.88%	46.57%	.54%	47.12%	100%	919
	MOD	65.40%	7.29%	72.69%	26.55%	.76%	27.31%	100%	919
	LEAST	76.82%	6.53%	83.35%	15.13%	1.52%	16.65%	100%	919
California	STRICT	43.09%	7.72%	50.80%	48.87%	.32%	49.20%	100%	311
	MOD	65.59%	7.40%	72.99%	26.37%	.64%	27.01%	100%	311
	LEAST	76.85%	7.07%	83.92%	15.11%	.96%	16.08%	100%	311
Florida	STRICT	48.03%	8.66%	56.69%	40.94%	2.36%	43.31%	100%	127
	MOD	67.72%	8.66%	76.38%	21.26%	2.36%	23.62%	100%	127
	LEAST	81.89%	7.87%	89.76%	7.09%	3.15%	10.24%	100%	127
New York	STRICT	35.46%	9.22%	44.68%	54.61%	.71%	55.32%	100%	141
	MOD	60.28%	9.22%	69.50%	29.79%	.71%	30.50%	100%	141
	LEAST	75.89%	7.09%	82.98%	14.18%	2.84%	17.02%	100%	141
North Carolina	STRICT	38.75%	6.25%	45.00%	55.00%	.00%	55.00%	100%	80
	MOD	55.00%	6.25%	61.25%	38.75%	.00%	38.75%	100%	80
	LEAST	60.00%	6.25%	66.25%	33.75%	.00%	33.75%	100%	80
Ohio	STRICT	57.81%	4.69%	62.50%	37.50%	.00%	37.50%	100%	128
	MOD	71.88%	4.69%	76.56%	23.44%	.00%	23.44%	100%	128
	LEAST	83.59%	3.13%	86.72%	11.72%	1.56%	13.28%	100%	128
Pennsylvania	STRICT	50.76%	7.58%	58.33%	41.67%	.00%	41.67%	100%	132
	MOD	68.18%	6.82%	75.00%	24.24%	.76%	25.00%	100%	132
	LEAST	76.52%	6.82%	83.33%	15.91%	.76%	16.67%	100%	132

Table 13

Consistencies and inconsistencies between self-reports and government of records of registration when adding additional phone number matches to original criteria

State	Criteria plus telephone	Consistencies between self-reports and government records			Inconsistencies between self-reports and government records				
		Self-reports and government records agree the respondent is registered	Self-reports and government records agree that the respondent is not registered	Total	Respondent reports registered, government records suggest not registered	Respondent reports not registered, government records suggest registered	Total	Total	N
All target states with telephone numbers in government records	STRICT	14.31%	8.15%	22.46%	77.38%	.15%	77.54%	100%	650
	MOD	21.69%	7.85%	29.54%	70.00%	.46%	70.46%	100%	650
	LEAST	25.85%	7.85%	33.69%	65.85%	.46%	66.31%	100%	650
California	STRICT	18.01%	8.04%	26.05%	73.95%	.00%	73.95%	100%	311
	MOD	28.94%	7.72%	36.66%	63.02%	.32%	63.34%	100%	311
	LEAST	34.08%	7.72%	41.80%	57.88%	.32%	58.20%	100%	311
Florida	STRICT	12.60%	10.24%	22.83%	76.38%	.79%	77.17%	100%	127
	MOD	18.11%	10.24%	28.35%	70.87%	.79%	71.65%	100%	127
	LEAST	23.62%	10.24%	33.86%	65.35%	.79%	66.14%	100%	127
North Carolina	STRICT	11.25%	6.25%	17.50%	82.50%	.00%	82.50%	100%	80
	MOD	16.25%	6.25%	22.50%	77.50%	.00%	77.50%	100%	80
	LEAST	20.00%	6.25%	26.25%	73.75%	.00%	73.75%	100%	80
Pennsylvania	STRICT	9.09%	7.58%	16.67%	83.33%	.00%	83.33%	100%	132
	MOD	11.36%	6.82%	18.18%	81.06%	.76%	81.82%	100%	132
	LEAST	12.12%	6.82%	18.94%	80.30%	.76%	81.06%	100%	132

