

Measuring youth and college student voter turnout

Bernard Fraga ^a, John Holbein ^{b,*}

^a Indiana University, Woodburn Hall 210, Bloomington, IN, 47405, United States

^b University of Virginia, 111 Garrett Hall, Charlottesville, VA, 22901, United States

ARTICLE INFO

Keywords:

Youth voter turnout
College student voter turnout
Nationwide voter files
Turnout measurement

ABSTRACT

Young adulthood is a critical period in civic development. However, measuring electoral participation among this group generally—and the many young people who go to college in particular—is fraught with potential pitfalls stemming from a reliance on survey-based measures of voting. In this note, we compare patterns of youth turnout in two large-scale, survey-based datasets commonly used to measure voting, the Current Population Survey and the Cooperative Congressional Election Study, to two voter-file based datasets: the National Study of Learning, Voting, and Engagement (NSLVE) and a comprehensive nationwide voterfile provided by the Data Trust. We find high levels of concordance between measures in the NSLVE, Data Trust, and the CPS. However, despite linking their sample to validated voter records, the CCES does not mirror these benchmarks. We conclude by discussing the challenges and opportunities that shape the study of youth turnout.

Education and age have long been viewed as key predictors of political participation (Verba and Nie, 1972; Wolfinger and Rosenstone, 1980; Verba et al., 1995; Plutzer, 2002; Leighley and Nagler, 2013). With few exceptions (e.g. Kam and Palmer, 2008), young people vote less often than older people, and levels of turnout increase with educational attainment. Previous research has shown that youth who establish voting habits early on are much more likely to vote in the future (Plutzer, 2002; Gerber et al., 2003; Coppock and Green, 2016; Meredith et al., 2009), and that the transition from high school to a college environment may mitigate the negative effect of youth on turnout (Strate et al., 1989: 454; Highton and Wolfinger, 2001: 207). However, several challenges confront researchers who seek to study youth political participation. Young people are highly mobile (Ansobehere et al., 2012), have high non-contact rates in surveys (Abraham et al., 2006), and nearly 70% of high school students transition to colleges and universities after graduation. Yet, most previous studies of youth and college student turnout measure voting with survey self-reports, where misreporting of voter turnout forms yet another barrier (Ansobehere and Hersh, 2012). To what extent do commonly used data sources accurately measure participation for these vital sub-populations?

In this research note we leverage linked administrative records to assess survey-based measures of youth and college student voter turnout. As we discuss in more detail below, such a task is necessary as there is a tradeoff between covariate breadth and accuracy when

choosing between survey and administrative measures of turnout. We compare state-level patterns of turnout for 18–24 year olds in the Current Population Survey (CPS) and the Cooperative Congressional Election Study (CCES)—two widely-used political surveys—to individual-level, comprehensive, nationwide voter file data from the vendor The Data Trust and college student voter records provided by the National Study of Learning, Voting, and Engagement (NSLVE), an individual-level dataset covering approximately 10 million students nationwide.

We find high levels of concordance between youth turnout as measured in the NSLVE, The Data Trust, and the CPS. The CPS's strong relationship with these two sources of validated voter turnout is notable given that the CPS relies on *self-reported* turnout. We also show that despite linking their sample to validated voter records, the CCES does not mirror any of these benchmarks for state-level youth or college student voter turnout. We attribute this difference to sample size limitations in the CCES. Our work thus provides insights into the various datasets that scholars might use to study the participation of young people, and the tradeoffs that must currently be made when linking participation to political opinions for youth voters.

1. Approaches to studying youth turnout

For decades, the dominant approach to studying the turnout patterns of young voters in the United States has been to examine survey data. In particular, the Current Population Survey (CPS) forms the bedrock of

* Corresponding author.

E-mail addresses: bfraga@indiana.edu (B. Fraga), jh5ak@virginia.edu (J. Holbein).

our understanding of who votes and what differentiates young voters from other populations (Wolfinger and Rosenstone, 1980; Rosenstone and Hansen, 1993; Highton and Wolfinger, 2001; Richman and Pate, 2010; Leighley and Nagler, 2013).¹ This approach has some distinct advantages: for example, allowing us to isolate our estimates to sub-populations of interest (e.g. current or former college students), and permitting nationwide analyses with a relatively rich set of demographic covariates.

While surveys are a vital tool in the study of political behavior, they also come with distinct drawbacks. The most commonly cited issue is that turnout is measured through self-reports (Traugott and Katosh, 1979; Cassel, 2004; Hur and Achen, 2013). Over-reporting of voter turnout shapes both estimates of both how many individuals participate (Burden, 2000; McDonald, 2007), and what factors predict participation (Silver et al., 1986; Bernstein et al., 2001; Ansolabehere and Hersh, 2012, but see Berent et al., 2016). With this in mind, some attempts have been made to match survey respondents to their voter registration records; the Cooperative Congressional Election Studies (CCES) is the most prominent source of vote validated survey data, and measures a large set of political attitudes and behaviors. However, even when high-quality surveys like the CCES are matched to voter records, issues remain. The design of the surveys themselves and their comparatively small sample sizes constrain the use of these samples to draw inferences about sub-state-level differences or subgroups of interest (Burden, 2018; Grimmer et al., 2018).

Instead, scholars have turned to examining the voter files themselves. In recent years the use of voter files for behavior research has proliferated (Sigelman et al., 1985; Gimpel et al., 2004; Dyck and Gimpel, 2005; Green and Gerber, 2005; McDonald, 2007; Cooper et al., 2009; Ansolabehere and Hersh, 2012; Hersh, 2015; Fraga 2016a, b; Holbein and Hillygus 2016). Voter file data allows researchers to avoid using self-reported voting, and ameliorates the sample size/coverage issues that are key when studying sub-populations. Some analysts have used voter files to study patterns of youth voter turnout, examining mobility associated with transitioning from high school to colleges and universities (Ansobehere et al., 2012). Yet, voter files provide a limited amount of information about each registrant (Hersh, 2015), lacking many of the demographic factors that political scientists take for granted when analyzing surveys. For example, if one wants to study the important population of college student voters, raw voter files are of limited use as educational status is not included in registration records. That said, administrative records do allow us to measure rates of validated voter turnout among young people; birth year is almost always included in voter records.

