

Leveraging Historical Voter Files as Accurate Measures of Who Votes: Analyzing and Disseminating Voter File Data to Enhance Understanding of Elections*

Seo-young Silvia Kim[‡], Bernard L. Fraga[†], and Daron Shaw[§]

[‡]American University

[†]Emory University

[§]University of Texas at Austin

Background on Proposal

Voter files are essential tools for election administrators, campaigns, and, increasingly, election researchers. With changes to election laws due to and in the wake of the COVID-19 pandemic, however, individual-level records of who registers, who votes, and how individuals vote are under increased scrutiny and heightened demand from the public. In this project, we seek to develop best practices related to the use of voter file data through the construction and analysis of a multi-state historical voter file. In addition to developing best practices for researchers, we seek to answer two key research questions of importance for both academics and administrators: what is the effective rate of voter registration at the county level proximate to a given election, and how much individual-level variation do we see in the methods used to vote in recent elections?

Project Status

Thanks to funding from MEDSL, we have partnered with a voter file vendor (L2) and gained access to their repository of state-level historical voter registration and vote history lists. These raw, unprocessed lists are in the format they were delivered to L2 by state officials; they are not the processed and standardized versions generally used by academics conducting national analyses. The snapshots include at least one version of the state's registration list for every year since 2016, and many states have multiple snapshots per year from 2013 onward.

We gained access to the data in the first week of January 2023. So far, we have downloaded and begun to analyze each state's raw registration list closest to the November 2020 election. This will help us answer our first research question voter registration rates proximate to election day, while developing an understanding of how to work with unprocessed voter file data at a national level.

*Prepared for presentation at the 2023 MEDSL Pre-Conference Workshop at the Southern Political Science Association Annual Meeting, St. Petersburg, FL.

An Initial Look at Voter Registration

We begin by assessing the correspondence between official counts of registered voters as relayed by election officials in the EAVS and official counts as reflected in voter registration files contemporaneous to the November 2020 election. As these files are not standardized, we anticipated challenges in processing the raw data to a point where we could verify that the lists were complete and organized in a rectangular, individual-level fashion. Indeed, we understood this standardization to be a major value add for third-party data vendors (Hersh 2015). However, the relatively recent data that we worked with were more similar than we expected. In this initial report, we draw on the raw snapshots from 21 states, covering 53,981,912 registered voters.

For this preliminary analysis, we selected state-level snapshots of the voter registration lists as close as possible to November 2020. In approximately half of the states these snapshots reflect the voter registration list as of October 2020. Most of the remaining state snapshots are from September, November, or December of 2020. In rare instances we relied on snapshots from the summer of 2020 or early 2021.

Our methodology for generating counts of registered voters is simple: count the number of rows in the state’s voter registration list. This does not account for active versus inactive registrants, the potential presence of canceled entries on the registration list, duplicate entries, or “deadwood,” all of which contribute to discrepancies between official records of, e.g., number of voting-eligible residents in a state and the number of registrants (Ansolabehere and Hersh 2014). However, at this stage the preliminary nature of our analysis aligns with a desire to reduce researcher degrees of freedom.

We compare these raw counts to administrator-reported voter registration totals as reported in the U.S. Election Assistance Commission 2020 Election Administration and Voting Survey (EAVS). Specifically, we rely on the first question on the survey, which asks election officials to “report the total number of people (not votes or ballots) who were registered and eligible to vote in the November 2020 general election.” The EAVS reports results at the sub-state electoral jurisdiction level, generally the county or equivalent sub-state unit responsible for voter registration processing. We sum the values for subpart A of Question A1 across jurisdictions within a state as our baseline indicator of the state-reported number of voting-eligible registrants for the November 2020 election.

In Figure 1 we compare the EAVS estimates to the raw voter registration list counts. The circles and corresponding percentages indicate the raw file total count divided by the EAVS estimate. In 19 of the 21 states we examine, the raw, unprocessed file count is within 90% of the EAVS report. In DC (84%) and UT (78%), the percentages are substantially lower. One explanation for the lower figure in Utah is the ease by which a registrant can opt out of having their registration record provided in the public voter file.¹ The broad pattern, though, shows that even with no processing whatsoever a simple count of records in the raw files corresponds closely to election administrators’ stated understanding of how many persons are registered and eligible to vote.