Table 14

Consistencies and inconsistencies between self-reports and government of records of registration when adding matches based on identical phone numbers to the original criteria

State	Criteria or telephone	Consistencies between self-reports and government records			Inconsistencies between self-reports and government records				
		Self-reports and government records agree the respondent is registered	Self-reports and government records agree that the respondent is not registered	Total	Respondent reports registered, government records suggest not registered	Respondent reports not registered, government records suggest registered	Total	Total	N
All target states with telephone numbers in government records	STRICT	46.31%	7.69%	54.00%	45.38%	.62%	46.00%	100%	650
	MOD	66.92%	7.23%	74.15%	24.77%	1.08%	25.85%	100%	650
	LEAST	77.23%	6.92%	84.15%	14.46%	1.38%	15.85%	100%	650
California	STRICT	45.02%	7.72%	52.73%	46.95%	.32%	47.27%	100%	311
	MOD	67.52%	7.40%	74.92%	24.44%	.64%	25.08%	100%	311
	LEAST	78.46%	7.07%	85.53%	13.50%	.96%	14.47%	100%	311
Florida	STRICT	48.82%	8.66%	57.48%	40.16%	2.36%	42.52%	100%	127
	MOD	68.50%	8.66%	77.17%	20.47%	2.36%	22.83%	100%	127
	LEAST	82.68%	7.87%	90.55%	6.30%	3.15%	9.45%	100%	127
North Carolina	STRICT	40.00%	6.25%	46.25%	53.75%	.00%	53.75%	100%	80
	MOD	58.75%	5.00%	63.75%	35.00%	1.25%	36.25%	100%	80
	LEAST	63.75%	5.00%	68.75%	30.00%	1.25%	31.25%	100%	80
Pennsylvania	STRICT	50.76%	7.58%	58.33%	41.67%	.00%	41.67%	100%	132
	MOD	68.94%	6.82%	75.76%	23.48%	.76%	24.24%	100%	132
	LEAST	77.27%	6.82%	84.09%	15.15%	.76%	15.91%	100%	132

Table 15

Probit Regression Estimates Using Different Measures of Registration

Explanatory Variable	Self-reported registration (a)	Registration based on government records using different matching criteria		
		STRICT (b)	MOD (c)	LEAST (d)
Income	.01	.20	.26	.03
	(.33)	(.19)	(.20)	(.22)
Education	1.31***$_{bc}$	-.19$_{acd}$	.42^+_b	.63*$_b$
	(.36)	(.24)	(.25)	(.26)
Unemployed	-.03	-.44$_c$	.14$_b$	-.08
	(.42)	(.29)	(.30)	(.31)
Age	.95	-1.23	.97	1.01
	(2.28)	(1.35)	(1.40)	(1.51)
Age squared	.36	1.24	-.74	-.68
	(2.54)	(1.37)	(1.43)	(1.55)
External Efficacy[49]	.25	.07	.03	-.10
	(.36)	(.20)	(.21)	(.24)
Strength of party identification	.99***$_{bcd}$	.10$_a$	.09	.32
	(.28)	(.17)	(.18)	(.19)
Preference for a party	-.28	-.12	-.26	-.19
	(.30)	(.17)	(.18)	(.20)
Care who wins the presidential election	1.76***$_b$	.67**$_{ac}$	1.11***$_b$	1.08***
	(.32)	(.22)	(.22)	(.24)
Preference for a presidential candidate	-.70*	-.15	-.26	-.25
	(.32)	(.18)	(.19)	(.21)
Church attendance	-.30	-.12	-.12	-.17
	(.20)	(.11)	(.12)	(.13)
Homeowners	.38$^+$	.34**	.24$^+$	.21
	(.19)	(.12)	(.13)	(.14)
Currently working	.04	-.14	-.16	-.09
	(.21)	(.12)	(.13)	(.14)
Perceived closeness of the election	.18	.23*	.31**	.22$^+$
	(.20)	(.12)	(.12)	(.13)
Live in a Southern state	-.39*	-.09	-.17	-.18
	(.19)	(.11)	(.12)	(.13)
Blacks	-.23	.09	-.04	-.19
	(.27)	(.17)	(.18)	(.19)
Hispanics	-.08	.07	.08	-.02
	(.29)	(.19)	(.20)	(.21)
Constant	-1.85***	-.68*	-1.21***	-.92*
	(.52)	(.34)	(.35)	(.37)
Model χ^2	124.66	32.79	61.83	57.39
N	818	818	818	818