2. Data

Our goal in this note is straightforward: benchmark state-level measures of youth and college student voting from commonly-used survey data to that found in administrative records. To accomplish this task, we draw on a unique combination of four primary data sources: survey data from the Current Population Survey (CPS), voter file validated survey data from the Cooperative Congressional Election Study (CCES), nationwide voter file data from The Data Trust, and college student voter records from the National Study of Learning, Voting, and Engagement (NSLVE).

We focus on validating state-level turnout across these datasets for three reasons. First, media organizations and nonprofits often quantify turnout rates below the national level when describing patterns of youth

voting.² Second, analysts rarely examine turnout with surveys at the sub-state level, as even the largest-scale surveys of political behavior (i.e. the CCES and CPS) are not designed to be representative at finer levels of geography (Cooper et al., 2009). Hence, benchmarking at the district or county level is not justifiable.³ Finally, many scholars use state-level variation to estimate the effect of electoral rules and other complementary policies on voter turnout (Keele and Minozzi, 2013), including impacts on youth in particular.⁴ If our estimates of state turnout are off, our conclusions regarding the effect of these policies may be spurious as well. Altogether, validating state-level turnout rates is thus an important first step for those interested not only in understanding how many young Americans vote, but also for research that examines individual-level factors that impact the turnout calculus for youth and college students.

The Current Population Survey November Voting and Registration Supplement serves as the primary source of survey estimates of voter turnout for young Americans. A national survey of over 130,000 individuals, the CPS is generally seen as providing the best estimates of voter turnout for small demographic groups (Wolfinger and Rosenstone, 1980; Highton, 2005; Rocha et al., 2010; McDonald, 2017) aside from voter file data (Fraga 2018). The CPS is a household-level survey, but queries the respondent regarding demographic and participatory variables for all members of a household. This includes questions regarding age and whether or not the individual is a part-time or full-time college student.⁵ However, the CPS relies on self-reported voting, and does not validate turnout by comparing stated behavior with official records. In this paper, we draw on CPS data from 2012, 2014, and 2016 in our analyses.

We also examine the CCES, a matched random sample survey that generates a nationally representative set of respondents from those participating in an online panel.⁶ While having a smaller sample size than the CPS ($N \approx 55,000$), the CCES is often favored over the CPS for research on voter behavior as it validates turnout measures with voter file data⁷ and provides information about vote choice, political attitudes, and a host of other variables of interest to political scientists. We use CCES data from 2012, 2014, and 2016.

We compare these surveys with two unique administrative data-based sources. The first is a nationwide voter file maintained by The Data Trust. The Data Trust is a private vendor that compiles discrete state voter files, combines these records into a national, individual-level database, and vends the resultant information to political campaigns.⁸ The Data Trust collects and cleans data on all registered voters in the

² e.g., CIRCLE (2019), Madhani (2018).

³ In benchmarking turnout measures across survey/administrative data sources, some aggregation *has* to occur in order to establish a common unit of analysis which each of these datasets is representative. Validating at the county or district level is not justifiable given that the CCES and CPS are not designed to be representative below the state level (Cooper et al., 2009). If we were, however, to still try and measure youth turnout at the county level despite the surveys not being designed to be used in this way, we would probably see a greater discrepancy between surveys and voter files given very small sample sizes at this level.

⁴ e.g., Wolfinger and Rosenstone (1980); Richman and Pate (2010); Holbein and Hillygus (2016).

⁵ During the period we study, student dormitories were included in the CPS sampling frame. However, as the CPS queries “usual residents” at sample addresses, the Census Bureau asserts that students living in dorms are more likely to be listed as “usual residents” in their pre-college place of residence (BLS, 2019; U.S. Census Bureau, 2018).

⁶ <http://cces.gov.harvard.edu>.

⁷ Through matching panelists with the Catalist, LLC database: the same voter file vendor used for the NSLVE. In our study, we follow guidelines provided by CCES researchers and consider individuals who could not be matched to voter files to be non-voters, regardless of their self-reported turnout.

⁸ Given that we rely solely on turnout, age, and geographic data, the data is unlikely to be biased due to the preferred clients of The Data Trust. Furthermore, no information was shared between the Data Trust and NSLVE.

¹ The American National Election Study (ANES) is also common in political science research on voting behavior (e.g., Strate et al., 1989). However, this survey is not well suited to studying youth turnout in that its (comparatively) small sample size restricts our ability to draw meaningful inferences.

United States, as well as many other adults. Our file has data through the 2016 election (in many, but not all, states) and was compiled in September 2017.⁹ We draw on The Data Trust's turnout indicators for in 2012, 2014, and 2016. Unlike other academic subscriptions to voter file vendor data, ours is not a 1% sample; it is the entire individual-level data file ($N \approx 200$ million). With this dataset, we use contemporaneous Census estimates of the voting-age population in each state as our denominator, subsetting to young people as described in more detail below.¹⁰ While this dataset does not have a measure of whether young people are students or not, it does provide us with a broader scope and a larger sample size with official turnout records for all 18–24 year olds in the U.S.

