One Shift, Two Shifts, Red Shift, Blue Shift: Reported Election Returns in the 2020 Election

John A. Curiel
jcuriel@mit.edu

Charles Stewart III
cstewart@mit.edu

Jack Williams
jackrw@mit.edu

Massachusetts Institute of Technology
MIT Election Data and Science Lab

Draft of July 9, 2021

Abstract

Shifting vote margins in the hours and days following the close of polls was a highlighted issue in the 2020 election. The tendency of vote margins in recent presidential elections to trend in favor of the Democratic presidential candidate as the count proceeds has been previously studied and given the label "blue shift." We address how these shifts occurred during the 2020 election with contemporaneous National Election Pool (NEP) data reported by county. States were slower to report overall vote counts if they had large volumes of mail ballots, prohibited preprocessing mail ballots, and allowed mail ballots to arrive after Election Day. We also find that both between- and within-county differences drove partisan trends in the 2020 vote count. Counties Biden won took longer to complete their counts than counties Trump won. Within the average county, Biden's votes took longer to count completely than Trump's. Nonetheless, in the first couple of hours after polls closed, there was actually a disproportionate number of Biden votes reported, which we show were likely due to preprocessed mail ballots being reported first.

Prepared originally for presentation at the annual meeting of the Midwest Political Science Association, April 14 – 17, 2021. Updated for presentation at the summer meeting of the Election Science, Reform, and Administration Conference, July 19 – 21, 2021.

One Shift, Two Shifts, Red Shift, Blue Shift: Reported Election Returns in the 2020 Election

John A. Curiel
Charles Stewart III
Jack Williams

The 2020 presidential election proved to be among the most contentious in U.S. history. One of the most notable features was the pre-emptive delegitimization of the election results by incumbent President Donald Trump and his supporters in the months leading up to the election, followed by post-election charges that longitudinal patterns in the reported results indicated fraud. The most prominent of these so-called suspicious patterns was a "blue shift" on election night; late-arriving results were more favorable to the Democratic candidate Joe Biden than reported results soon after the polls were closed (Bronner, Matlin, and Tabb, 2020).

The presence of a blue shift is not new to American politics. Foley (2013) noted that increasingly after 2000, later-counted ballots tended to arise from Democratic-leaning sources. Because of the partisan divide that opened up over the use of mail ballots in 2020, the shift to voting by mail promised to exacerbate the blue-shift phenomenon.

Ahead of the election, election administrators and the academic community were aware of the possibility of a blue shift and the opportunities for public confusion, misinformation, and disinformation that could follow if the count had a slow, pronounced partisan time trend. Consequently, efforts were made to speed up the vote count, especially by allowing the preprocessing of mail ballots. Attempts were also made to educate the public that votes reported on election night (and the days immediately after that) are unofficial and that the non-random release of returns was because of unremarkable administrative practices.

Although accusations and retorts were hurled around in the intense days that followed the November election concerning the blue shift, several empirical questions remain about the pace of election-return reporting in the 2020 presidential election and its partisan composition over time. Among these questions are: How did the blue shift affect the declaration of election results for the 2020 presidential election? What was the size of the shift, if any? How did the size and trajectory of the shift vary geographically, within and between states, and within and between voting modes? Did policies that allowed for the preprocessing of mail ballots speed up the count?

We provide answers to these questions by employing time-series analysis of unofficial election night returns reported by national media organizations. Our analysis of national trends is based on web-scraping election results gathered and distributed by the National Election Pool (NEP) and reported by the *New York Times* at the county level. We use these results to identify the percentage of the two-party vote reported by counties on an hourly basis and identify the uniformity and source of shifting vote totals. We use data scraped from the Georgia Secretary of State's website to analyze shifting vote totals at the precinct level and identify trends between vote modes.

We show that a small nationwide blue shift occurred in the 2020 election, although more states had "red shifts" than in recent elections. The reporting speed of election results varied across states and local jurisdictions within states. This variability can be explained in a statistical sense by differences in state laws, such as those governing the preprocessing of election results. However, there are enough exceptions that variability in less-well-documented administrative practices must also play a role.

These results are in line with what we know about election machine capacity and operations. We discuss how these results can be better framed and communicated to the public and reporting outlets regarding election integrity.

Tabulation over Time

An election is not over until all the votes have been counted and certified. Even so, there has been a public hunger for election results immediately upon polls closing, a need that the media have been willing to feed for nearly two centuries. The advent of rail travel and the telegraph made the nearly-instantaneous reporting of unofficial election returns possible. As early as 1844, James K. Polk could wake up the morning following Election Day and know that he had won a sufficient number of states to be the next president (Borneman 2008). Even so, the quick reporting of unofficial results may not provide equally quick insight into who the likely winner is if the election is close enough (Chervinsky 2020; Mitchell 2020).

Although election returns are unofficial until they have been certified, election integrity scholars have been interested in using unofficial election returns and real-time reports of those returns to check against election fraud. Hyde (2008, pp. 204 – 5) notes dozens of indicators of "irregularities" that might be used as evidence of fraud. Two of these are based on patterns in unofficial election returns: (1) "systematically late or missing materials in opposition strongholds" and (2) "parallel vote tabulation that differs significantly from official results (determines winner within margin of error)."

These indicators might hint at fraud if four conditions hold. The first is that fraudulent activities occur primarily on Election Day, taking advantage of the government's weak centralized control of the electoral system. The second is that the perpetrators of fraud do not control the authentic ballots that were cast, and therefore need to delay the process to replace

legitimate ballots with fraudulent ones. The third assumption is that the unofficial/parallel count is the best unbiased estimate of the sincere voting patterns of the populace, under an assumption that the reporting of results from polling places is random. A fourth necessary condition implied in most of the comparative research on fraud that relies on unofficial returns is that the fraudulent actors actually have the reach required to target and conduct fraud, even though the nation has direct control over the reporting of election results (Bjornlund 2020; 2011).

While these conditions might seem unrealistic, parallel vote tabulation (PVT) aided in catching authorities in the Philippines (1986) and Panama (1989) attempting to commit fraud. Garber and Cowan (1993) also note that PVT methods tend to increase the confidence in election results by the losing parties and candidates, a determining factor in reducing post-electoral violence (Smidt 2016; Daxecker 2012).

However, it would be wrong to directly apply the principles of PVT to the U.S. practice of unofficial vote totals released during election night. The unofficial vote count issued in the U.S. consists of media reports of election results from official reporting centers. They are not a true independent count. The use of exit polls falls more in line with traditional PVTs as a sample of voters independent of election officials. However, these exit polls are far from authoritative and are susceptible to extreme error due to spatio-temporal variation and cluster sampling issues (Pavia, 2005; 2011).

Instead, within the U.S. and other nations with a running unofficial vote total, Pettigrew and Stewart (2020) note that the process of returns reporting must be thought of on two separate tracks. The first is the official count, tabulated continually until completion, albeit in a less visible manner. The second is the media tracking of the unofficial results, focused on real-time reporting and updated projections of winners as part of horse race coverage. There is a heavy

emphasis upon the media track in the U.S., which is derived from, but still an imperfect reflection of, the official count.

Beyond the technical complications and non-suitability for election night results to replace a PVT, systematic differences between voters and election infrastructure across space result in what has been called the "blue shift." Foley coined the term in noting the marginal shift of the proportion of the two-party vote towards Democratic candidates, presumably due to the non-random selection of Democratic voters relative to Republican voters into longer-to-tabulate provisional and absentee ballots following the implementation of the Help America Vote Act (Foley 2013, p. 518). This blue shift rose to prominence in the close 2018 Florida U.S. Senate and gubernatorial races, which prompted tweets from President Trump proposing that votes not counted on Election Night be discarded (Foley 2019).

Although the mechanism of provisional votes initially seems to have started the blue shift, the same process should occur whenever partisans self-select into ballot types that take longer to tabulate or report. Unofficial reports will also have non-random partisan trends if supporters of one party tend to live in the types of communities that take longer to count their ballots. This is likely true in the U.S., where Republicans tend to live in smaller counties that also tend to report election results faster than large metropolitan counties. Therefore, partisan swings and even secular trends in the reported count of the vote will often occur as a function of time.