Note: Cell entries are probit coefficients with standard errors in parentheses underneath. Subscripts on a coefficient indicate which other column's coefficients in that row are significantly different from it (p<.05). $^+$ p<.10, * p<.05, ** p<.01, *** p<.001

[49] A comparison of coefficients for the MOD and LEAST was unavailable due to a convergence failure of the constrained model

Table 16

Turnout Rates Estimated by Various Different Methods

State	Government Statistics: Percent of citizens over 18 years of age who lived in households served by a landline telephone as of November, 2008 who voted	Percent of survey respondents whose Knowledge Networks address was in the state who said they voted in that state	Percent of survey respondents whose Knowledge Networks address was in the state who said they voted in any state	Percent of survey respondents whose Knowledge Networks address was in the state who voted in that state using different matching criteria			Study respondents whose Knowledge Networks address was in the state
				STRICT	MOD	LEAST	
All target states	63.01%	78.47%	80.58%	39.64%	56.34%	65.90%	994
California	61.36%	82.11%	82.99%	39.88%	59.82%	68.62%	341
Florida	66.15%	78.99%	81.88%	44.93%	62.32%	76.09%	138
New York	58.58%	70.59%	73.86%	26.80%	46.41%	60.13%	153
North Carolina	65.82%	74.16%	77.53%	23.60%	32.58%	37.08%	89
Ohio	67.07%	80.74%	82.96%	50.37%	62.96%	71.85%	135
Pennsylvania	63.19%	78.26%	80.43%	47.83%	61.59%	68.12%	138

Table 17

Comparison of the Percent of Registered People Who Turned out to the Percent of Study Respondents Who Said They Were Registered and Who Said They Turned out and What the Government Records Suggest about the Percent of Registered Study Respondents Who Turned Out

State	Percent of registered citizens over 18 years of age who lived in households served by a landline telephone as of November, 2008 who voted	Percent of survey respondents whose Knowledge Networks address was in the state who said they voted in that state	Percent of survey respondents whose Knowledge Networks address was in the state who said they voted in any state	What government records suggest about the percent of survey respondents whose Knowledge Networks address was in the state who voted in that state using different matching criteria		
				STRICT	MOD	LEAST
All target states	71.75%	94.71%	95.63%	89.85%	89.23%	88.44%
	[74.74%]	*756*	*756*	*453*	*650*	*770*
California	80.40%	96.54%	97.31%	94.56%	93.33%	91.25%
	[80.66%]	*260*	*260*	*147*	*225*	*263*
Florida	67.27%	97.06%	98.04%	94.03%	95.74%	94.83%
	[73.03%]	*102*	*102*	*67*	*94*	*116*
New York	66.23%	89.91%	90.83%	83.33%	83.52%	84.75%
	[69.83%]	*109*	*109*	*54*	*91*	*118*
North Carolina	70.74%	92.42%	92.42%	64.71%	63.83%	67.31%
	[74.42%]	*66*	*66*	*34*	*47*	*52*
Ohio	70.01%	93.58%	95.41%	88.75%	89.80%	86.96%
	[70.61%]	*109*	*109*	*80*	*98*	*115*
Pennsylvania	71.00%	95.45%	96.36%	94.37%	90.53%	90.57%
	[75.87%]	*110*	*110*	*71*	*95*	*106*

Notes: Sample sizes are italicized. Turnout rates after eliminating estimated deadwood are in brackets.