We also compare registration counts to the 2020 Current Population Survey Voting and Registration Supplement (CPS). As noted by several previous researchers, the CPS’s method of querying

¹see, e.g., <https://slco.org/clerk/elections/voter-registration/voter-privacy-form/>

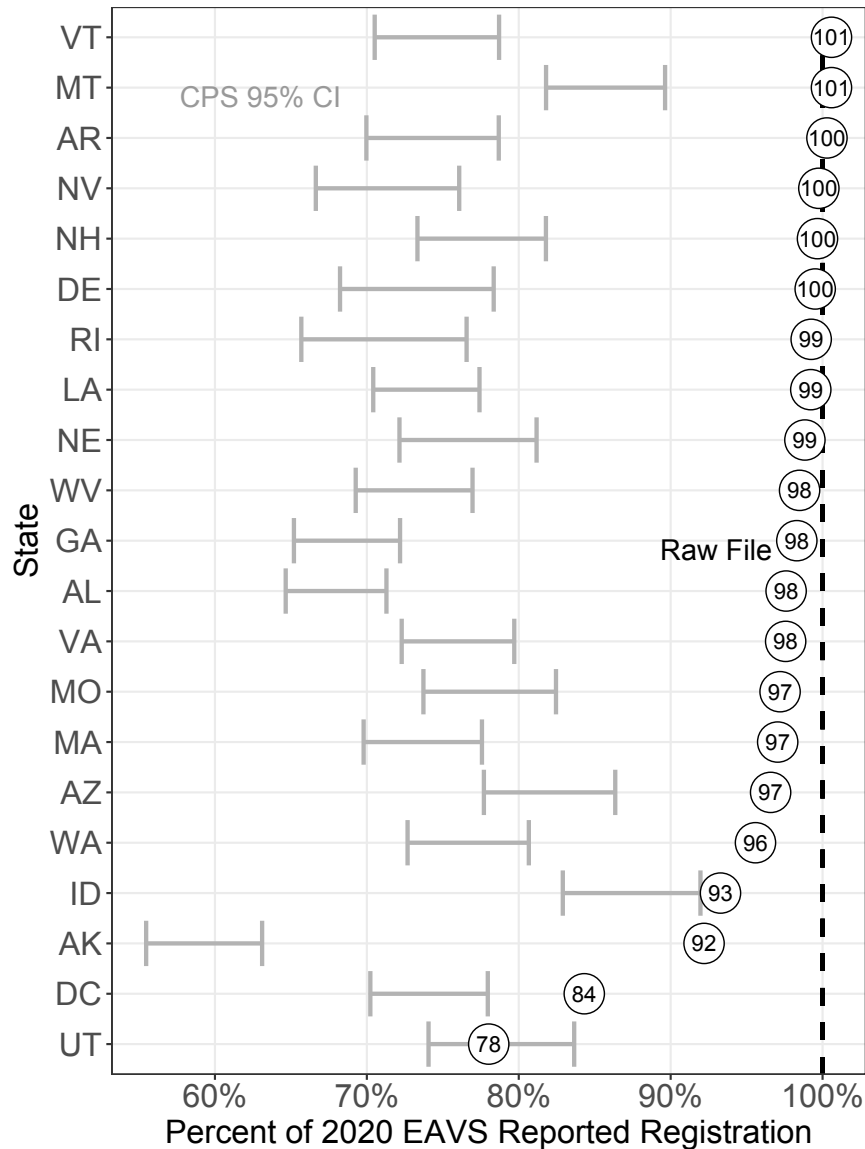


Figure 1: **Comparison of Raw Voter Registration File Count to EAVS Reports, 2020.** Baseline is the sum of county-level EAVS reports for the total number of registered voters (Q.A1a) in the state, represented by a dashed line. Circles and numbers indicate the number of lines in the raw statewide voter registration file closest to the November 2020 election, as a percentage of the EAVS total. Gray error bars indicate 95% confidence interval for Current Population Survey estimate of the number of registered voters (PES1 = 1 or PES2 = 1) as a percentage of the EAVS total.

registration is the reverse of most political science surveys (e.g., the ANES); respondents are only asked if they are registered to vote *after* they state that they did not vote in the election (Hur and Achen 2013; Pettigrew and Stewart 2018). We provide CPS estimates of registration totals by state via the gray 95% confidence intervals for these estimates in Figure 1 (Ansolabehere, Fraga

and Schaffner 2022). In nearly all cases the CPS does a far worse job of estimating the total number of registrants in a state, off by an average of 30%. This is all the more notable given the well-documented tendency of surveys, including the CPS, to over-report voter turnout.

References

- Ansolabehere, Stephen, Bernard L. Fraga and Brian F. Schaffner. 2022. “The Current Population Survey Voting and Registration Supplement Overstates Minority Turnout.” *Journal of Politics* 84(3):1850–1855.
- Ansolabehere, Stephen and Eitan Hersh. 2014. “Voter Registration: The Process and Quality of Lists”. In *The Measure of American Elections*, ed. Barry C. Burden and Charles Stewart, III. Cambridge: Cambridge University Press.
- Hersh, Eitan. 2015. *Hacking the Electorate: How Campaigns Perceive Voters*. Cambridge: Cambridge University Press. Chapter 6.
- Hur, Aram and Christopher H. Achen. 2013. “Coding Voter Turnout Responses in the Current Population Survey.” *Public Opinion Quarterly* 77(4):985–993.
- Pettigrew, Stephen and Charles Stewart, III. 2018. “Moved out, Moved On: Assessing the Effectiveness of Voter Registration List Maintenance.” Massachusetts Institute of Technology Political Science Department Research paper No. 2018-1.