However, lack of widespread public knowledge about the factors that might lead to partisan swings and trends in the reported vote in even a clean election can make these patterns fodder for mistaken suspicions of fraud or even disinformation campaigns by the losers. Mistaking what appears to have been a partisan shift with fraud indirectly contributed to a coup

during the 2019 Bolivian Presidential election (Idrobo et al., 2020). The partisan time path of the reported vote in states such as Arizona, Georgia, Michigan, Nevada, and Pennsylvania helped form an integral part of the Trump campaign's challenges to the results of those states.

Therefore, there is a need to systematically understand the time path of election results in the United States, both the overall pace and partisan composition. Because no credible charges of widespread fraud have been leveled in the 2020 U.S. presidential election, especially fraud that depends on the manipulation of returns and substitution of fake ballots for real ones, it is a good case study to explore what these trends look like in a no-to-low-fraud election.

Seven Hypotheses about Vote Counting in 2020

The underlying trends responsible for a blue shift interacted with political polarization during the contentious 2020 U.S. presidential election. Republican Party actors had already sowed doubt concerning election integrity related to the blue shift with allegations of fraud in Florida and New Mexico in 2018 (Foley 2019). Partisan disputes over changes in election administration practices related to the pandemic led many Republican leaders, spurred on by President Trump, to allege before the election that any drawn-out report of the election results would constitute proof of fraud. Trump acted on his claims and attempted to declare himself the winner of the 2020 presidential election on election night, based on an election night lead (Burns and Martin, 2020).

Controversy over the pace of the vote count was not confined simply to partisan sniping, however. For instance, the political data firm Hawkfish rebranded the blue shift the "red mirage" in an August report publicized by Axios, which led to a steady stream of journalistic attempts to explain the concept to the public in a more balanced way (Talev 2020). The pre-election warning about over-estimating the significance of this blue shift/red mirage was

repeated by numerous national media outlets, such as Politico, NBC News, and Reuters (Graff 2020; Kahn and Lang 2020; Wasserman 2020).

An important detail in describing and explaining the blue shift and its relation to election integrity is *where* the blue shift occurs, if it occurs. Of course, cities tend to be more Democratic than the suburbs and mainly rural areas. If there are reasons related to election administration that would lead to slower counting and/or reporting in cities, then early vote reports would be "redder" than later ones. As Foley (2013; 2015) notes, the mechanism of provisional ballots cast would be concentrated mainly among voters without the proper identification and confusion about their correct precinct, which is more likely to occur in cities.

Furthermore, although the Republican-Democratic balance in the casting of mail ballots in recent presidential elections had been relatively close, the politicized nature of the response to the pandemic resulted in Democrats choosing to vote by mail at much greater rates than Republicans in 2020 (Stewart 2021). This, too, would exacerbate the blue shift, assuming mail ballots took longer to tabulate than in-person ballots.

The widespread Democratic preference for mail ballots in 2020 could not only have heightened the urban/rural divide that leads to the blue shift, but it could also have increased *intracounty* variation in the blue shift. For example, even in highly Republican counties, if the many Republicans all cast ballots on Election Day and the few Democrats all cast ballots by mail (to state a strong counterfactual), and if Election Day votes were counted and reported before mail votes, then even Republican counties would see blue shifts. If Democrats did not prefer mail ballots at a greater rate than Republicans in this example, any blue shift observed in the state would be attributable only to intercounty variation, not intracounty variation.

These considerations give rise to our first two hypotheses.

Hypothesis 1: Counties that Biden won took longer to report results than counties that Trump won. This derives from three observations, one specific to 2020 and the other more general since the passage of HAVA. The observation specific to 2020 is that more Democratic counties likely saw proportionately more mail ballots in 2020. The two more general observations are (1) provisional ballots are generally counted days after Election Day, and Democrats are more likely to cast provisional ballots, and (2) the greater administrative complexity of large urban areas plus the sheer volume of ballots to count slow down the count in cities. These three factors together should lead Democratic strongholds to take longer to report election results.

Hypothesis 2: Counties reported votes for Biden slower than they reported votes for Trump. This derives from the observation that Democrats voted by mail at a higher rate than Republicans in 2020 and the assumption that absentee ballots took longer to count than Election Day votes. This seems like a commonsense assumption. However, it needs to be justified through a specification of the workflow of processing mail ballots compared to Election Day votes.

We provide this justification in Appendix A. That discussion emphasizes the greater complexity and number of discrete actions taken to deliver, receive, validate, and count mail ballots than in-person votes. The two processes, Election Day and mail voting, are discussed as ideal types. In practice, states and local authorities can make decisions about these counting processes that can cause counting and reporting of election results faster or slower. For instance, Election Day votes can be scanned centrally, slowing down the reporting of results compared to precinct counting. In this scenario, the ballots, not the memory cards from scanners, are delivered to the central office, where they are scanned in a process similar to that described for

the last third of the mail-ballot process. Or, local jurisdictions can be allowed to process and scan mail ballots as they arrive, which, under certain conditions, can lead to reporting of election returns very quickly on election night.

These observations together lead to our third hypothesis.

Hypothesis 3: Completing the count of mail ballots should take longer than Election Day votes. This derives from the observation that the mail-ballot process involves more steps than the Election Day process. Unless significant ballot preparation and scanning is allowed before Election Day, it will be challenging for mail ballots to be counted and reported as quickly as Election Day ballots.

There is a third mode of voting we have not addressed, early in-person voting. In terms of the procedures outlined for Election Day and mail voting, the ideal-typical early in-person voting process is a hybrid of the two. On the one hand, with early voting, the voter's identity is verified personally, making the check-in process similar to Election Day voting and less involved than mail voting. There is the complication of whether early in-person ballots are scanned on-site or transported to a central location for scanning. Still, that complication also applies to Election Day voting. On the other hand, if the local jurisdiction has multiple ballot styles, it may be necessary to sort those ballots before they are scanned, which is one delay that faces many (but not all) mail-ballot processes. Even if early voting centers can tabulate multiple ballot styles seamlessly, the memory cards of the scanners in a vote center will have votes from multiple precincts on them; those memory cards will be delivered to the central tabulation facility at different times.

This discussion leads to our fourth hypothesis.

Hypothesis 4. Completing the count of early in-person ballots will take longer than Election Day votes but less time than mail votes.

The final three hypotheses pertain to the pace of reporting election results at the state level in general. In 2020, considerable attention was paid to the unprecedented growth in mail ballots' use and the interstate variability in policies related to whether states could preprocess mail ballots. This attention leads to the final three hypotheses.

Hypothesis 5. The more absentee ballots cast in a state, the slower the reporting of election results overall.

Hypothesis 6. Allowing local jurisdictions to preprocess mail ballots will speed up reporting the vote count.

Hypothesis 7. Allowing mail ballots received after Election Day to be counted will slow the reporting of the vote count.

Data and Measures

To address the hypotheses just stated, we rely upon data from the National Election Pool (NEP) and reported by the *New York Times* and upon time-stamped election results reported by the Secretary of State of Georgia.

National Election Pool (NEP) and unofficial election returns data

The National Election Pool (NEP) uses stringers and input from state and local automated feeds to collect election returns and then distribute them to subscribing news organizations. This data collection and reporting starts from the moment polls are closed, and initial results are reported, through to the certification of elections, often many weeks later. Many of the organizations that subscribe to the NEP feed, in turn, repackage those reports through their websites and other avenues. One such organization is the *New York Times,* which we rely on here.

