THE PRACTICAL RESEARCHER

The Value of Voterfiles for U.S. State Politics Research

Christopher A. Cooper, *Western Carolina University*
Moshe Haspel, *Emory University*
H. Gibbs Knotts, *Western Carolina University*

ABSTRACT

State voter registration lists, or "voterfiles," provide a rich and under-utilized data source for state politics scholars. Voterfiles present a number of advantages over traditional surveys for answering important questions about state politics and policy. We present an overview of state voterfiles and include a short discussion of the history of voterfiles. Through internet searches and structured interviews with state election officials, we compile, for the first time, a complete account of the information that is available on voterfiles in the fifty states and the District of Columbia. We also present a number of best practices for obtaining and working with state voterfiles.

FOR STATE POLITICS SCHOLARS, studies of voter turnout have taken on renewed importance because of new and innovative state laws designed to increase voter registration and participation. States have employed early voting methods, same-day registration, liberalized absentee voting, and precinctless voting methods such as voting by mail and Election Day Vote Centers outside of traditional precinct locations. While research has assessed the success of these new laws, studies frequently rely on survey data, presenting a number of limitations for state politics scholars.

To overcome these limitations, we suggest that researchers diversify their data portfolios and consider the use of statewide voter registration lists or "voterfiles." Voterfiles provide an account of all of the registered voters in a state and often include a number of important characteristics for each voter (McDonald 2007). Used alone or in conjunction with other data sources, voterfiles can help researchers better understand the impact of state voting laws as well as assess a number of other questions of interest to state politics scholars.

State Politics and Policy Quarterly, Vol. 9, No. 1 (Spring 2009): pp. 102–121

THE LIMITATIONS OF SURVEY DATA

Surveys, like the American National Election Studies (ANES), Current Population Surveys (CPS), and exit polls, have fueled the academic literature since the early 1950s. Studies based on these surveys have produced a rich tradition of research, covering a range of questions about the views and characteristics of individual citizens. In addition to demographic traits such as race, sex, education, and income, the surveys include questions about ideology, efficacy, and a variety of other political and policy questions. To improve on the work done by survey researchers, scholars have married survey data with environmental and contextual factors to develop a more complete understanding of voter turnout and political participation (Huckfeldt and Sprague 1995; Oliver 2000).

This reliance on surveys has three key limitations. First, survey respondents self report their political activities, such as voting. The tendency to overreport—for a voter to claim that he voted when in fact he did not—has been well documented in the literature (McDonald 2003; Silver, Anderson, and Abramson 1986; Wright 1992; but see also, Gronke 1992). Given recent evidence that the level of overreporting varies systematically by context (Karp and Brockington 2005), this limitation may lead to misleading conclusions about electoral reforms and political participation.

Second, the quality of the sample could be problematic when using survey data. Even when analyzing the well-respected ANES, researchers often find it difficult to make meaningful inferences about the political activities of subgroups such as racial and ethnic minorities, young adults, or the very wealthy. Sample quality is particularly troublesome in an age where cell phones, caller ID, and a general culture of distrust have reduced response rates, thereby making accurate and representative samples more difficult to achieve (Groves 2006; Keeter et al. 2006).

A third limitation of survey research is of particular interest to state politics scholars. The analysis of national polls makes it difficult to understand the decision to participate in state or local elections. For example, the ANES provides too few observations to analyze participation in the North Carolina governor's race, a key state legislative election, or a local school board contest. Even when researchers pool state-level survey data over time to create larger sample sizes, they are still faced with relatively small samples that vary across states (Erikson, Wright, and McIver 1993; Brace et al. 2002). Scholars interested in state legislative district races or other elections in small political units, face an even more daunting set of obstacles.

HISTORY AND USE OF VOTERFILES

Voterlists, an alternative research tool to survey data, have existed as long as there has been voter registration. The history of voterlists dates back to the early nineteenth century (Keyssar 2001). During this time, white male property owners were on the lists and other citizens were excluded. Voter registration typically occurred at the local level and the voterlist was maintained using pen and paper. In the late 1920s, many states proposed laws requiring permanent voter registration at a centralized location rather than registering at precincts (Harris 1928).

More recently, federal legislation has greatly influenced the creation and maintenance of voterfiles. The National Voter Registration Act (NVRA) of 1994, also known as Motor Voter, required states to allow citizens to register to vote when they obtained or renewed their driver's license, or when they received other types of public assistance. However, by moving the power to register from election officials to other bureaucrats, clerical mistakes in voter registration occur more frequently (Alvarez 2005). The Motor Voter legislation also established guidelines to maintain and purge voterlists, although states still enjoy some latitude in these areas.

The Help America Vote Act (HAVA) of 2002 included a number of provisions to improve the creation and maintenance of voter registration lists. Title III, Section 303, "Computerized Statewide Voter Registration List Requirements and Requirements for Voters Who Register by Mail," requires each state to "implement, in a uniform and nondiscriminatory manner, a single, uniform, official, centralized, interactive computerized state-wide voter registration list defined, maintained, and administered at the State level that contains the name and registration information of every legally registered voter in the State and assigns a unique identifier to every legally registered voter in the State." States have scrambled since 2002 to become HAVA compliant, and they have succeeded to varying degrees.[1] The Brennan Center's "Making the List" report provides an excellent overview of how states create and maintain their database records (Levitt, Weiser, and Munoz 2006).

Because of their rich content, a number of important uses for these data have developed over time. For example, political candidates have long relied on voterfiles for campaign mailings and fundraising efforts. The best way to determine likely supporters is to know who is registered to vote and who voted in previous elections. Not surprisingly, voterfiles are important resources for campaign consultants as well. Companies such as Aristotle have provided information to candidates, based largely on voterfiles, for more than 20 years. In addition, candidates and parties have increasingly used voterfiles to micro-

target potential supporters (Green 2006). Recent elections have been marked by increased use of voterfiles by both Democrats and Republicans, although Republicans have led the way (Vander Veen 2006). The power of the information contained in voterfiles has even been used in consumer marketing, although some states do have restrictions on this use, as we note below.