Our second administrative data source, the NSLVE, consists of college student enrollment data matched to voter registration records.¹¹ The National Study of Learning, Voting, and Engagement (NSLVE) is produced by the Institute for Democracy & Higher Education (IDHE), a research center housed in the Tisch College of Civic Life at Tufts University. The IDHE entered into a partnership with the National Student Clearinghouse (NSC) to acquire student enrollment data and match this information with voter registration records. The NSC is the largest, and most institutionally diverse, database of student enrollment, covering thousands of public, non-profit private, and for-profit private institutions. Through the NSLVE partnership, the NSC matches student records to a voter file database maintained by Catalist, LLC.¹² Participating institutions receive information about student voting participation in a given election in the form of a report, with aggregate statistics also provided on the NSLVE website. Just over half of postsecondary students nationwide are in one of the 1020 NSLVE-participating institutions, making the roughly 10 million individuals in the NSLVE the largest national database of student voting participation.¹³ Here we use individual-level NSLVE data corresponding to the 2012, 2014, and 2016 elections.¹⁴ In conducting the state-level analyses that we provide in this note, we use the “home” state of the student, registrant, or respondent in all analyses.¹⁵

Given our interest in understanding youth turnout, we subset to individuals who were 18–24 on Election Day in each of our datasets.¹⁶ We do this for three reasons. First, disparities in voter registration and turnout are most severe for individuals who are in their teens and early twenties, with 18–21 year olds having particularly low rates of voter registration according to voter file data (Ansolabehere et al., 2012).

⁹ In the [Online Appendix](#) we compare this snapshot with archival data from 2012, 2014, and 2016, and find our results are not different when using historical snapshots.

¹⁰ While it is advisable to remove non-citizens from estimates of voter turnout (McDonald and Popkin, 2001), the enrollment data used in the NSLVE does not consistently denote non-citizen students. Thus in order to provide a consistent measure of turnout across the datasets, we do not remove non-citizens and estimate the voting-age population denominator for the Data Trust analysis using U.S. Census Bureau Population Estimates Program (PEP) figures interpolated to November of each election year.

¹¹ <http://idhe.tufts.edu/nslve>.

¹² <https://idhe.tufts.edu/nslve/frequently-asked-questions>.

¹³ In fall 2017, there were approximately 18.5 million individuals enrolled in degree-granting postsecondary institutions in the United States, 16 million of which were pursuing undergraduate education <https://nscresearchcenter.org/current-term-enrollment-estimates-fall-2017/>.

¹⁴ In the [Online Appendix](#), we indicate that the coverage of the NSLVE dataset is independent of broader patterns of voter turnout in the state. We also show that rates of student participation in a given state-year are strongly related to overall rates of youth turnout in that state-year.

¹⁵ For the NSLVE, this was ascertained by the listed home ZIP code in the enrollment record, the state the student is registered to vote in if home ZIP is unavailable, and finally, the state of the postsecondary institution if neither of these sources is provided.

¹⁶ We test the robustness of our results to this age range in the [Online Appendix](#).

Second, the inclusion of 22–24 year olds covers differences in age-of-entry for college students, some of which may be attributable to variation in the birthdate required for kindergarten enrollment, variation in the timeline for completion of undergraduate degree requirements, and other factors associated with states or undergraduate institutions that are non-random. Finally, such an age range seeks to limit the extent to which a person's decision to enroll in a college or university, conditional on age, is also related to the decision to vote.¹⁷ The decision to enroll in college is itself non-random; despite the growth of the college-educated population in the United States, college students still come from disproportionately wealthy, highly educated, and politically engaged families. Our analysis seeks to examine a population that is roughly comparable across states and surveys, and can inform our understanding of patterns of undergraduate student electoral population. As noted above, students are a subgroup of particular interest in the broader youth voter population, and the NSLVE and CPS data allow us to identify this group directly. However, the CCES only has an indirect measure of whether an individual is a college student: combining information about educational attainment (high school graduates), age (18–24 years old), and employment status (“student”), we infer which respondents are likely to be enrolled in postsecondary institutions.

3. Results

In any benchmarking exercise it is useful to establish a working theory of which dataset is likely to be closest to the “truth.” Ex-ante there are strong reasons to suspect that The Data Trust's nationwide voter file best captures youth turnout. Possessing validated voting rates across all states with high rates of coverage/accuracy in birth dates, The Data Trust data we employ relies on no assumptions about sampling or survey-response (present in both the CPS and CCES), nor self-reported voting (present in the CPS). In the [Online Appendix](#) we demonstrate that overall (i.e. all voting age adults) state-level estimates of turnout in the Data Trust align closely with the state-level turnout figures estimated by McDonald (2017), especially in recent years ($r \approx 0.99$). For these reasons, we argue for the use of The Data Trust data as the most useful benchmark. The question then becomes: to what extent do the CPS and CCES accurately align with nationwide measures of validated youth turnout?

3.1. Benchmarking youth voter turnout

We start by giving a sense of overall patterns of youth turnout in the United States. Fig. 1 maps youth turnout rates in the 2012, 2014, and 2016 elections across states. This helps to establish a baseline of where youth turnout tends to be low, moderate, and high. Though discussing these patterns might seem somewhat elementary, doing so is important given that comprehensive nationwide voter files are just becoming available to researchers. Again, most of what we know about levels of youth turnout comes from the CPS, which instead relies on turnout as measured through self-reports.

In 2012, Minnesota (44.1%), Maine (43.4%), Iowa (43.2%), Ohio (41.9%), and Virginia (40.1%) led the nation in youth voter turnout.

¹⁷ Individuals enrolled in undergraduate coursework who are older than 24 may have delayed college entry due to family or work obligations, both of which may be related to the decision to vote. On the other hand, individuals continuing schooling beyond the undergraduate level may have greater socioeconomic resources than those who are unable to do so. Notably, the CPS and CCES do not ask respondents to indicate whether they are enrolled in undergraduate or graduate coursework. We encourage future work to examine age differences and patterns that may manifest for “non-traditional” student populations.