Starting on election night, we scraped the *New York Times* election results at five-minute intervals between pulls, going through the 50 states and the District of Columbia until the final tally was reported.[1] The script cycled through the state landing pages; the program run-time ranged between ten and twenty minutes. Excluding failed pulls at specific points in the program's life of 158 hours with 473 snapshots; the program averaged a snapshot every 20 minutes and 2 seconds. We also used the Internet Archive's Wayback Machine to get time-stamped results to make up for errors or missed updates.[2] Furthermore, starting on November 12, we switched over to scraping the JSON source of the *New York Times* data, instead of the web page itself, at three-minute intervals.[3]

Before proceeding, we must be very clear about one thing. The data we gathered are the election results *as they were reported by the NEP and distributed by the New York Times.* We did not systematically gather election results directly from state and local election offices for comparison, nor did we compare the results with those collected by the competing organization, the Associated Press. We did scrape the websites of a small number of official state election-night reporting sites and have spot-checked those reports against the NEP data.[4]

The research in this paper proceeds under the assumption that the timing of the release of the reported results corresponds very closely to the release of election results by the states and localities. Because of competitive pressures between the NEP and the AP, in addition to those

---

[1] "Presidential Election Results: Biden Wins," *New York Times,* https://www.nytimes.com/interactive/2020/11/03/us/elections/results-president.html?action=click&pgtype=Article&state=default&module=styln-elections-2020&region=TOP_BANNER&context=storyline_menu_recirc.

[2] Internet Archive, "Wayback Machine," https://archive.org/web/.

[3] The JSON source is at this URL: https://static01.nyt.com/elections-assets/2020/data/api/2020-11-03/national-map-page/national/president.json.

[4] The New York Times state level data appears to have 158 observations where votes were retracted, which amounts to 0.11 percent of the data.

among all the subscribing news organizations, this assumption seems reasonable because there is a large premium on speed.[5] However, we have spotted occasional errors, some of which went uncorrected for several hours and even days. (Some of these errors have been fodder for conspiracy theorists.) The errors appear to be infrequent and do not affect the overall analysis presented here.

Beginning on election night and continuing until November 12, we downloaded state election results from the *New York Times* election result website every 10-20 minutes, yielding 473 datasets with vote totals for the major presidential candidates. Further, we added Wayback Machine snapshots to fill in missed places, contributing an additional 4,865 snapshots of the *New York Times* state results webpages, national results webpages, and original data files between November 3 and November 12. Once we switched over to the JSON source on November 12, we began scraping the source data in 4 – 5 minute intervals, which yielded 7,554 snapshots between November 12 and December 11.

We can compare the 5,098 unique updates we recorded to the 10,963 records in the "time series" section of the source data, which appears to have a complete record of their updates. Typically, the election results for state *s* at time *t* were identical to those reported 15 minutes before. In these cases, we removed duplicates, allowing us to focus on changes to vote counts. Changes to the vote counts can show a decline in votes from one time to the next, owing to error

[5] We feel it necessary to make this point because one of the conspiracy theories floating around after the election related to changed election results that were detected by others who were scraping the same data, either from the *New York Times* or other media outlets. State and local officials make mistakes in reporting election results, and sometimes those changes reflect correction of mistakes made by those officials. In other instances, the changes are necessary because of data-entry errors on the part of NEP Research that were corrected. For those looking at these data, hoping to audit election results, it is important to keep in mind that data-entry errors are likely to be greater than errors made by election officials, mainly because data-entry by NEP is more likely to be done manually.

corrections, but such instances were infrequent. (Only 0.014 percent of reports show a decline in votes from the previous report.)

We analyze two major metrics in this paper, (1) total votes reported by state *s* by time *t* and (2) the two-party vote share *difference* in state *s* at time *t*. When comparing across states, it is convenient to normalize these measures. Total votes reported by time *t* are normalized to be the percentage of the total number of votes in the final vote count. In other words,

$$Normalized\ vote_{s,t} = \frac{v_{s,t}}{V_s}, \text{ where} \quad (1)$$

$v_{s,t}$ = votes reported in state *s* by time *t* and V_s = final votes counted in state *s*.

The two-party vote share is normalized to the percentage-point difference between the two-party vote share at time *t* and the percentage-point difference in the final count. Positive values of the measure indicating Biden's vote share at time *t* was greater than the final count, and negative values indicating Trump's vote share was greater than the final count. In other words,

$$Normalized\ two-party\ vote\ share_{s,t} = \frac{v_{Biden,s,t}}{v_{Biden,s,t}+v_{Trump,s,t}} - \frac{V_{Biden,s}}{V_{Biden,s}+V_{Trump,s}}, \text{ where} \quad (2)$$

$v_{Biden,s,t}$ = votes for Biden reported in state *s* by time *t*, $v_{Trump,s,t}$ = votes for Trump reported in state *s* by time *t,* $V_{Biden,s}$ = final votes for Biden counted in state *s,* and $V_{Trump,s}$ = final votes for Trump counted in state *s*.

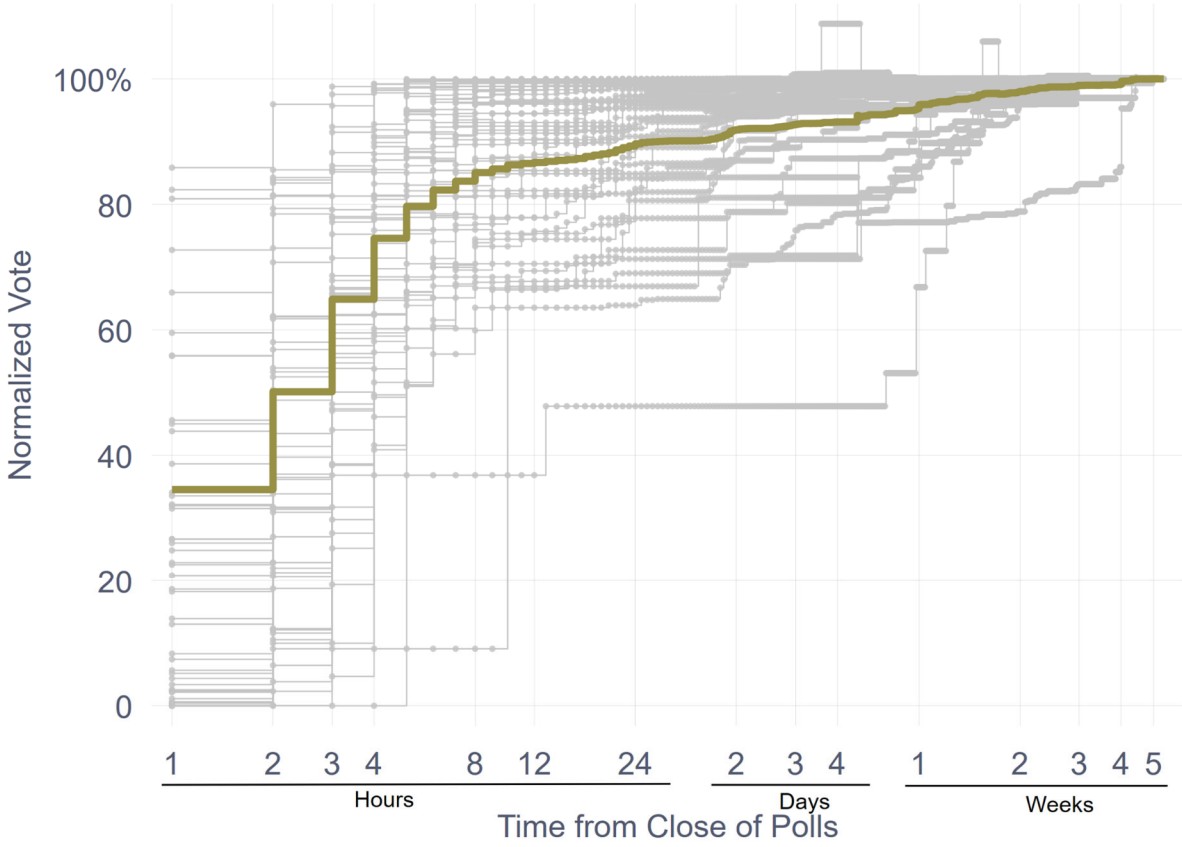
In Appendix B, we illustrate the NEP data stream using the votes reported from Georgia as an example. But, Georgia's vote reporting trajectory is just one of several seen in the data. We summarize the ballot-reporting paces of all states in Figures 1 and 2, using the normalized vote measures described above. We have performed two additional normalizations in Figures 1 and 2 that carry through to the end of the paper. First, it is often convenient for us to characterize the state of vote reporting at particular time slices. To do so, we created a dataset that records

the two normalized measures at one-hour intervals for every state, which is what we display in Figures 1 and 2. For comparison, we show versions of Figures 1 and 2 with the actual reporting times in Appendix C. Second, we draw attention to the normalization along the *x*-axis, as well. Because of time-zone and poll-closing-time differences, we have normalized the clock of each state to start at its poll-closing time. This means that when we take a slice of time, we are analyzing an aggregate that no television viewer of voting returns on election night would have experienced, since, on election night, the public is viewing an aggregate outcome that is based on a collection of vote reports that reflect different time zones and closing times.