Academic research using voter registration lists dates back to at least 1926. Early studies used voterfiles to better understand voter mobilization (Gosnell 1926) and to determine the accuracy of survey organizations (Cantril 1937). There is also a long history of using voterlists data to validate self-reported voting information from survey respondents (Kitt and Gleicher 1950; Traugott 1989; Traugott, Traugott, and Presser 1992). Voterlist maintenance in the 1980s has also been analyzed by Abramson and Claggett (1992).

More recently, scholars have used voterfile data to address a range of issues related to political behavior. As examples, voterfile data has been used to better understand concentric electorates (Sigelman and Jewell 1986); the impact of distance to the polling place on turnout (Dyck and Gimpel 2005; Gimpel and Schuknecht 2003; Haspel and Knotts 2005); absentee and early voting methods (Stein and Vonnahme 2008); the influence of gentrification on voter turnout (Knotts and Haspel 2006); the political behavior of late registrants (Gimpel, Dyck, and Shaw 2007); the spatial composition of convenience voting (Gimpel, Dyck, and Shaw 2006); the influence of statewide registration portability on voter turnout (McDonald 2008); the prevalence of voter fraud (McDonald and Levitt 2008); the composition of the electorate in California's 2003 recall (Arbour and Hayes 2005); and the effects of politically heterogeneous households on vote choice (Belanger and Eagles 2007). Increasingly, scholars have used voterfiles to generate lists of registered voters for field experiments (Arceneaux, Gerber, and Green 2006; Gerber and Green 2000; Nickerson 2007). Companies such as Polimetrix also use registration-based sampling in their surveys. Voterfiles have been used in other disciplines ranging from economics (Hastings et al. 2007) to epidemiology (Adimora et al. 2001; Foote et al. 2003). The breadth of studies above suggests that voterfiles are not new to political science or to academic researchers in general. Nonetheless, they have been used to investigate only a small number of the potential research questions that can be answered using such data.

DATA AND METHODS

An important factor inhibiting research using state voterfiles is the lack of a central information repository for academics. To determine what information is available in state voterfiles, we first searched state board of election

websites in the fifty states and the District of Columbia. Some of the sites contained all of the information that we needed. For example, in Ohio the entire voterfile can be downloaded free of charge. In most cases, however, we had to follow-up our internet search with structured phone interviews with election officials and sometimes with subsequent email correspondence.

Although time-consuming, this process has provided us with the most up-to-date and comprehensive data available about state voterfiles. We asked election officials about five types of variables included in the voterfile: access, contact information, district information, demographics, and information about voting. We should note that our research focuses specifically on statewide voterfiles. Voterfiles are often available at the local level and include more information, but they might have different restrictions than statewide files. For instance, the Massachusetts voterfile cannot be used by academic researchers but files from particular townships may be used for scholarly purposes.

INFORMATION AVAILABLE IN VOTERFILES

Table 1 reports access to statewide voterfiles in the United States. All states, except Arizona, have a statewide voterfile with some degree of public availability.[2] In Arizona, a good deal of voterfile data can still be obtained from each of the 15 county election offices. Table 1 also demonstrates that the price of voterfiles ranges dramatically. The mean cost is $2,279, and the median cost is $300. Voterfiles in Ohio, Vermont, and Massachusetts are free. In 16 cases the statewide voterfile costs $100 or less, and in 33 instances the price is $1000 or less. While many states are relatively inexpensive, some states, such as Alabama, West Virginia, and Wisconsin, are quite costly. The Alabama voterfile, at $27,000, is the most expensive.

Academic researchers will be particularly interested in restrictions placed on the use of voterfile data. As detailed in Table 1, 19 states place no restrictions on the use of voterfiles. The most common restriction is that voterfiles not be used for commercial purposes. In total, 17 states placed this restriction on voterfiles. Ten states indicate that their voterfile may be used for political purposes only. Kentucky and Virginia include a special provision indicating that their voterfile may not be used for academic purposes. Overall, the news about access is good for political scientists. There are no restrictions placed upon using voterfile data for academic purposes in 36 states, opening up a rich data source for the majority of states.[3]

Table 2 describes voter contact information available in statewide voterfiles. This information is important to candidates and political consultants,

Table 1. Access, Cost, and Restrictions

State	Voterfile Available?	Cost[1]	Restrictions[2]	State	Voterfile Available?	Cost	Restrictions[2]
Alabama	Yes	$27,000	A	Montana	Yes	$1,000	A
Alaska	Yes	$178	A	Nebraska	Yes	$500	B
Arizona	No	N/A	N/A	Nevada	Yes	$106.62	A
Arkansas	Yes	$2.50	A	New Hampshire	Yes	$17,837.50	D
California	Yes	$30	B	New Jersey	Yes	$53	A
Colorado	Yes	$500	A	New Mexico	Yes	$4,000	B
Connecticut	Yes	$300	A	New York	Yes	$0	F
Delaware	Yes	$250	A	North Carolina	Yes	$25	A
D.C.	Yes	$60	A	North Dakota	Yes	No price currently set	D
Florida	Yes	$10	A	Ohio	Yes	$0	A
Georgia	Yes	$500	B	Oklahoma	Yes	$150	A
Hawaii	Yes	$250	B	Oregon	Yes	$500	B
Idaho	Yes	$20	B	Pennsylvania	Yes	$20	B
Illinois	Yes	$2,000	B	Rhode Island	Yes	$50	A
Indiana	Yes	$500	H	South Carolina	Yes	$1,795	D
Iowa	Yes	$1,000	B	South Dakota	Yes	$2,500	F
Kansas	Yes	$200	B	Tennessee	Yes	$2,500	F
Kentucky	Yes	$450	C	Texas	Yes	$1,100	B
Louisiana	Yes	$5,000	A	Utah	Yes	$1,050	A
Maine	Yes	$2,000	B	Vermont	Yes	$0	B
Maryland	Yes	$125	B	Virginia	Yes	$4,330	G
Massachusetts	Yes	$0	D	Washington	Yes	$30	F
Michigan	Yes	$22.08	A	West Virginia	Yes	$14,025	B
Minnesota	Yes	$46	E	Wisconsin	Yes	$12,500	A
Mississippi	Yes	$2,100	B	Wyoming	Yes	$50	F
Missouri	Yes	$127	B	Percent Yes	98%		

1. Costs for New Hampshire and West Virginia were calculated using the formula they provided, assuming the number of registered voters as reported in 2004 by the U.S. Census Bureau.