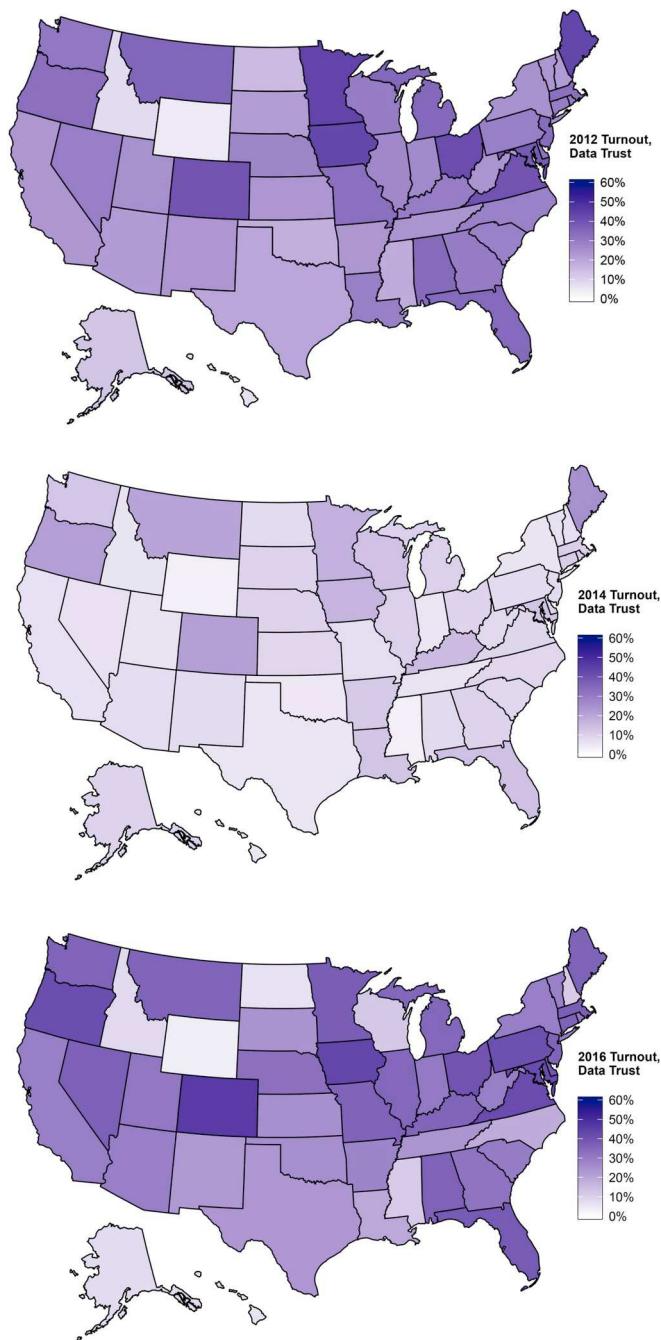


Fig. 1. Voter Turnout Among Youth, 2012–2016. Note: Maps present state-level voter turnout rate among youth (18–24). Numerator is the number of registration records in The Data Trust flagged as voters in general elections held in November 2012, 2014, and 2016. Denominator is the Census Population Estimates Program estimate of the 18–24 year old population in each state in November of the election year.

However, states like Wyoming (4.7%), Hawaii (7.0%), and Idaho (8.0%) all lagged well behind their counterparts.¹⁸ Similar patterns appear in 2016. In that year, Colorado (46.4%), Iowa (43.4%), and Virginia (42.2%) led the nation in youth turnout, while again Wyoming (4%), Hawaii (6.3%), and North Dakota (6.5%) made up the bottom of the list.¹⁹ In these latter states, the vast majority of young people do not vote in presidential elections.

As expected, youth turnout in 2014 was lower than in the two presidential elections in our sample. States that had especially high levels of youth turnout in 2014 included Maine (25.1%), Oregon (21.8%), Colorado (21.5%), Montana (20.5%), and Minnesota (17.7%). In contrast, states like Wyoming (3.5%), Mississippi (3.6%), Hawaii (5.0%), D.C. (5.1%), and Oklahoma (5.4%) all struggled to get even a small fraction of their young people to turn out to vote.²⁰ States that had an especially acute drop-off in youth turnout from the 2012 to the 2014 elections include Ohio (32.6 percentage point decline), Virginia (30.9 point decline), Iowa (26.5 point decline), Minnesota (26.4 point decline), and Alabama and New Jersey (26.1 point decline). While these patterns are likely due to myriad factors—including, but not limited to, electoral competition, cohort dynamics, and electoral rules—they indicate where validated voter turnout among young people is high, low, and especially variable across elections.

Fig. 2 compares CPS and CCES youth turnout rates to those in The Data Trust.²¹ Points above the reference line indicate states with higher turnout in the CPS/CCES than in The Data Trust, whereas points below the line indicate that turnout was higher in The Data Trust. Among the two surveys, the CPS benchmarks considerably better to The Data Trust's voter turnout data than the CCES. In virtually all cases, the CPS turnout numbers are above those from The Data Trust, as we would expect based on self-reported voting.²² However, the relationship between the measures is strong ($r = 0.7$). In contrast, there is a considerably weaker relationship between the CCES's matched validated voter turnout estimates and The Data Trust figures. While there is still a moderate correlation ($r = 0.37$), the CCES has substantially more dispersion from the reference line. This is true *despite* the CCES containing validated measures of individual-level voting.

Why do the CCES's youth turnout numbers appear to be so far from voter-file based measures of youth turnout? Taking a closer look at **Fig. 2**, we see several states that have 0% youth turnout in the CCES. These implausible estimates suggest that the CCES might suffer from small samples in particular states, suggesting that there are too few

¹⁸ As we discuss further in the [Appendix](#), one challenge in working with voter files is that age is missing from more than 1% of voters in a state voter file in 10 states (NC, WI, NH, MS, DC, ID, HI, ND, AK, and WY). This results in an underestimate of turnout at the youth turnout given that we can only count the number of votes among individuals with birthdates. If we down-weight our VAP denominator by the fraction of missing birthdates in the file, turnout is still lowest in Wyoming (7.7%), Hawaii (9.2%), and Idaho (10.5%). Such an adjustment does not change the rank-ordering of our youth turnout estimates a great degree; Indeed, the correlation between adjusted and unadjusted turnout is high (2012: 0.98, 2014: 0.98, 2016: 0.999).

¹⁹ The turnout numbers for these states adjusting for missing date of birth are Wyoming (6.6%), Hawaii (8.3%), and North Dakota (8.9%).