In Figure 1, the thin grey lines indicate the normalized state votes at one-hour intervals, while the thick gold line indicates the normalized vote across all states. The interstate variability in the pace of vote reports was quite substantial.[6] For instance, at one hour after polls closed, 34.6 percent of all votes had been reported. However, no votes had been reported from eight states, while Florida had already reported 9,502,374 votes, 85.9 percent of its final tally. The standard deviation across all states at that point was 25.0 points. Even at the twenty-four-hour mark, when 89.7 percent of all votes had been reported, Alaska was stuck at 47.8 percent, while eighteen states were above 99 percent. The standard deviation was 11.8 points.

[6] A few reports resulted in the cumulative number of votes at a point in time being greater than the final total certified by the state. With one exception, these were by small amounts that were corrected quickly. The one exception was a report of 50,238 new votes for Trump and 26,691 new votes for Biden in West Virginia, issued at 9:18:12 a.m. on November 7, which took the total number of votes to 108 percent of what was eventually certified in West Virginia. That report was rescinded at 11:44:48 a.m. the next day.

Figure 1. Normalized votes reported for each state on an hourly basis.



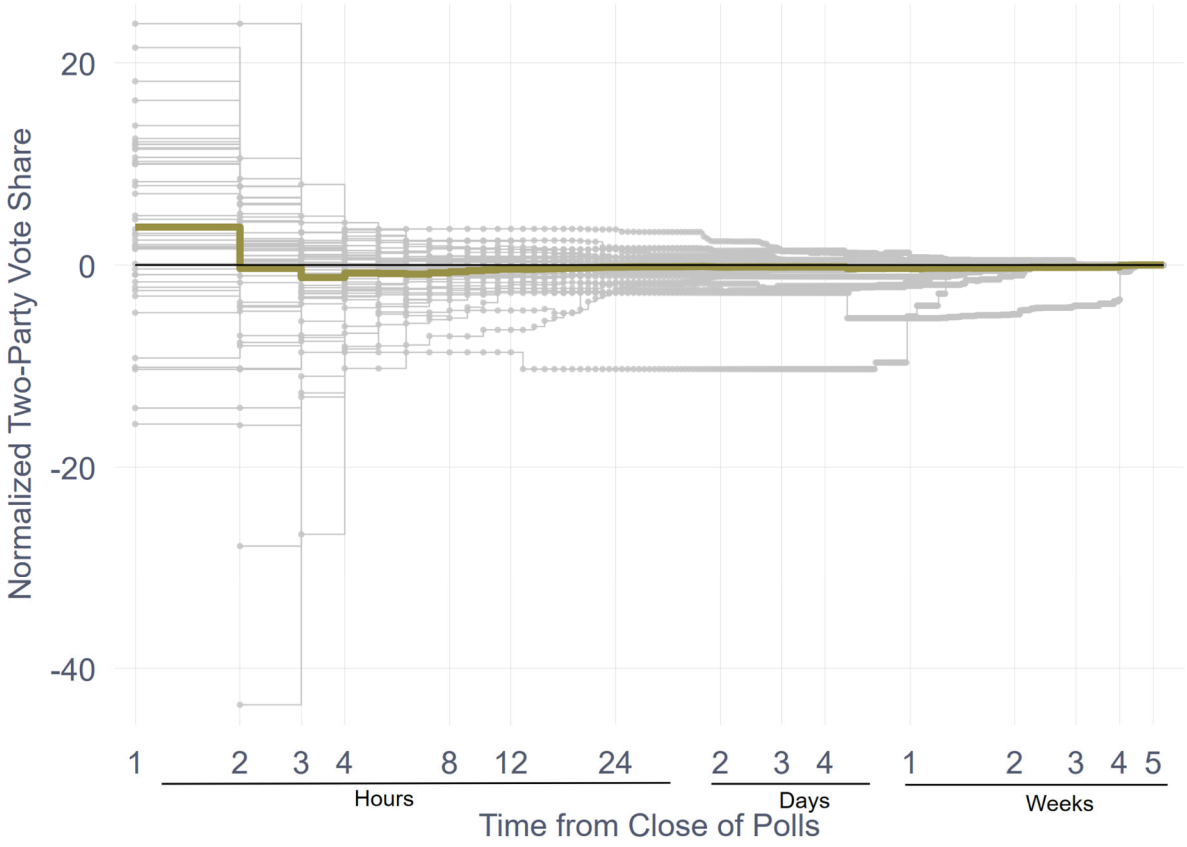
Note: The grey lines indicate the cumulative number of votes reported in each state, as a percentage of the final vote total in that state, with the circles indicating when new votes were reported. The gold line indicates the nationwide normalized vote, calculated hourly starting one hour after polls closed. Note that the *x*-axis is the logarithm of the number of hours from poll closings.

Data source: NEP via the New York Times

In Figure 2, we summarize the normalized two-party vote share. As with the total number of votes reported, there was considerable interstate variability in the two-party vote share, particularly in the first eight hours after polls closed. The votes reported one hour after polls closed gave Biden 56.8 percent of the vote, 3.9 points more than his final 52.3 percent of the two-party vote. However, in the second hour, Biden fell to near equality to his final share. By

hour 6, he was down by 0.91 points, at which point he began to recover. Still, the climb was slow, and Biden did not reach his final vote share until all the votes were counted.

Figure 2. Normalized two-party vote share in each state on an hourly basis.



Note: The grey lines indicate the normalized vote share in each state, with the circles indicating when new votes were reported. The gold line indicates the nationwide cumulative vote share, calculated hourly starting one hour after polls closed. Note that the *x*-axis is the logarithm of the number of hours from poll closings.

Data source: NEP via the New York Times

The nationwide deviation in Biden's two-party vote share is directly analogous to the blue shift, although to convert the deviation shown in Figure 2 to the Foley blue shift measure requires us to reverse signs. The national time series in Figure 2 shows how the magnitude, and even valence, of the blue shift depends on the starting point. For instance, if we compare the vote share at the one-hour mark to the final vote share, there was a *red* shift in the 2020 election.

However, for most of the counting period, Biden was trailing his final marks. If we judge the change from the 30-hour mark, which is close to Foley's (Foley 2013; Foley and Stewart 2015, 2020) benchmark of Thursday morning, then the nationwide blue shift in 2020 was 0.17 points, much less than 2016 (1.02 points) and 2012 (0.75 points).[7]

Probably the most significant takeaway from Figure 2 is the interstate variability in the blue shift. This variability diminished quickly, in contrast to the interstate variability in the pace of reporting overall. At the one-hour mark, the standard deviation of the data displayed in Figure 2 was 9.0 points. It was down to 3.3 points at four hours, dropping to 2.2 points at 8 hours and 1.8 points at 24 hours. Furthermore, whether a state would exhibit a long-term blue- or red shift was set mainly by Wednesday around noon. In the previous paragraph, we noted that Foley and Stewart set the benchmark for the size of the blue shift at approximately 30 hours after polls had closed. Looking at the more complete time series, we note that the sign of the blue shift at the 16-hour mark — roughly Wednesday at noon — was identical to the sign of the blue shift at the 30-hour mark for 45 states. In other words, nearly ninety percent of states showed a consistent pro-Biden or pro-Trump drift in vote share from Wednesday at noon until the final certification.[8]

Georgia Secretary of State election return data

We rely upon data from the Georgia Secretary of State's election reporting website for our deeper analysis into voting mode. Georgia is one of several states that relies upon an election-night reporting system provided by the company Scytl. These files identified election returns by

[7] To be clear, the measure of the 2020 blue shift we offer here, which is 30 hours after polls closing, is not strictly comparable to Foley's, because we have not taken time-zone differences into account.