2. A=None; B=No commercial use; C=At their discretion. Cannot use for scholarly, journalistic, political, or governmental purposes; D=Parties and/or Official candidates only; E= Elections, political activities, or law enforcement only. Must be a registered voter in Minnesota; F= Political purposes only; G= Only for courts for jury selection, candidates or political parties, party committees, PACs, incumbents, non-profit organizations, local government—Academics cannot use; H=available in extremely limited circumstances.

Table 2. Registrant Contact Information

State	Residential Address	Mailing Address	Telephone Number	Voter Registration Number
Alabama	Yes	Yes	Yes	Yes
Alaska	Yes	Yes	No	Yes
Arkansas	Yes	Yes	Yes	Yes
California	Yes	Yes	Yes	Yes
Colorado	Yes	Yes	Yes	Yes
Connecticut	Yes	Yes	Yes	Yes
Delaware	Yes	Yes	Yes	Yes
D.C.	Yes	Yes	No	No
Florida	Yes	Yes	Yes	Yes
Georgia	Yes	Yes	No	Yes
Hawaii	Yes	Yes	No	No
Idaho	Yes	Yes	Yes	Yes
Illinois	Yes	Yes	Yes	Yes
Indiana	Yes	Yes	Yes	Yes
Iowa	Yes	Yes	Yes	Yes
Kansas	Yes	Yes	Yes	Yes
Kentucky	Yes	Yes	No	No
Louisiana	Yes	Yes	Yes	Yes
Maine	Yes	No	No	Yes
Maryland	Yes	Yes	No	Yes
Massachusetts	Yes	Yes	Yes	Yes
Michigan	Yes	Yes	No	Yes
Minnesota	Yes	No	Yes	Yes
Mississippi	Yes	Yes	No	Yes
Missouri	Yes	No	No	Yes
Montana	Yes	Yes	Yes	Yes
Nebraska	Yes	Yes	No	Yes
Nevada	Yes	Yes	Yes	Yes
New Hampshire	Yes	Yes	No	Yes
New Jersey	Yes	Yes	No	Yes
New Mexico	Yes	Yes	Yes	Yes
New York	Yes	Yes	No	Yes
North Carolina	Yes	Yes	Yes	Yes
North Dakota	Yes	Yes	No	Yes
Ohio	Yes	Yes	No	Yes
Oklahoma	Yes	Yes	No	Yes
Oregon	Yes	Yes	Yes	Yes
Pennsylvania	Yes	Yes	No	Yes
Rhode Island	Yes	Yes	Yes	Yes
South Carolina	Yes	Yes	No	Yes
South Dakota	Yes	Yes	Yes	Yes
Tennessee	Yes	Yes	No	Yes
Texas	Yes	Yes	Yes	Yes
Utah	Yes	Yes	Yes	Yes
Vermont	No	Yes	Yes	Yes
Virginia	Yes	Yes	No	No
Washington	Yes	Yes	No	Yes
West Virginia	Yes	Yes	No	No
Wisconsin	Yes	Yes	No	No
Wyoming	Yes	Yes	No	No
Percent Yes	98%	94%	50%	86%

but it might also be useful for researchers who wish to marry voterfile data with survey or content data. As displayed in Table 2, nearly all states report both a residential and a mailing address. The residential address could be useful for those who wish to learn more about contextual influences on voting behavior, while the mailing address might be of interest to those who wish to use voterfiles to survey registered voters, or to contact them in a field experiment (e.g. Gerber and Green 2000). While addresses are reported in virtually every state, phone numbers are not. In total, 25 states provide a phone number on the voterfile and 25 states do not. Obviously, phone numbers are important for those who want to poll registered voters or who wish to follow up with them personally in order to assess the influence of experimental manipulations. In addition, 43 states have a unique voter registration number. In what promises to be a more common practice, Rhode Island and California currently include an optional email address field for registrants, providing a potentially low cost means of surveying registered voters or attempting to turn out the vote.[4]

Table 3 displays precinct and district information included in each state's voterfile. We indicate whether the information is present or absent, even if the variable in question does not apply to that state. For example, not all states have judicial elections, and some states only have one member of Congress, making congressional district lines unnecessary. As Table 3 reports, all but two states indicate the voting district or precinct on the statewide voterfile. Considering recent studies showing the importance of distance to the polling place (Dyck and Gimpel 2005; Haspel and Knotts 2005) and even the type of polling place (Berger, Meredith, and Wheeler 2006; Stein and Vonnahme 2008), this variable promises to be important for future research. Forty-one states also report a registrant's state house and senate district. This information is particularly important for scholars of state elections and legislatures. Whereas state politics scholars have long relied on static measures, such as the *Almanac of State Legislatures* (Lilley and DeFranco 1998; Lilley et al. 2007), to determine state legislative district preferences, voterfiles provide an opportunity for scholars to easily (and in many states, quite cheaply) determine the distribution of party identification by state legislative districts. Table 3 also suggests that the vast majority of states include congressional districts, and many also include judicial and other districts. In fact, 19 states include judicial district on the voterfile. Some voterfiles even include school board and other minor districts, rapidly expanding the data availability for scholars interested in political behavior and representation in local elections.