Table 18

Consistencies and Inconsistencies Between Self-Reports and Government Records of Turnout For All Respondents

State	Criteria	Consistencies			Inconsistencies				
		Self-reports and government records agree: turned out	Self-reports and government records agree: no turn out	Total	Respondent: turned out Govt Records: did not	Respondent: did not Govt Records: turned out	Total	Total	N
All target states	STRICT	41.65%	9.59%	51.24%	48.76%	.00%	48.76%	100%	886
	MOD	59.59%	9.59%	69.19%	30.81%	.00%	30.81%	100%	886
	LEAST	69.86%	9.59%	79.46%	20.54%	.00%	20.54%	100%	886
California	STRICT	40.91%	8.12%	49.03%	50.97%	.00%	50.97%	100%	308
	MOD	62.34%	8.12%	70.45%	29.55%	.00%	29.55%	100%	308
	LEAST	72.08%	8.12%	80.19%	19.81%	.00%	19.81%	100%	308
Florida	STRICT	47.15%	8.13%	55.28%	44.72%	.00%	44.72%	100%	123
	MOD	67.48%	8.13%	75.61%	24.39%	.00%	24.39%	100%	123
	LEAST	82.93%	8.13%	91.06%	8.94%	.00%	8.94%	100%	123
New York	STRICT	28.24%	13.74%	41.98%	58.02%	.00%	58.02%	100%	131
	MOD	50.38%	13.74%	64.12%	35.88%	.00%	35.88%	100%	131
	LEAST	64.89%	13.74%	78.63%	21.37%	.00%	21.37%	100%	131
North Carolina	STRICT	26.58%	12.66%	39.24%	60.76%	.00%	60.76%	100%	79
	MOD	36.71%	12.66%	49.37%	50.63%	.00%	50.63%	100%	79
	LEAST	43.04%	12.66%	55.70%	44.30%	.00%	44.30%	100%	79
Ohio	STRICT	53.28%	8.20%	61.48%	38.52%	.00%	38.52%	100%	122
	MOD	65.57%	8.20%	73.77%	26.23%	.00%	26.23%	100%	122
	LEAST	74.59%	8.20%	82.79%	17.21%	.00%	17.21%	100%	122
Pennsylvania	STRICT	50.41%	9.76%	60.16%	39.84%	.00%	39.84%	100%	123
	MOD	63.41%	9.76%	73.17%	26.83%	.00%	26.83%	100%	123
	LEAST	69.11%	9.76%	78.86%	21.14%	.00%	21.14%	100%	123

Table 19

Consistencies and Inconsistencies Between Self-Reports and Government Records of Turnout For Respondents Matched To a Government Record