[8] Another way to measure the consistency of state reports over the course of vote counting is to measure the point at which either Biden or Trump took the lead in a state, never to relinquish it. By this measure, the reported results in 2020 were even more consistent than what we report here. Within four hours of polls closing, the leader in 45 states (including D.C.) was the eventual victor in that state. By eight hours, that number had risen to 48 states. Within twenty-four hours, the only two states that were showing a lead in the reported vote for the candidate who eventually lost the state were Georgia and Pennsylvania.

county and mode — Election Day, early in-person (advance), absentee (by mail), and provisional. We downloaded all of the site's updates following the closing of the polls, 213 in all. The data provide 213 modal updates following the initial close of the polls for the 159 counties, providing 169,335 groupings of county-time-mode observations. In Figure 3, we have graphed the normalized vote for Georgia, taking counties (Figure 3a) and precincts (Figure 3b) as the basic unit of analysis.

Figure 3. Normalized votes reported in Georgia on an hourly basis.

a. County data displayed

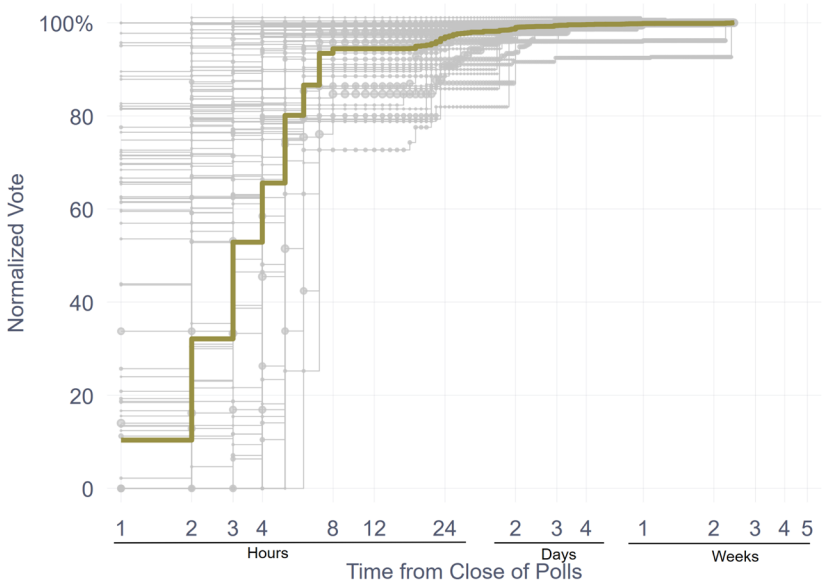
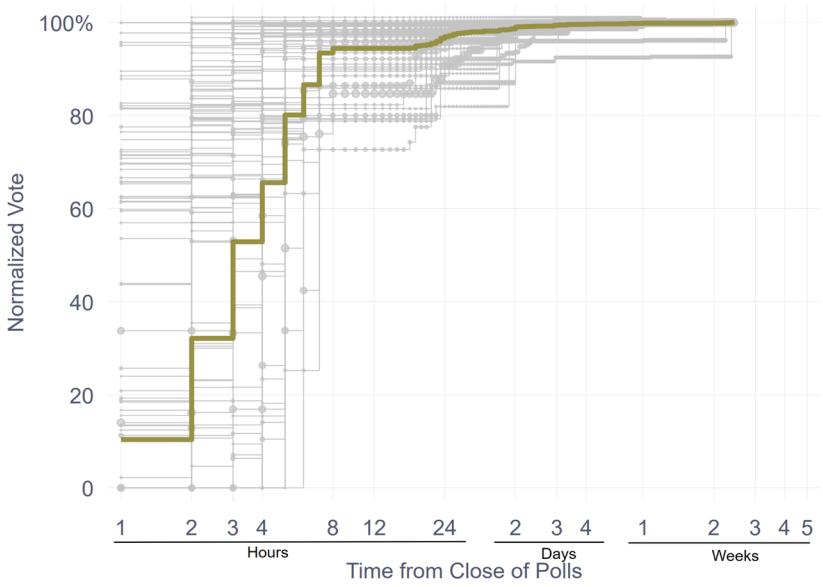


Figure 3 continued.

b. Precinct data displayed



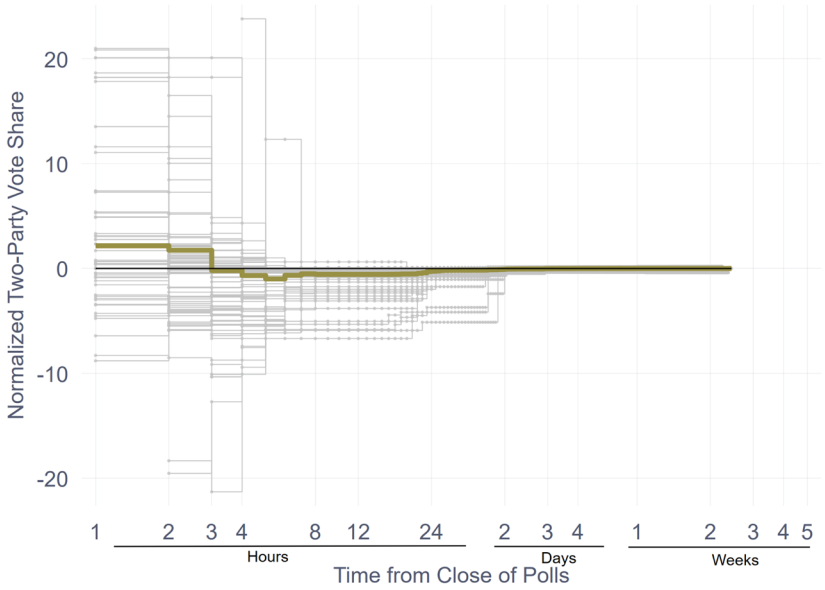
Note: The grey lines indicate the cumulative number of votes reported in each county or precinct as a percentage of the final vote total in that county or precinct. The circles indicate when new votes were reported. The gold line indicates the normalized statewide vote, calculated hourly starting one hour after polls closed. Note that the *x*-axis is the logarithm of the number of hours from poll closings.

Data source: Georgia Secretary of State

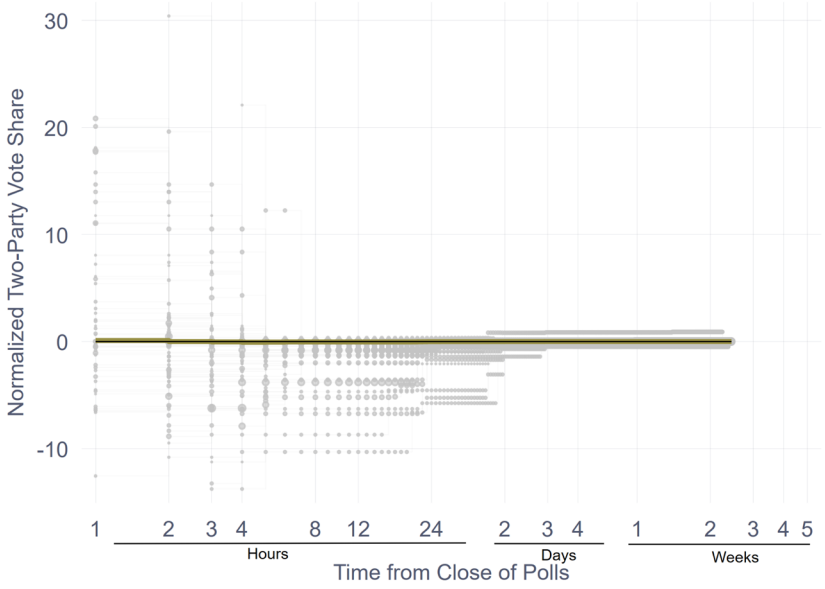
Figure 4, likewise, graphs the normalized two-party vote on an hourly basis, with data for counties (Figure 4a) and precincts (Figure 4b) shown.