Voterfiles provide a wealth of consistent information about the location and address of voters, but surveys still provide the most extensive demographic

Table 3. District Information

State	Precinct	Congressional	State Legislative	Judicial	Other Districts
Alabama	Yes	Yes	Yes	No	Yes
Alaska	Yes	Yes	Yes	No	No
Arkansas	Yes	Yes	Yes	Yes	Yes
California	Yes	No	No	No	No
Colorado	Yes	Yes	Yes	No	Yes
Connecticut	Yes	Yes	Yes	No	No
Delaware	Yes	No	Yes	No	No
D.C.	Yes	No	No	No	Yes
Florida	Yes	Yes	Yes	No	Yes
Georgia	Yes	Yes	Yes	Yes	Yes
Hawaii	Yes	No	Yes	No	Yes
Idaho	Yes	No	Yes	No	No
Illinois	Yes	Yes	Yes	Yes	No
Indiana	Yes	Yes	Yes	No	No
Iowa	Yes	Yes	Yes	Yes	Yes
Kansas	Yes	Yes	Yes	Yes	Yes
Kentucky	Yes	No	No	No	No
Louisiana	Yes	Yes	Yes	Yes	Yes
Maine	Yes	No	Yes	No	Yes
Maryland	Yes	Yes	Yes	No	No
Massachusetts	Yes	Yes	Yes	No	No
Michigan	Yes	Yes	Yes	No	Yes
Minnesota	Yes	Yes	Yes	No	Yes
Mississippi	Yes	No	No	No	No
Missouri	Yes	Yes	Yes	No	No
Montana	Yes	No	No	Yes	No
Nebraska	Yes	Yes	Yes	Yes	Yes
Nevada	Yes	Yes	Yes	Yes	No
New Hampshire	No	No	No	No	No
New Jersey	Yes	No	No	No	No
New Mexico	Yes	Yes	Yes	Yes	No
New York	Yes	Yes	Yes	No	No
North Carolina	Yes	Yes	Yes	Yes	Yes
North Dakota	No	No	No	No	No
Ohio	Yes	Yes	Yes	Yes	Yes
Oklahoma	Yes	Yes	Yes	No	No
Oregon	Yes	Yes	Yes	Yes	Yes
Pennsylvania	Yes	Yes	Yes	Yes	No
Rhode Island	Yes	No	Yes	No	Yes
South Carolina	Yes	Yes	Yes	No	No
South Dakota	Yes	No	Yes	No	Yes
Tennessee	Yes	Yes	Yes	Yes	Yes
Texas	Yes	Yes	Yes	Yes	Yes
Utah	Yes	Yes	Yes	Yes	No
Vermont	Yes	No	Yes	No	Yes
Virginia	Yes	Yes	Yes	No	No
Washington	Yes	No	No	No	No
West Virginia	Yes	Yes	Yes	Yes	No
Wisconsin	Yes	Yes	Yes	Yes	Yes
Wyoming	Yes	No	Yes	No	Yes
Percent Yes	96%	66%	82%	38%	48%

Table 4. Demographics

State	Gender	Age/Year Born	Race	Birthplace
Alabama	Yes	Yes	Yes	Yes
Alaska	Yes	No	No	No
Arkansas	No	Yes	No	No
California	Yes	Yes	No	Yes
Colorado	Yes	Yes	No	No
Connecticut	Yes	Yes	No	No
Delaware	No	No	No	No
D.C.	No	No	No	No
Florida	Yes	Yes	Yes	No
Georgia	Yes	Yes	Yes	No
Hawaii	Yes	No	No	No
Idaho	Yes	Yes	No	No
Illinois	Yes	Yes	No	No
Indiana	Yes	Yes	No	No
Iowa	Yes	Yes	No	No
Kansas	Yes	Yes	No	No
Kentucky	Yes	Yes	No	No
Louisiana	Yes	Yes	Yes	No
Maine	No	Yes	No	No
Maryland	Yes	Yes	No	No
Massachusetts	Yes	Yes	No	No
Michigan	Yes	Yes	No	No
Minnesota	No	Yes	No	No
Mississippi	No	No	No	No
Missouri	No	No	No	No
Montana	No	Yes	No	No
Nebraska	No	Yes	No	No
Nevada	No	Yes	No	No
New Hampshire	No	No	No	No
New Jersey	Yes	Yes	No	No
New Mexico	Yes	Yes	No	No
New York	Yes	Yes	No	No
North Carolina	Yes	Yes	Yes	Yes
North Dakota	No	No	No	No
Ohio	No	Yes	No	No
Oklahoma	No	Yes	No	No
Oregon	No	Yes	No	No
Pennsylvania	Yes	Yes	No	No
Rhode Island	Yes	Yes	No	No
South Carolina	Yes	Yes	Yes	No
South Dakota	No	Yes	No	No
Tennessee	Yes	Yes	Yes	No
Texas	Yes	Yes	Yes	Yes
Utah	No	Yes	No	No
Vermont	No	Yes	No	No
Virginia	Yes	Yes	No	No
Washington	Yes	Yes	No	No
West Virginia	No	No	No	No
Wisconsin	No	No	No	No
Wyoming	No	No	No	No
Percent Yes	58%	78%	16%	8%

information.[5] As Table 4 suggests, the demographic information available in state voterfiles varies tremendously by states. Three states (Alabama, North Carolina, and Texas) include all of the demographic information we recorded. Eight states (Delaware, Mississippi, Missouri, New Hampshire, North Dakota, West Virginia, Wisconsin, and Wyoming) and the District of Columbia include none of this information. In terms of specific demographics, 29 include gender. The registrant's age or birth year is included in 39 states, a sizeable number, but not as many as suggested by McDonald's (2007) sample of 11 states. Although we do not list it here, some states even include the birth day, month, and year. Eight states include race (Alabama, Florida, Georgia, Louisiana, North Carolina, South Carolina, Tennessee, and Texas). All of these states are required by the Federal Voting Rights Act to register voters by race and are clustered in the South.[6] Alabama, California, North Carolina, and Texas also include place of birth. Although it is available only in a limited number of states, this variable can allow scholars of state politics to learn about the differences between natives and non-natives.