²⁰ The turnout numbers for these states adjusting for missing date of birth are Wyoming (5.7%), Mississippi (4.5%), Hawaii (6.6%), D.C. (6.7%), and Oklahoma (5.4%).

²¹ Online Appendix [Figure A2](#) describes the relationship between year of age and turnout across the three datasets. There we see that the deviation between the Data Trust's estimates of turnout and the CPS or CCES estimates of turnout is especially large for young voters.

²² Here we use the CPS's convention of counting non-respondents to the Voting and Registration Supplement as non-voters. However, as noted by [Hur and Achen \(2013\)](#) and [McDonald \(2017\)](#), survey researchers usually treat non-respondents as missing data. In the [Online Appendix](#), we reproduce our main text analyses with non-respondents removed from the analyses of the CPS, and find the same patterns.

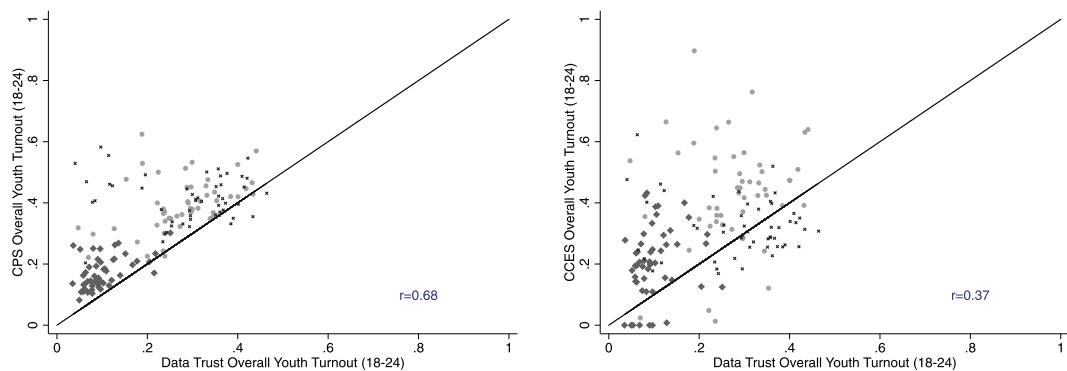


Fig. 2. Overall Youth Turnout From the CPS/CCES Benchmarked to The Data Trust. Note: Points indicate a single state-year observation. The lightest points indicate turnout estimates for 2012, while the darkest points indicate estimates for 2016. Circles also indicate 2012, Diamonds 2014, and X-marks 2016.

18–24 year olds in the CCES to allow for accurate state-level measurement of voter turnout. Thus, a problem with the sample composition negates some of the value of having validated turnout merged into the CCES survey. Analyses in the [Online Appendix](#) further corroborate this story, as deviations from The Data Trust estimates are most severe in states with small sample sizes.

3.2. Benchmarking college student voter turnout

Next we explore an important subgroup of young people: those enrolled in college. As with The Data Trust, we begin by describing patterns of turnout in the NSLVE data before benchmarking it to other samples. [Fig. 3](#) maps voter file-derived student turnout from the NSLVE by state over the past three elections.²³ As can be seen, there is a significant amount of variation across states within a given year and within individual states over time. One example of a state that has had consistently high student voter turnout over time is Colorado, which had a turnout rate of 58.5% (first place among states) in 2016, 31.9% (second) in 2014, and 56.2% in 2012 (first). The District of Columbia, perhaps due to its noncompetitive elections, lack of statehood, and many out-of-District students, consistently ranks at the bottom of youth student voting: coming in last in 2012, 2014, and 2016. Between these two bookends, however, there is a significant amount of state by state variation even when examining just these three elections. Virginia (34 position fall in the rank ordering) and Alabama (27 position fall) had especially steep relative declines in student voter turnout from 2012 to 2014, while Alaska (32 position rise) and Hawaii (32 position rise) had especially large increases.

Some of these changes can be attributed to the differences in races on the ballot over time. However, even if we look within presidential elections, the changes are stark. North Carolina (23 spot fall), Arkansas (20 position fall), and Mississippi (18 position fall)—all southern states—saw a marked relative decline in student voting rates from 2012 to 2016. Pennsylvania (18 position rise), Montana (17 position rise), and Connecticut (15 position rise) all saw meaningful increases over these two election cycles.

We now turn to benchmarking the NSLVE measure of college student voter turnout to survey-based estimates. Given that nationwide voter file databases do not have information about student status, the matched student data in the NSLVE provides crucial information about this subgroup of youth voters. [Fig. 4](#) shows the relationship between state-level estimates of college student turnout from the CPS (y-axis) and the NSLVE (x-axis). Along with the corresponding state-level points, in the

figure is a reference line where turnout in the NSLVE and CPS/CCES would be if there were no difference between the measures.

As with youth voting overall, turnout among college students in the CPS is generally higher than in the NSLVE. This is true at the national level: In the CPS, college student turnout in 2012 clocked in at 45.0%, whereas in the NSLVE the number was 41.3%. In 2014, the CPS had college student turnout at 17.5%, while in NSLVE it was 13.1%. In 2016, the CPS (44.3%) and NSLVE (44.7%) were very closely aligned. While the national turnout numbers came closest in 2016, this came at the expense of of larger error across individual states. The correlation between measures across states is relatively strong in 2012 and 2014, but there is a significant dip in 2016. Overall, however, the CPS and NSLVE seem to be picking up on similar state-level variation, with the correlation when pooling all years exceeding $r = 0.85$.²⁴

We engage in the same exercise with the CCES data in [Fig. 5](#). Again, the relationship is weak between the NSLVE's validated voting rate among college students ages 18–24 and the CCES's validated voter turnout rate among current students who have completed high school and are 18–24. This can be seen in the large degree of stochastic dispersion from the reference line. As discussed in the [Online Appendix](#), small sample sizes again appear to be an important factor producing this effect.

3.3. Dealing with turnout discrepancies

To what degree do turnout discrepancies across various datasets influence estimation of effects of interest? What should researchers do to compensate for these differences? While strategies are likely specific to the research question, we can make some brief methodological recommendations that should apply generally to those studying youth turnout at the sub-national level.