Figure 4. Normalized two-party vote share reported in Georgia on an hourly basis.

a. County data displayed



b. Precinct data displayed



Note: The grey lines indicate the normalized vote share in each county or precinct, with the circles indicating when new votes were reported. The gold line indicates the statewide cumulative vote share, calculated hourly starting one hour after polls closed. Note that the *x*-axis is the logarithm of the number of hours from poll closings.

Data source: Georgia Secretary of State

State Reporting and Mail Ballots

One of the most significant issues related to the speed of counting that occupied public discourse before the 2020 election was mail ballots. In many quarters it was believed that three factors would lead to long delays in counting them: (1) the longer time it was assumed that it would take to count mail ballots than in-person ballots, (2) the limitations on, and even prohibitions against, processing mail ballots before the Election Day polls had closed, and (3) allowances by roughly twenty states for mail ballots to arrive after Election Day if they were postmarked by Election Day. These assumptions about mail ballots led to Hypothesis 5, which is that states that had more mail ballots to count would be slower to report, Hypothesis 6, which is that states that prohibiting the preprocessing of mail ballots would also slow the vote report, and Hypothesis 7, which is that allowing mail ballots that arrived after Election Day to be counted would slow down the count.

We can test these three hypotheses with the NEP data described in the previous section. We measure the speed of counting as the percentage of ballots reported by the states at five different time markers after the close of polls: 4 hours, 8 hours,12 hours, 24 hours, and 48 hours. We measure the percentage of voters casting ballots by mail using responses to the November Supplement of the U.S. Census Bureau's Current Population Survey, which asked voters which mode they used to cast ballots.[9] We measure which states prohibited preprocessing of ballots and whether late-arriving ballots could be counted using information collected and reported by the National Conference of State Legislatures. For the latter measure, we coded the number of

[9] We did not use administrative data because vote-by-mode is not reported by all states. The EAC's Election Administration and Voting Survey (EAVS), which records state reports of this statistic, has not been released for 2020 at the time of the writing of this paper. Our experience from previous years is that self-reports of voting modes and administrative statistics are highly correlated.

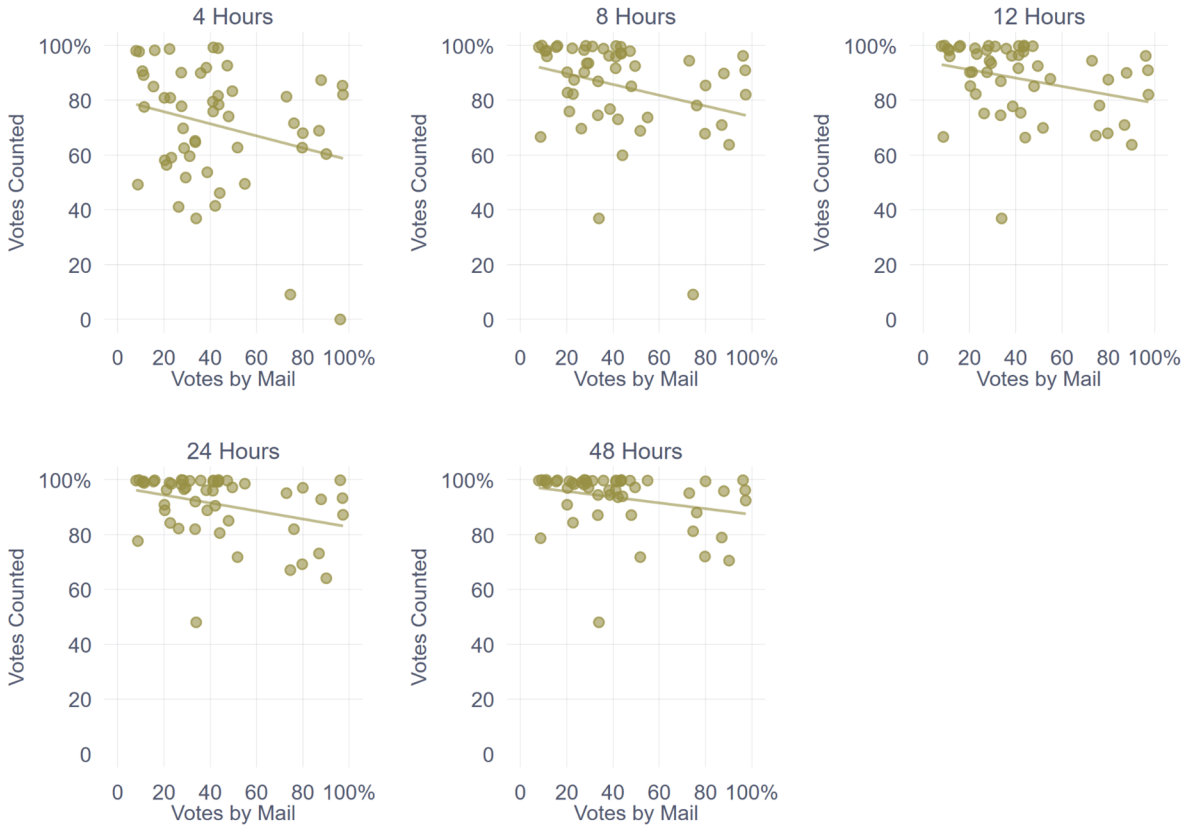
days after Election Day that ballots could arrive, so that we could distinguish between, for instance, Texas, which allowed mail ballots to be counted if they arrived the day after Election Day, and California, which had a mail-ballot-return window of seventeen days.[10]

Figure 5 illustrates the simple bivariate relationship between the percentage of votes reported at the various time points and the percentage of ballots cast by mail. (The lines represent the least-squares fit to the data.) As time progressed, the percentage of votes reported increased, of course, except for the outlier, Alaska.[11] At the four-hour mark—roughly midnight—there was a weak negative relationship and considerable variability in the dependent variable. At the other time marks, the negative relationship becomes more apparent. Still, outliers are also evident, which provides an initial indication that the speed of counting was not tied only to the volume of mail ballots.

[10] Alaska's 10-day window to return ballots is the longest under normal circumstances, but California's deadline was extended to 17 days due to the pandemic. See NCSL, "Absentee and Mail Ballot Policies in Effect for the 2020 Election," November 3, 2020, https://www.ncsl.org/research/elections-and-campaigns/absentee-and-mail-voting-policies-in-effect-for-the-2020-election.aspx.

[11] Alaska is the outlier that becomes apparent at the 8-hour mark. Under Alaska's mail-ballot-counting process, all mail ballots have to be processed and compared against poll books before the ballots can be counted. This delayed Alaska's mail-ballot counting until ten days after the Election. This means that for the period covered in the regression analysis to follow, Alaska's reported vote totals could not contain mail ballots. For that reason, the regressions omit Alaska. See Brooks (2020).

Figure 5. Progress of Results Reporting, as a Function of Mail Ballot Volume



Data sources: NEP via the New York Times; National Conference of State Legislatures

Table 1 reports the results of regressions that test the three hypotheses directly. (Alternative specifications are explored in Appendix D.) We excluded Alaska for reasons explained in footnote 11. Across all five regressions, the signs of the coefficients are in the expected direction. The impacts of the mail-ballot percentage and the preprocessing ban both declined over time, but even at the end of 48 hours, they still had an impact. The effect of extended deadlines to return mail ballots remained negative and relatively stable, starting at four hours and lasting to 48.