Perhaps the most valuable information available on voterfiles is past voting history. Election scholars often lack data on down-ballot elections in odd-numbered years, primary voting, and individual-level participation over multiple elections. Most studies are drawn from cross-sectional surveys, which rarely include data about down-ballot elections or primary voting and are unable to measure participation over time. Panel studies measure participation over time but still rely on self-reported behavior. In addition, they are typically concerned with presidential voting and are prohibitively expensive. Fortunately, as Table 5 suggests, voterfiles can solve many of these problems. Voter histories are available in almost every state voterfile.[7] Some states include history for 10–15 years, while others only cover the most recent election.[8] The date a person registers to vote appears in 37 states, which opens up a number of potential research questions regarding the relationship between registration and turnout in the states. In addition, party affiliation is available in many states. Measuring party registration creates a number of problems, however. In some states, people are not allowed to register by party. In others, people declare their partisanship at the time of registration. In a third group of states, partisanship is reported based on voting in the most recent party primary. Because of these inconsistencies, we do not include party in our tables, but we recommend that interested researchers contact the state boards of elections directly to determine the availability and format of this information.

Research on early voting, absentee voting, and electoral reform has recently exploded as states experiment with new policies aimed at increasing turnout. This research generally relies on either aggregated voter data or reports from

Table 5. Voting

State	Date Registered	Voter History	Absentee	Other Special Voting
Alabama	Yes	Yes	No	No
Alaska	No	Yes	Yes	No
Arkansas	Yes	Yes	Yes	Yes
California	Yes	Yes	No	No
Colorado	Yes	No	No	No
Connecticut	Yes	No	Yes	No
Delaware	Yes	Yes	No	No
D.C.	Yes	Yes	No	No
Florida	Yes	Yes	Yes	Yes
Georgia	Yes	Yes	Yes	No
Hawaii [1]	Yes	Yes	Yes[1]	Yes[1]
Idaho	Yes	Yes	No	No
Illinois	Yes	Yes	No	No
Indiana	Yes	Yes	Yes	No
Iowa	No	Yes	Yes	Yes
Kansas	Yes	Yes	No	No
Kentucky	Yes	Yes	No	No
Louisiana	Yes	Yes	No	No
Maine	Yes	Yes	Yes	Yes
Maryland	Yes	Yes	No	No
Massachusetts	No	Yes	No	No
Michigan	Yes	Yes	Yes	No
Minnesota	No	Yes	Yes	Yes
Mississippi	No	Yes	No	No
Missouri	No	Yes	No	No
Montana	Yes	No	No	No
Nebraska	Yes	Yes	No	No
Nevada	No	No	No	No
New Hampshire	No	No	No	No
New Jersey	No	Yes	No	No
New Mexico	Yes	Yes	Yes	Yes
New York	Yes	Yes	No	No
North Carolina	Yes	Yes	Yes	Yes
North Dakota	No	Yes	No	No
Ohio	Yes	Yes	No	No
Oklahoma	Yes	Yes	Yes	Yes
Oregon	Yes	Yes	No	No
Pennsylvania	Yes	Yes	Yes	No
Rhode Island	Yes	Yes	Yes	Yes
South Carolina	Yes	Yes	No	No
South Dakota	Yes	Yes	No	No
Tennessee	Yes	Yes	Yes	Yes
Texas	Yes	Yes	Yes	Yes
Utah	Yes	Yes	Yes	No
Vermont	Yes	No	No	No
Virginia	Yes	Yes	No	No
Washington	Yes	Yes	No	No
West Virginia	No	Yes	No	No
Wisconsin	No	Yes	No	No
Wyoming	No	No	No	No
Percent Yes	74%	86%	38%	24%

1. Hawaii lumps absentee and early voting together

survey data. Aggregated voter data can help answer these questions, but it leaves scholars unable to determine reliably the influence of individual-level factors. Inferences from survey data can include the biases mentioned above. Fortunately, voterfiles provide a wealth of information about different voting methods. For instance, information about whether an individual voted absentee appears in 19 states. At least 11 states report other types of special voting, such as early voting.

To gain a better overall sense of which variables are included most frequently in voterfiles, we calculated the percent of states that include each in their voterfiles. Figure 1 presents a dot plot summarizing this information. As you can see, contact information, such as residential address, precinct number, and mailing address is readily available, especially compared with demographic information such as birthplace and race. Voter history data is easily available in most states. Unfortunately, the availability of district information varies considerably across the states.

To understand which states report the most complete information in their voterfiles, we present the proportions in Figure 2. Each state's proportion is calculated by dividing the number of variables included on each state's voterfile by 17, the total number of variables we accounted for in our study. As you can see, the amount of information reported by each state varies