Our first recommendation is that scholars exercise caution when using surveys to examine youth turnout. Even if a.) the survey is designed to be representative at the state level (among all adults), b.) the survey is large, and c.) the survey is matched to voter files, this does not mean that turnout for the sub-sample of young people will mirror true rates of youth turnout at the state level. Deciding which data to use comes with inherent tradeoffs, as the administrative records-based measures provided by nationwide voter files also results in a small number of available covariates. However, scholars should be cognizant of survey data's inherent limitations in the study of youth turnout.

Second, researchers studying youth turnout would do well to benchmark state and national turnout estimates to official or semi-official estimates such as those provided by [McDonald \(2017\)](#). Doing so is not common practice when analyzing survey data at present.

²³ [Fig. 3](#) does not display data for North Dakota, New Mexico, and Delaware to avoid identifying the turnout rate of specific NSLVE-participating institutions. These states are shaded in gray in the map. Including them does little to change our results below.

²⁴ See [Figure A1](#) for maps that indicate the deviation between NSLVE and CPS measures of student turnout by year.

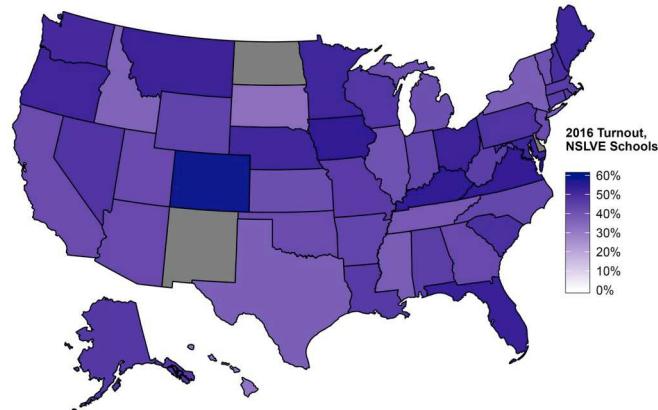
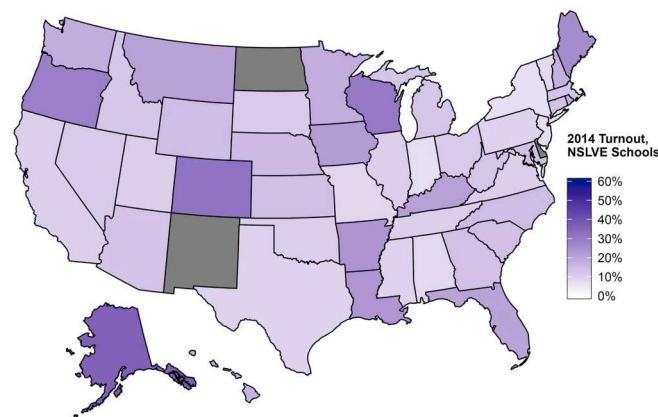
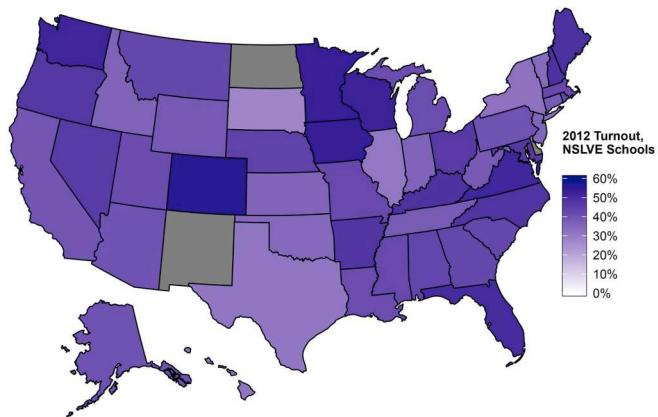


Fig. 3. Voter Turnout Among College Students, 2012-2016. Note: Maps present state-level voter turnout rate among 18-24 year old college students in the NSLVE data for November 2012, 2014, and 2016. Figure does not display data for North Dakota, New Mexico, and Delaware to avoid identifying the turnout rate of specific NSLVE-participating institutions.

However, researchers may be eliding important limitations on the external validity of their observational research with the misguided assumption that “nationally representative” surveys allow for accurate subgroup analyses. Hence, we recommend that this become a standard check in voter turnout studies, especially studies using samples that have not been benchmarked in the past.

Third, scholars should explore how the propensity to mismeasure a state’s turnout varies with the treatment of interest. If a researcher is concerned that mismeasurement is influencing her estimates, one option is to include the state-level deviation from an official estimate as a control in statistical models. This deviation measure would vary over

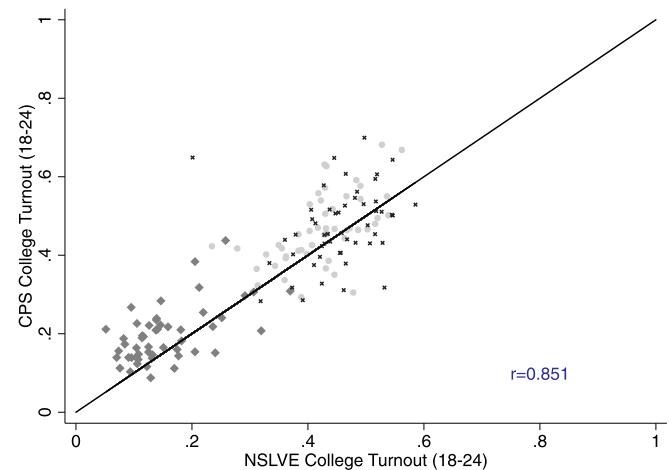


Fig. 4. College Student Turnout in the NSLVE Benchmarked to the CPS. Note: Points indicate a single state-year observation. The lightest points indicate estimates for 2012, while the darkest points indicate estimates for 2016. Circles also indicate 2012, Diamonds 2014, and X-marks 2016.