Table 1. Progress of Results Reporting, as a Function of Mail Ballot Volume, Preprocessing Rule, and Ballot Return Deadline. (Robust standard errors. Alaska excluded. N = 50)

	4 Hours	8 Hours	12 Hours	24 Hours	48 Hours
Mail pct.	-0.311* (0.153)	-0.219** (0.074)	-0.155*** (0.045)	-0.128** (0.043)	-0.0817** (0.0346)
No preprocessing	-17.7** (5.71)	-15.1** (5.2)	-8.65** (2.68)	-5.73** (2.03)	-3.38* (1.47)
Return deadline	0.079 (0.638)	-1.35** (0.41)	-1.14*** (0.25)	-1.28*** (0.20)	-1.19*** (0.18)
Intercept	90.6*** (6.3)	105.0*** (4.42)	102.0*** (2.07)	103.6*** (1.86)	102.8*** (1.37)
RMSE	19.4	12.4	8.52	6.81	5.31
R^2	.226	.440	.469	.561	.597

***$p < .001$ **$p < .01$ * $p < .05$

The effect of the extended return deadline on ballot counting is puzzling in one regard. It is understandable why the extended receipt deadline would have slowed down vote counting 24- and 48-hours after polls closed. Still, it is not obvious why the deadline would have had an effect at eight and twelve hours since the only ballots that could have been counted at that point had arrived on Election Day.[12] The best explanation is that the extended return deadline gave election officials license in these states to count mail ballots more slowly than they might have otherwise, even mail ballots that had arrived by Election Day.

Overall, these results demonstrate that state policy choices, whether emergency or long-standing, had consequences for the speed of vote reporting in different ways. Immediately after

[12] Of course, it is possible that some ballots arrived after polls closed on Election Day, but this number must have been trivial.

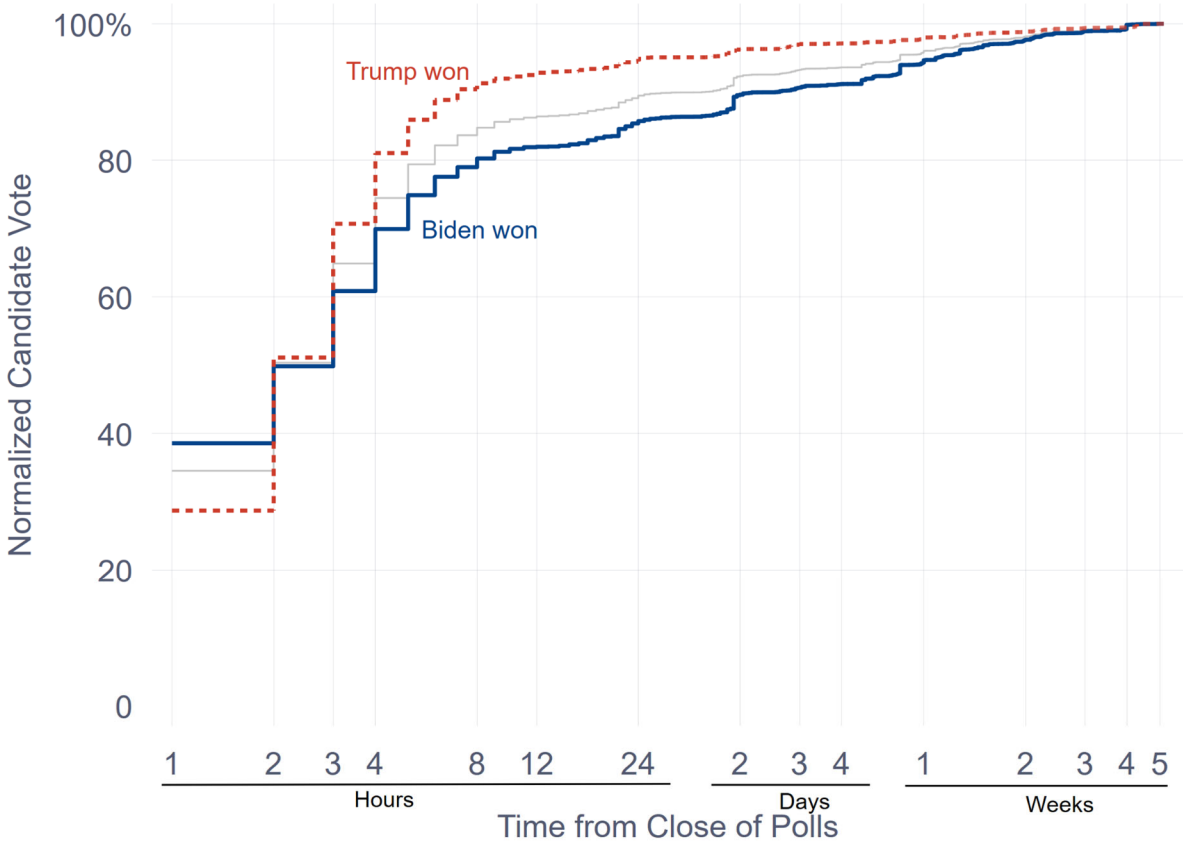
polls closed, the volume of mail balloting and the prohibition against preprocessing slowed down vote reporting, but not the ballot-return window. Eventually, the influence of these policy choices declined, but they were persistent, at least through the first two days of counting.

County Variation in Reporting

In general, local governments, not states, are responsible for counting and reporting election results. Counties within states vary. As far as the pace of reporting election results and partisan trends seen in those results are concerned, the three most important sources of variation are partisan disposition, size, and institutional capacity. Hypothesis 1 addressed the first of these factors by stating that we should expect Democratic counties to report more slowly than Republican counties. The simplest way to illustrate this hypothesis is provided in Figure 6, where we have plotted the normalized vote for Biden and Trump on an hourly basis separately.[13] (We show both the data and the averages in Appendix E.)

[13] In this and all of the analyses that rely on the NEP county dataset, we have removed counties whose final vote count is more than one percentage point different from the maximum vote count in the time series, or whose final vote count for one of the two candidates is greater than one percentage point from the maximum vote count. This is to eliminate counties for which a major reporting error necessitated a correction. Of the 3,159 counties in the dataset, only 69 were thus eliminated.

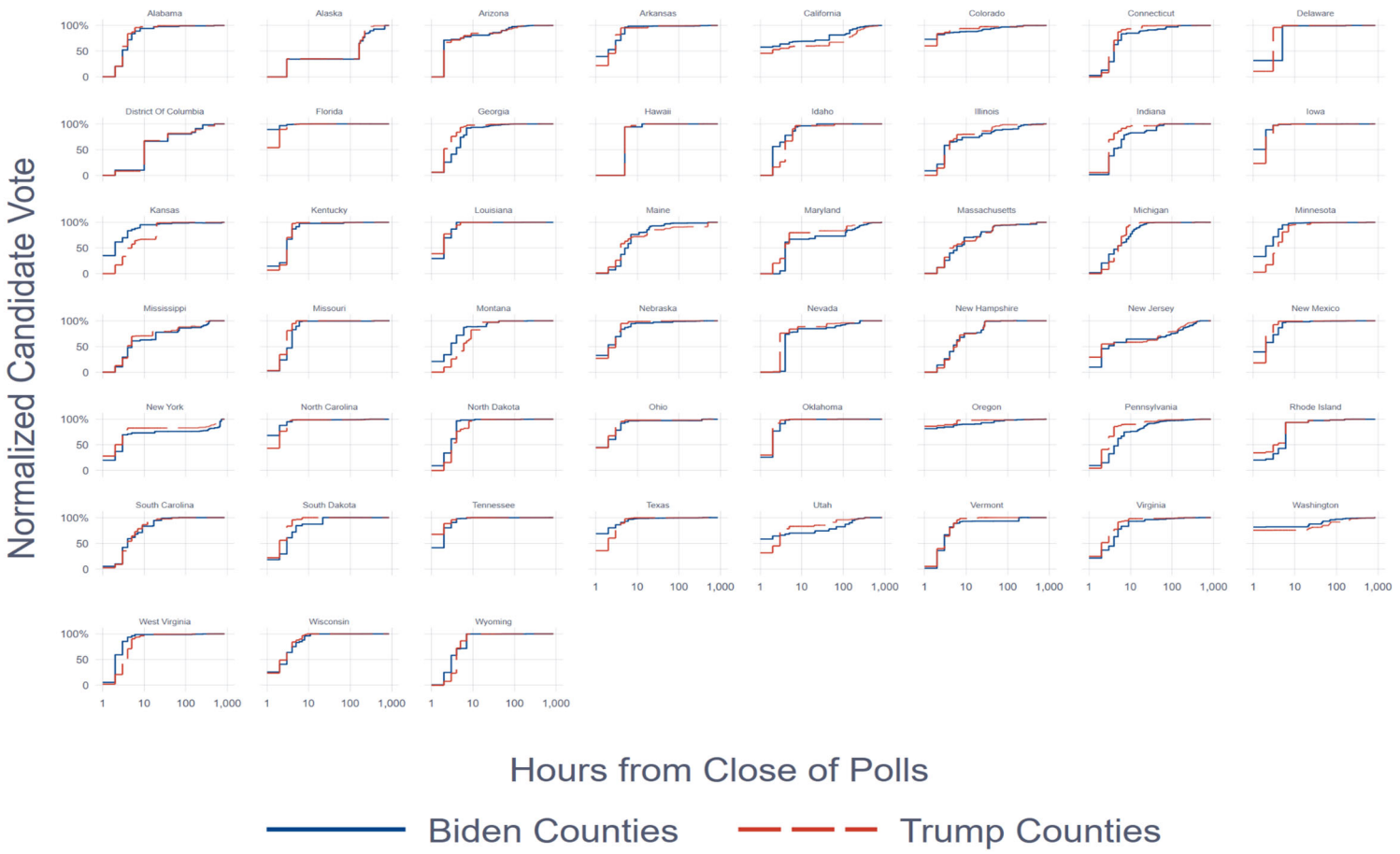
Figure 6. Percentage of County Votes Reported, by Time Since Polls Closed