Figure 1. Proportion of States that Include Variables

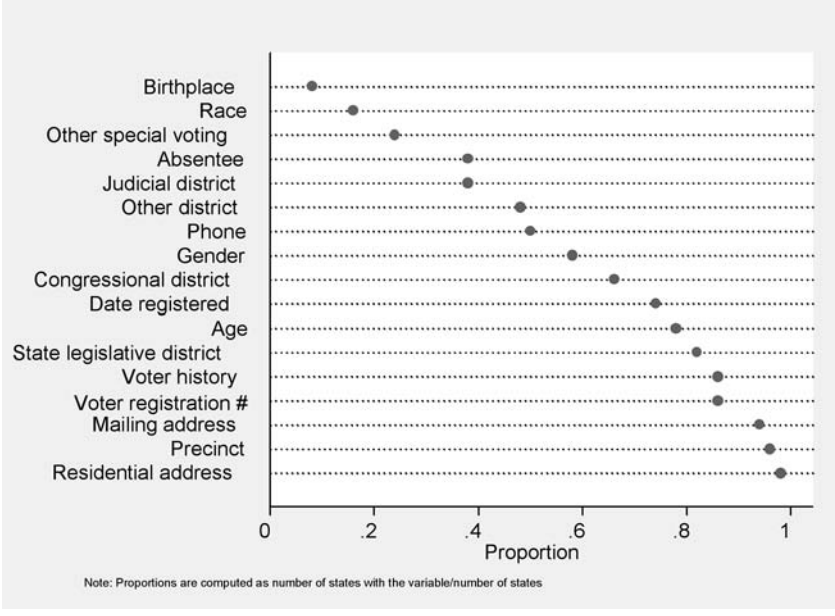
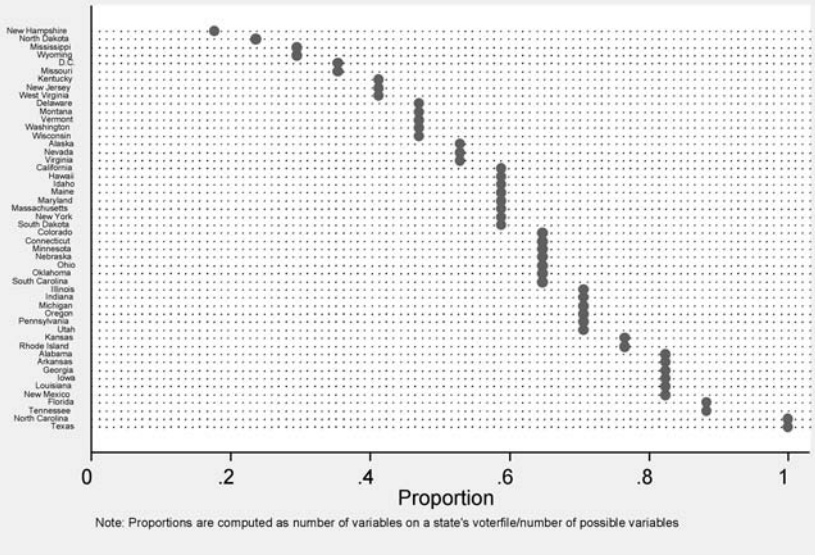


Figure 2. Proportion of Possible Variables by State



considerably. States such as Texas, North Carolina, and Tennessee include a great deal of information, while others like New Hampshire, North Dakota, and Mississippi include only about one-fifth of the variables.

We have reviewed the most commonly listed information on voterfiles, but some states include a variety of other useful data. For instance, South Dakota includes the date of last jury service, offering a valuable resource for those interested in jury service as an independent or dependent variable. The California voterfile includes information indicating whether the person asked for assistance in voting, as well as the language (English or some other named language) included on the ballot. Minnesota reports when absentee ballots are requested, transmitted, and returned.[9] Clearly, there are many interesting and important research questions that can be addressed using voterfiles.

PRACTICAL ADVICE FOR WORKING WITH VOTERFILES

When working with voterfiles, researchers must always keep in mind that they are conducting secondary analyses of data originally collected for administrative purposes, which has a number of implications. First, voterfiles are con-

stantly changing: new people register and others change residences every day. Therefore, one must be mindful of the timing of the voterfile request itself. Indeed, one might even need to acquire a voterfile twice: once immediately after the deadline to register for a particular election and a second time after the voter history for that same election has been added to the file.

Second, the manner in which voterfile data are collected could shift over time as administrative needs change. For example, the Georgia voterfile now includes Asian and Hispanic as racial categories. Yet these categories are not of much use to researchers as voters have only had the opportunity to register as Asian or Hispanic for less than a decade. Because of this, Asian and Hispanic voters today are systematically undercounted because one must re-register only upon changing addresses. The Voting Rights Act (for the states to which it applies), Motor Voter, and HAVA have all reduced the latitude that individual states have in how they handle voter registration issues, yet there is still variation across states in terms of eligibility (e.g., felon disenfranchisement), registration requirements (e.g., the deadline to register before an election), and how inactive voters are identified and purged.[10] The lesson here is that each file, and even seemingly straightforward variables in these files, might have quirks that must be investigated thoroughly.

Third, voterfiles tend to be noisy. When you register to vote, you fill out a paper form and the data from this form must be entered into a computer. In pre-Motor Voter days, you would have had to appear personally at the voter registration office or mail in your registration form, and if your form was not filled out legibly, an election official had the opportunity to verify your information. As mentioned above, the increased convenience that Motor Voter has brought in terms of registration almost certainly comes at some cost to measurement accuracy. Typographical errors can and do occur during the data entry process, and in contrast to the data entry process for a survey, there are typically few procedures in place for quality control or data cleaning. Voter history tends to be particularly noisy because most polling places are understaffed. Election Day workers are poorly paid and at best only moderately well-trained, and the data entry in many cases is conducted by temporary workers who have no incentive to enter the data well. However, registration cards provided to voters and the ability to look up voter registration information at many state board of election websites allow voters to identify these types of mistakes. Moreover, these data entry errors are mostly stochastic.[11] Voterfiles might also include errors based on which people are included. For example, there are often extraneous records in the data, including people who have moved out of the state or have died. These errors are likely not distributed evenly across the states

and it is highly likely that some states have more "deadwood" on their rolls than others.

Fourth, voterfiles are quite large. Whereas the ANES might have approximately 2,700 respondents, a voterfile will typically have orders of magnitude more data. The voterfile for Washington, D.C. alone has about 360,000 records, while that of North Carolina has over 5.4 million records. This cuts both ways. On one hand, this helps offset the noisy data issue. But it can also present special technical issues to the researcher as well. The full voterfile for many states is too big to store in a Microsoft Access database; even a pared-down voterfile often flirts with the upper limits of Stata. In addition to the costs of the voterfiles themselves, scholars seeking to work with voterfiles should budget for the purchase of a reasonably high-end computer with ample memory. MySQL, one of the popular open-source databases, is a good choice for data management because it is designed to accommodate very large datasets.

Fifth, voterfiles in many states have data structures unfamiliar to many scholars. Political scientists are accustomed to dealing with what are known as "flat files." But most voterfiles are broken into two pieces: the voter information, such as name and address, will be found in one data table, and voter histories will be stored in one or more separate tables. While this separation actually makes it easier to work with the data files, given their size, those unfamiliar with the concept of relational databases will need to become familiar with new concepts before conducting their analyses.