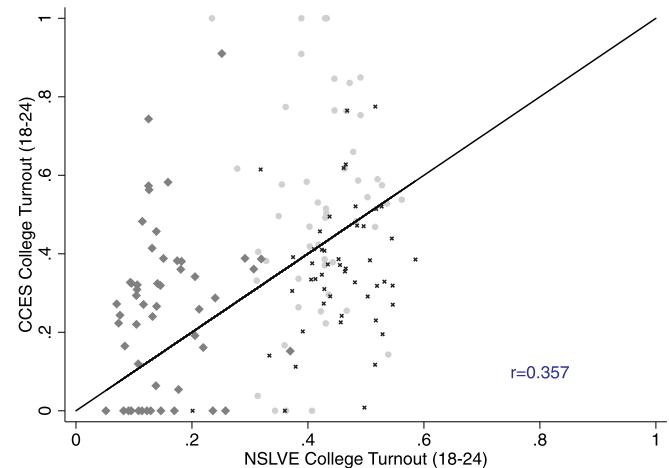


Fig. 5. College Student Turnout in the NSLVE Benchmarked to the CCES. Note: Points indicate a single state-year observation. The lightest points indicate estimates for 2012, while the darkest points indicate estimates for 2016. Circles also indicate 2012, Diamonds 2014, and X-marks 2016.

time and geography and, as such, could be included in e.g., the canonical two-way fixed effects model with state and year fixed effects. This control variable strategy could be expanded to include benchmarked differences between subgroups if measurement of turnout for these subgroups is of interest to the individual researcher. Even if mismeasurement does not vary by the treatment of interest, scholars may want to consider strategies for limiting the statistical noise that comes from introducing measurement error into the dependent variable.

4. Conclusion

In this research note, we explored the properties of four data-sets—including two newly available administrative records-based sources—that measure youth voting. We find high levels of concordance between state-level measures of youth turnout in the NSLVE, The Data Trust, and the CPS, despite the fact that the CPS relies on self-reported turnout. While linking their sample to validated voter records, the CCES does not mirror any of these benchmarks for youth voter turnout. Thus, a trade-off emerges when using the rich set of political covariates available in the CCES to study correlates of voting behavior,

as CCES turnout measures for youth and college students are not well aligned with state-level administrative records or the CPS.

Our note has implications beyond the four datasets we analyze here. In the study of youth turnout in particular, some scholars have turned to specialty surveys which, like the CCES, provide a set of covariates in addition to participation measures (e.g., Niemi and Hanmer, 2010). We would encourage analysts of observational data to verify the broad turnout patterns they find with voter file-based sources, but also to develop new ways of measuring covariates of interest that do not lead to small sample concerns. An example of this is a resource like the NSLVE: student status cannot be ascertained with raw voter file data alone, but the linking of registration records to student administrative data has enhanced experimental work (e.g., Bennion and Nickerson, 2011) and may provide insights into broader patterns of turnout as well.

We outline other methodological checks in the [Online Appendix](#) that researchers may wish to keep in mind when working with these datasets in the future, but leave the broader theorizing and analysis of additional correlates to future work. Here we have focused on learning which datasets best measure turnout among the vital subpopulation of young people. Much remains to be done to understand what helps set young people on a path to becoming active participants in democracy; however, in this task accurate measurement cannot be ignored.

Acknowledgement

The funder included the National Science Foundation (SES-1657821).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.electstud.2019.102086>.