State	Criteria	Consistencies			Inconsistencies				
		Self-reports and government records agree: turned out	Self-reports and government records agree: did not	Total	Respondent: turned out Govt Records: did not	Respondent: did not Govt Records: turned out	Total	Total	N
All target states	STRICT	89.56%	3.16%	92.72%	7.28%	.00%	7.28%	100%	412
	MOD	90.26%	2.39%	92.65%	7.35%	.00%	7.35%	100%	585
	LEAST	89.71%	2.75%	92.46%	7.54%	.00%	7.54%	100%	690
California	STRICT	94.74%	1.50%	96.24%	3.76%	.00%	3.76%	100%	133
	MOD	93.20%	2.43%	95.63%	4.37%	.00%	4.37%	100%	206
	LEAST	91.36%	2.47%	93.83%	6.17%	.00%	6.17%	100%	243
Florida	STRICT	95.08%	1.64%	96.72%	3.28%	.00%	3.28%	100%	61
	MOD	96.51%	1.16%	97.67%	2.33%	.00%	2.33%	100%	86
	LEAST	96.23%	1.89%	98.11%	1.89%	.00%	1.89%	100%	106
New York	STRICT	80.43%	15.22%	95.65%	4.35%	.00%	4.35%	100%	46
	MOD	91.67%	4.17%	95.83%	4.17%	.00%	4.17%	100%	72
	LEAST	91.40%	4.30%	95.70%	4.30%	.00%	4.30%	100%	93
North Carolina	STRICT	65.63%	6.25%	71.88%	28.13%	.00%	28.13%	100%	32
	MOD	65.91%	4.55%	70.45%	29.55%	.00%	29.55%	100%	44
	LEAST	69.39%	4.08%	73.47%	26.53%	.00%	26.53%	100%	49
Ohio	STRICT	87.84%	1.35%	89.19%	10.81%	.00%	10.81%	100%	74
	MOD	88.89%	1.11%	90.00%	10.00%	.00%	10.00%	100%	90
	LEAST	86.67%	2.86%	89.52%	10.48%	.00%	10.48%	100%	105
Pennsylvania	STRICT	93.94%	.00%	93.94%	6.06%	.00%	6.06%	100%	66
	MOD	89.66%	2.30%	91.95%	8.05%	.00%	8.05%	100%	87
	LEAST	90.43%	2.13%	92.55%	7.45%	.00%	7.45%	100%	94

Table 20

Probit Regressions Estimating Turnout Using Various Measures

| | Self-report turnout | Turnout based on matching criteria | | |
| | | STRICT | MOD | LEAST |
Explanatory Variable	(a)	(b)	(c)	(d)
Income	.29	.19	.29	.20
	(.31)	(.19)	(.19)	(.20)
Education	.92**$_b$	-.18$_{acd}$	.30$_b$	.51*$_b$
	(.34)	(.24)	(.24)	(.25)
Unemployed	-.36	-.45$_c$	-.06$_b$	-.21
	(.35)	(.30)	(.28)	(.28)
Age	1.35	-.23	1.77	2.20
	(2.03)	(1.37)	(1.38)	(1.42)
Age squared	-.06	.48	-1.15	-1.41
	(2.18)	(1.39)	(1.40)	(1.46)
External Efficacy	.08	.01	.09	.05
	(.34)	(.20)	(.21)	(.22)
Strength of party identification	1.02***$_{bcd}$	.02$_a$	.05	.23$_a$
	(.26)	(.17)	(.17)	(.18)
Preference for a political party	-.04	-.08	-.16	-.10
	(.28)	(.17)	(.17)	(.18)
Care who wins presidential election	1.85***$_{bcd}$	.74***$_{ac}$	1.01***$_{ab}$	.89***$_a$
	(.30)	(.22)	(.22)	(.22)
Preference for a presidential candidate	-.36	-.05	-.15	-.12
	(.29)	(.18)	(.18)	(.19)
Church attendance	-.16	-.10	-.08	-.17
	(.19)	(.11)	(.12)	(.12)
Homeowners	.27	.33*	.22[+]	.23[+]
	(.19)	(.13)	(.12)	(.13)
Currently working	.28	-.07	-.03	-.05
	(.19)	(.12)	(.12)	(.13)
Perceived closeness of election	-.09	.19	.23[+]	.07
	(.21)	(.12)	(.12)	(.13)
Live in a Southern state	-.26	-.17	-.23*	-.21[+]
	(.18)	(.11)	(.11)	(.12)
Blacks	-.23	.02	-.22	-.33[+]
	(.25)	(.18)	(.17)	(.17)
Hispanics	.57	.11	.19	.13
	(.36)	(.19)	(.20)	(.20)
Constant	-2.11***	-1.05**	-1.57***	-1.38***
	(.51)	(.35)	(.35)	(.35)
Model χ^2	146.77	35.77	66.32	75.44
N	818	818	818	818

Note: Cell entries are probit coefficients with standard errors in parentheses underneath. Subscripts on a coefficient indicate which other column's coefficients in that row are significantly different from it (p<.05).
[+] p<.10, * p<.05, ** p<.01, *** p<.001