Sixth, voterfiles come in a variety of forms. As mentioned previously, the Ohio voterfile can be easily downloaded from the Internet. Most other voterfiles must be requested and are sent on a compact disk. The file formats include Access, fixed ASCII, delimited text, and DBASE. People who wish to work with voterfiles in multiple states would need to be proficient in multiple data types.

Finally, voterfiles only include people who register to vote. While this is an obvious point, the implications may be safely ignored or could be a fatal flaw depending on your particular research question. Surveys hold the clear advantage for answering any questions requiring data on people who are not in the electorate.

CONCLUSION

Research on voter turnout has been dominated by the analysis of survey data. However, survey data are not always appropriate because of the problems associated with small sample sizes, non-representative samples, and poten-

tially false reports of voter turnout. National surveys can be particularly problematic for state politics scholars because they do not allow researchers to analyze participation at state and local levels. In this paper, we argue that voterfiles are a rich and under-utilized data source that can solve many of the problems associated with traditional data.

Of course, voterfiles are not perfect. In addition to our cautionary notes, others point out that the quality of voterfiles varies by states (Atkeson et al. 2007; Green and Gerber 2006). Some states do an admirable job purging the voter rolls of inactive voters and voters who have moved. Others do much worse. Many voterfiles also contain a large amount of missing data on certain variables (McDonald 2007). Nonetheless, in a study of 11 states, McDonald (2007) finds that state voterfiles and the Current Population Survey have comparable demographics. As a result, we recommend that state politics scholars increase their use of voterfiles, both alone and alongside traditional data sources.

ENDNOTES

1. See electionline.org (2005) for an account of the ways states have approached HAVA compliance.

2. North Dakota does not require voters to register, but election officials keep a statewide record of who voted.

3. We encourage scholars to ask specifically about the use of voterfile data for academic research. In our discussions with election officials, "political purposes only" meant different things to different officials.

4. Of course, since phone numbers and email addresses are usually optional for voters to fill out, coverage will be uneven. It is best to consult with the elections division of the state of interest regarding the percentage of voters who have supplied this information before beginning research based on the use of these data.

5. We include information for the potential fields in a dataset. In some states, some of the demographic information is optional, at times producing unacceptable levels of missing data (McDonald 2007).

6. Mississippi is also covered by the Voting Rights Act but does not include race in the state voterfile made available to the public.

7. For some states, these data must be obtained at lower levels and then aggregated up to the state level.

8. It should be noted, however, that voter history compiled after HAVA tends to be much more reliable than data compiled in earlier years. In addition, voter histories that cover long time periods suffer from the fact that people are constantly being added to and removed from voterfiles, as discussed later in this paper.

9. We thank an anonymous reviewer for pointing out the resources available in Minnesota.

10. While Motor Voter set some standards for these procedures, meaningful variation

remains. Also some states, including Georgia, changed their purging plans in response to the law; this is another example of how data in a voterfile may not be perfectly comparable over time.

11. In states where voter histories are maintained by the counties, there may be a relationship between the economic well-being of a county and the integrity of the voter history data. Counties and municipalities that are pressed for resources are less likely to keep accurate records than those that can afford to do the job effectively. This means that one should be cautious when conducting validation studies examining the effects of race. Counties and municipalities with high percentages of minorities are likely to suffer from greater economic pressure and, as such, may not keep records as well as other, wealthier counties.

REFERENCES

Abramson, Paul R., and William Claggett. 1992. "The Quality of Record Keeping and Racial Differences in Validated Turnout." *The Journal of Politics* 54:871–80.

Adimora, Adaora A., Victor J. Schoenbach, Francis E. A. Martinson, Tonya R. Stancil, and Kathryn H. Donaldson. 2001. "Driver's License and Voter Registration Lists as Population-Based Sampling Frames for Rural African Americans." *Annals of Epidemiology* 11:385–8.

Alvarez, R. Michael. 2005. "Voter Registration: Past, Present, and Future." Written Testimony Prepared for the Commission on Federal Election Reform, Voter Technology Project, Working Paper No. 30.

Arbour, Brian K., and Danny Hayes. 2005. "Voter Turnout in the California Recall: Where did the Increase Come from?" *American Politics Research* 33:187–215.

Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2006. "Comparing Experimental and Matching Methods using a Large-Scale Voter Mobilization Experiment." *Political Analysis* 14:37–62.

Atkeson, Lonna Rae, Lisa A. Bryant, Thad E. Hall, Kyle L. Saunders, and R. Michael Alvarez. 2007. "New Barriers to Participation: Application of New Mexico's Voter Identification Law." Presented at the Meeting of the Annual Political Science Association Meeting, Washington, DC.

Belanger, Paul, and Munroe Eagles. 2007. "Partisan Cross-Pressure and Voter Turnout: The Influence of Micro and Macro Environments." *Social Science Quarterly* 88:850–67.

Berger, Jonah, Marc Meredith, and S. Christian Wheeler. 2006. "Can Where People Vote Influence How They Vote? The Influence of Polling Location Type on Voting Behavior." Research Paper Series, Stanford Graduate School of Business.

Brace, Paul, Kellie Sims-Butler, Kevin Arceneaux, and Martin Johnson. 2002. "Public Opinion in the American States: New Perspectives using National Survey Data." *American Journal of Political Science* 46:173–89.

Cantril, Hadley. 1937. "How Accurate Were the Polls?" *Public Opinion Quarterly* 1:97–109.

Dyck, Joshua J., and James G. Gimpel. 2005. "Distance, Turnout, and the Convenience of Voting." *Social Science Quarterly* 86:531–48.

Electionline.org. 2005. "Assorted Rolls: Statewide Voter Registration Databases Under HAVA." *Election Reform Briefing*. www.electionline.org/Portals/1/Assorted%20Rolls.pdf (June 2005).