References

- Abraham, Katharine G., Maitland, Aaron, Bianchi, Suzanne M., 2006. Nonresponse in the American time use survey: who is missing from the data and how much does it matter? *Public Opin. Q.* 70 (5), 676–703.
- Ansolabehere, Stephen, Hersh, Eitan, 2012. Validation: what big data reveal about survey misreporting and the real electorate. *Political Anal.* 20 (4), 437–459.
- Ansolabehere, Stephen, Hersh, Eitan, Shepsle, Kenneth, 2012. “Movers, stayers, and registration: why age is correlated with registration in the U.S.” *quarterly. J. Political Sci.* 7 (4), 333–363.
- Bennion, Elizabeth A., Nickerson, David W., 2011. The cost of convenience: an experiment showing E-mail outreach decreases voter registration. *Political Res. Q.* 64 (4), 858–869.
- Berent, Matthew K., Krosnick, Jon A., Arthur, Lupia, 2016. Measuring voter registration and turnout in surveys: do official government records yield more accurate assessments? *Public Opin. Q.* 80 (3), 597–621.
- Bernstein, Robert, Chadha, Anita, Montjoy, Robert, 2001. Overreporting voting: why it happens and why it matters. *Public Opin. Q.* 65 (1), 22–44.
- BLS, Bureau of Labor Statistics, 2019. Technical Documentation: Handbook of Methods. United States Department of Labor. <https://www.bls.gov/opub/hom/cps/pdf/cps.pdf>. (Accessed 20 May 2019).
- Burden, Barry C., 2000. Voter turnout and the national election studies. *Political Anal.* 8 (4), 389–398.
- Burden, Barry C., 2018. Disagreement over ID requirements and minority voter turnout. *J. Politics* 80 (3).
- Cassel, Carol A., 2004. Voting records and validated voting studies. *Public Opin. Q.* 68, 102–108.
- CIRCLE, Center for Information and Research on Civic Learning and Engagement, 2019. 2018 Youth Turnout Increase in Every State for Which We Have Data. Available at: <https://civicyouth.org/2018-youth-turnout-increased-in-every-state-for-which-we-have-data/>.
- Cooper, Christopher A., Haspel, Moshe, Gibbs Knotts, H., 2009. The value of voterfiles for U.S. State politics research. *State Politics Policy Q.* 9 (1), 102–121.
- Coppock, Alexander, Green, Donald P., 2016. Is voting habit forming? New evidence from experiments and regression discontinuities. *Am. J. Pol. Sci.* 60 (4), 1044–1062.
- Dyck, Joshua J., Gimpel, James G., 2005. Distance, turnout, and the convenience of voting. *Soc. Sci. Q.* 86 (3), 531–548.
- Fraga, Bernard L., 2016. Candidates or districts? Reevaluating the role of race in voter turnout. *Am. J. Pol. Sci.* 60 (1), 97–122.
- Fraga, Bernard L., 2016. Redistricting and the causal impact of race on voter turnout. *J. Politics* 78 (1), 19–34.
- Fraga, Bernard L., 2018. The Turnout Gap: Race, Ethnicity, and Political Inequality in a Diversifying America. Cambridge University Press, New York.
- Gerber, Alan S., Green, Donald P., Shachar, Ron, 2003. Voting may be habit-forming: evidence from a randomized field experiment. *Am. J. Pol. Sci.* 47 (3), 540–550.
- Gimpel, James G., Dyck, Joshua J., Shaw, Daron R., 2004. Registrants, voters, and turnout variability across neighborhoods. *Political Behav.* 26 (4), 343–375.
- Green, Donald P., Gerber, Alan S., 2005. Preface: recent advances in the science of voter mobilization. *Ann. Am. Acad. Pol. Soc. Sci.* 601 (September), 6–9.
- Grimmer, Justin, Hersh, Eitan, Meredith, Marc, Mummo, Jonathan, Nall, Clayton, 2018. “Obstacles to estimating voter ID laws’ effect on turnout. *J. Politics* 80 (3), 000–000.
- Hersh, Eitan D., 2015. Hacking the Electorate: How Campaigns Perceive Voters. Cambridge University Press.
- Highton, Benjamin, 2005. Self-reported versus proxy-reported voter turnout in the current population survey. *Public Opin. Q.* 69 (1), 113–123.
- Highton, Benjamin, Wolfinger, Raymond E., 2001. The first seven years of the political life cycle. *Am. J. Pol. Sci.* 45 (1), 202–209.
- Holbein, John B., Sunshine Hillygus, D., 2016. Making young voters: the impact of preregistration on youth turnout. *Am. J. Pol. Sci.* 60 (2), 364–382.
- Hur, Aram, Achen, Christopher H., 2013. Coding voter turnout responses in the current population survey. *Public Opin. Q.* 77 (4), 985–993.
- Kam, Cindy D., Palmer, Carl L., 2008. Reconsidering the effects of education on political participation. *J. Politics* 70 (3), 612–631.
- Keele, Luke, Minozzi, William, 2013. How much is Minnesota like Wisconsin? Assumptions and counterfactuals in causal inference with observational data. *Political Anal.* 21 (2), 193–216.
- Leighley, Jan E., Nagler, Jonathan, 2013. Who votes now? Demographics. In: *Issues, Inequality, and Turnout in the United States*. Princeton University Press, Princeton, NJ.
- Madhani, Aamer, 2018. “Early and Absentee Voting: Youth Vote Surges in Hotly Contested races.” *USA Today*. November 5, 2018. Available at: <https://www.usatoday.com/story/news/politics/elections/2018/11/05/midterm-election-early-voting-absentee-ballot-youth-vote-republican-democrat/1892285002/>.
- McDonald, Michael P., 2007. The true electorate: a cross-validation of voter registration files and election survey demographics. *Public Opin. Q.* 71 (4), 588–602.
- McDonald, Michael P., 2017. “Voter Turnout Demographics.” *United States Elections Project*. <http://www.electproject.org/home/voter-turnout/demographics>. (Accessed 7 March 2017).
- McDonald, Michael P., Popkin, Samuel L., 2001. The myth of the vanishing voter. *Am. Pol. Sci. Rev.* 95 (4), 963–974.
- Meredith, Marc, et al., 2009. Persistence in political participation. *Quart. J. Political Sci.* 4 (3), 187–209.
- Niemi, Richard G., Hanmer, Michael J., 2010. Voter turnout among college students: new data and a rethinking of traditional theories. *Soc. Sci. Q.* 91 (2), 301–323.
- Plutzer, Eric, 2002. Becoming a habitual voter: inertia, resources, and growth in young adulthood. *Am. Pol. Sci. Rev.* 96 (1), 41–56.
- Richman, Jesse, Pate, Andrew, 2010. “Can the college vote turn out? evidence from the U.S. States, 2000–08. *State Politics Policy Q.* 10 (1), 51–68.
- Rocha, Rene R., Caroline, J., Tolbert, Daniel C. Bowen, Clark, Christopher J., 2010. Race and turnout: does descriptive representation in state legislatures increase minority voting? *Political Res. Q.* 63 (4), 890–907.
- Rosenstone, Steven J., Hansen, John Mark, 1993. *Mobilization, Participation, and Democracy in America*. Macmillan, New York.
- Sigelman, Lee, Roeder, Philip W., Jewell, Malcolm E., Baer, Michael A., 1985. Voting and nonvoting: a multi-election perspective. *Am. J. Pol. Sci.* 29 (4), 749–765.
- Silver, Brian D., Anderson, Barbara A., Abramson, Paul R., 1986. Who overreports voting? *Am. Pol. Sci. Rev.* 80 (2), 613–624.
- Strate, John M., Parrish, Charles J., Elder, Charles D., Ford, Coit, 1989. Life span civic development and voting participation. *Am. Pol. Sci. Rev.* 83 (2), 443–464.
- Traugott, Michael W., Katosh, John P., 1979. Response validity in surveys of voting behavior. *Public Opin. Q.* 43, 359–377.
- U.S. Census Bureau, 2018. Group quarters/residence rules. <https://www.census.gov/topics/income-poverty/poverty/guidance/group-quarters.html>. (Accessed 20 March 2018).
- Verba, Sidney, Schlozman, Kay L., Brady, Henry E., 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Harvard University Press, Cambridge, MA.
- Verba, Sidney, Nie, Norman H., 1972. *Participation in America*. Harper and Row, New York.
- Wolfinger, Raymond E., Rosenstone, Steven J., 1980. *Who Votes?* Yale University Press, New Haven, CT.