Erickson, Robert A., Gerald C. Wright, and John P. McIver. 1993. *Statehouse Democracy: Public Opinion and Policy in the American States*. New York, NY: Cambridge University Press.

Foote, Janet A., Suzanne P. Murphy, Lynne R. Wilkens, Jean H. Hankin, Brian E. Henderson, and Laurence N. Kolonel. 2003. "Factors Associated with Dietary Supplement use among Healthy Adults of Five Ethnicities: The Multiethnic Cohort Study." *American Journal of Epidemiology* 157:888–97.

Gerber, Alan S., and Donald P. Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94:653–63

Gimpel, James G., and Jason E. Schuknecht. 2003. "Political Participation and the Accessibility of the Ballot Box." *Political Geography* 22:471–88.

Gimpel, James G., Joshua J. Dyck, and Daron R. Shaw. 2006. "Location, Knowledge, and Time Pressures in the Spatial Structure of Convenience Voting." *Electoral Studies* 25:35–58.

Gimpel, James G., Joshua J. Dyck, and Daron R. Shaw. 2007. "Election Year Stimuli and the Timing of Voter Registration." *Party Politics* 13:351–74.

Gosnell, Harold F. 1926. "An Experiment in the Stimulation of Voting." *American Political Science Review* 20:869–74.

Green, Donald A. 2006. "What do Social Scientists Know About What Makes Citizens Turn Out to Vote." *AAPSS Blog*. http://blog.aapss.org/index.cfm?catID=5.

Green, Donald P., and Alan S. Gerber. 2006. Can Registration-Based Sampling Improve the Accuracy of Midterm Election Forecasts? *Public Opinion Quarterly* 70:197–223.

Gronke, Paul. 1992. "Overreporting the Vote in the 1988 Senate Election Study: A Response to Wright." *Legislative Studies Quarterly* 17:113–29.

Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70:646–75.

Harris, Joseph P. 1928. "Permanent Registration of Voters." *American Political Science Review* 22:349–53.

Haspel, Moshe, and H. Gibbs Knotts. 2005. "Location, Location, Location: Precinct Placement and the Costs of Voting." *The Journal of Politics* 67:560–73.

Hastings, Justine S., Thomas J. Kane, Douglas O. Staiger, and Jeffrey M. Weinstein. 2007. "The Effect of Randomized School Admissions on Voter Participation." *Journal of Public Economics* 91:915–37.

Huckfeldt, Robert, and John Sprague. 1995. *Citizens, Politics, and Social Communication: Information and Influence in an Election Campaign.* New York, NY: Cambridge University Press.

Karp, Jeffrey A, and David Brockington. 2005. "Social Desirability and Response Validity: A Comparative Analysis of Over-Reporting Voter Turnout in Five Countries." *The Journal of Politics* 67:825–40

Keeter, Scott, Courtney Kennedy, Michael Dimock, Jonathan Best, and Peyton Craighill. 2006. "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey." *Public Opinion Quarterly* 70:759–79.

Keyssar, Alexander. 2001. *The Right to Vote: The Contested History of Democracy in America.* New York, NY: Basic Books.

Kitt, Alice S., and David B. Gleicher. 1950. "Determinants of Voting Behavior." *Public Opinion Quarterly* 14:393–412.

Knotts, H. Gibbs, and Moshe Haspel. 2006. "The Impact of Gentrification on Voter Turnout." *Social Science Quarterly* 87:110–21.

Levitt, Justin, Wendy R. Weiser, and Ana Munoz. 2006. "Making the List: Database Matching and Verification Processes for Voter Registration." New York, NY: Brennan Center for Justice at NYU School of Law. http://brennan.3cdn.net/96ee05284dfb6a6d5d_j4m6b1cjs.pdf.

Lilley, William III, and Laurence J. DeFranco. 1998. *Almanac of State Legislatures, Second Edition*. Washington, DC: CQ Press.

Lilley, William III, Laurence J. DeFranco, Mark F. Bernstein, and Kari L. Ramsby. 2007. *Almanac of State Legislative Elections, Third Edition*. Washington, DC: CQ Press.

McDonald, Michael P. 2003. "On the Overreport Bias of the National Election Study Turnout Rate." *Political Analysis* 11:180–6.

McDonald, Michael P. 2007. "The True Electorate: A Cross-Validation of Voter Registration Files and Election Survey Demographics." *Public Opinion Quarterly* 71:588–602.

McDonald, Michael P. 2008. "Portable Voter Registration." *Political Behavior* 30:491–501.

McDonald, Michael P., and Justin Levitt. 2008. "Seeing Double Voting: An Extension of the Birthday Problem." *Election Law Journal* 5:144–57.

Nickerson, David W. 2007. "Quality is Job One: Professional and Volunteer Voter Mobilization Calls." *American Journal of Political Science* 51:269–82.

Oliver, J. Eric. 2000. "City Size and Civic Involvement in Metropolitan America." *American Political Science Review* 94:361–73.

Sigelman, Lee, and Malcolm E. Jewell. 1986. "From Core to Periphery: A Note on the Imagery of Concentric Electorates." *The Journal of Politics* 48:440–9.

Silver, Brian D., Barbara A. Anderson, and Paul R. Abramson. 1986. "Who Overreports Voting?" *American Political Science Review* 80:613–24.

Stein, Robert, and Greg Vonnahme. 2008. "Engaging the Unengaged Voter: Voter Centers and Voter Turnout." *The Journal of Politics*. 70:487–97.

Traugott, Santa. 1989. "Validating Self-Reported Vote: 1964–1988." ANES Technical Report Series, No. nes010152.

Traugott, Michael W., Santa Traugott, and Stanley Presser. 1992. "Revalidation of Self-Reported Vote." ANES Technical Report Series, No. nes010160.

Vander Veen, Chad. 2006. "Zeroing In." *Government Technology*. www.govtech.com/gt/97723?id=&story_pg=2.

Wright, Gerald C. 1992. "Reported Versus Actual Vote: There is a Difference and it Matters." *Legislative Studies Quarterly* 17:131–42.