In Figure 6, we emphasize the hourly averages in the normalized votes, with the solid blue line showing votes in counties Biden won and the dashed red line showing votes in counties Trump won. In the first hour, counties that Biden won actually reported more quickly than counties that Trump won; by that first hour, Biden's counties had reported 37.6 percent of all the votes they would eventually report, compared to 28.7 percent for Trump's counties. The two sets of curves switched places starting in the second hour, but just barely. By the third hour, counties that Trump won had reported 70.4 percent of their votes, compared to 60.6 percent for Biden counties. It was not for another two weeks before Biden's and Trump's counties reached parity in the fraction of the vote they had counted.

In general, then, Hypothesis 1 is sustained when we look at the nationwide data. This pattern also holds for most states, although there are some for which the pattern does not fit. This is illustrated by Figure 7, where we have graphed the hourly average of the normalized vote for each state. Because there are states where one of the candidates won nearly every county, we have altered the definitions of "Biden counties" and "Trump counties." For Figure 7, a Biden county is one that gave Biden a vote share above the median of all counties in that state, with the rest defined as "Trump counties." For most states, the Biden counties outpaced the Trump counties in the first hour. By the third hour, Trump counties were ahead in two-thirds of states; Trump counties outpaced the Biden counties for the early hours before eventually converging. The gap is persistent in two states where the pace of vote reporting was controversial, Georgia and Pennsylvania. On the other hand, in other battleground states, such as Florida, Michigan, New Hampshire, and Wisconsin, Biden- and Trump counties reported at approximately the same rates. The one state where strong Biden counties notably reported votes before strong Trump counties was Minnesota.

Figure 7. Percentage of County Votes Reported, by Time Since Polls Closed, by state.



Hypothesis 2 relates to Hypothesis 1, except that it focuses on the pace of reporting votes for Trump and Biden separately *within counties*. Once we examine within county variation in vote reports, were Trump votes reported before Biden votes?

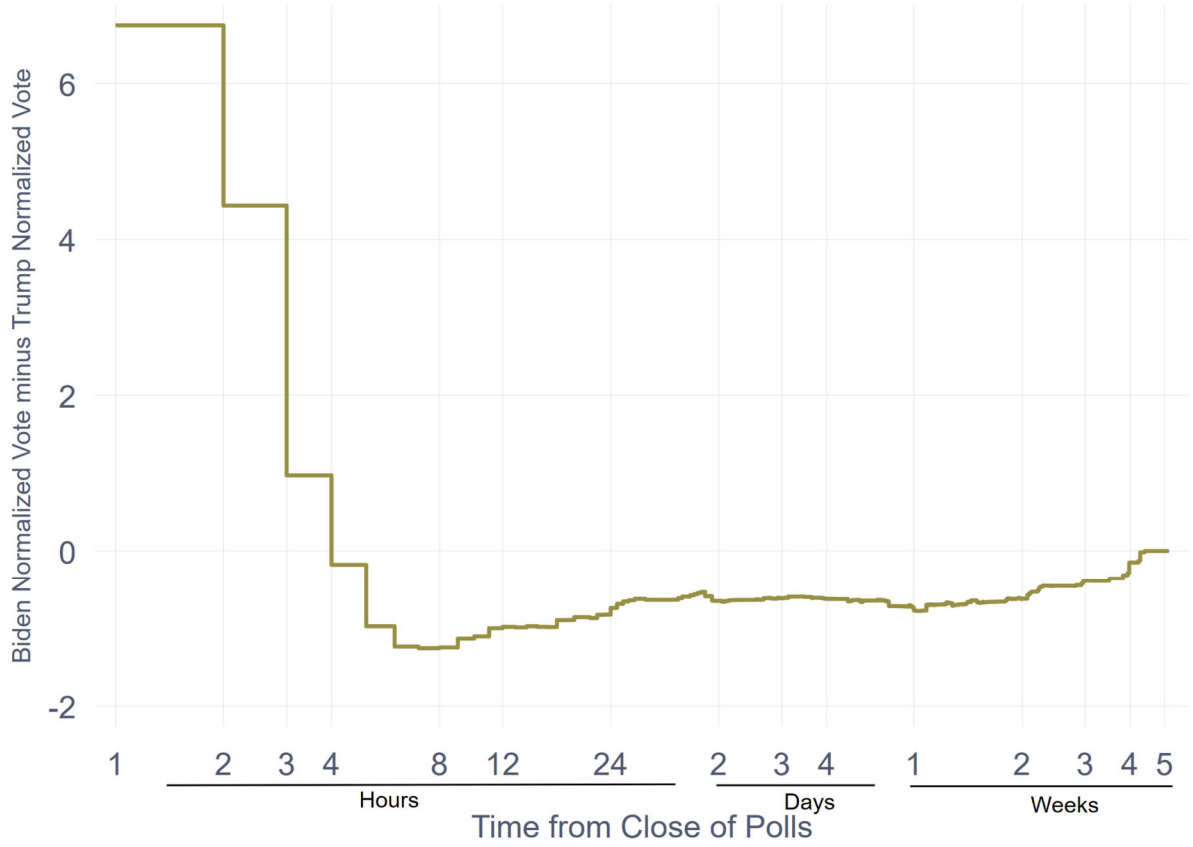
To answer this question, we started by adapting the normalized vote metric by creating a variable that measured the relative number of votes reported for each candidate in each county, which we term the normalized candidate vote. It is defined as follows:

$$Normalized\ candidate\ vote_{c,y,t} = \frac{v_{c,y,t}}{V_{c,y}}, \text{ where} \quad (3)$$

$v_{c,y,t}$ = votes reported for candidate $c \in$ (Biden, Trump) in county y by time t and $V_{c,y}$ = final votes for candidate c counted in county y. We then calculated the difference between the two normalized candidate vote values for each county, with positive values indicating that Biden's votes were counted more quickly, i.e., $\frac{v_{Biden,y,t}}{V_{Biden,y}} - \frac{v_{Trump,y,t}}{V_{Trump,y}}$. We then calculated the weighted average of these differences across counties for each hour for those counties still reporting votes.[14] Figure 8 shows the analysis graphically. (Appendix F shows Figure 8 with the county data.)

[14] This means that we exclude counties before they have reported their first vote and after they have reported their last vote. This allows us to focus on the period when counties were actively reporting votes, and also to account for the different canvassing periods of each state.

Figure 8. Average difference in reporting pace of Biden and Trump votes, by time since polls closed.



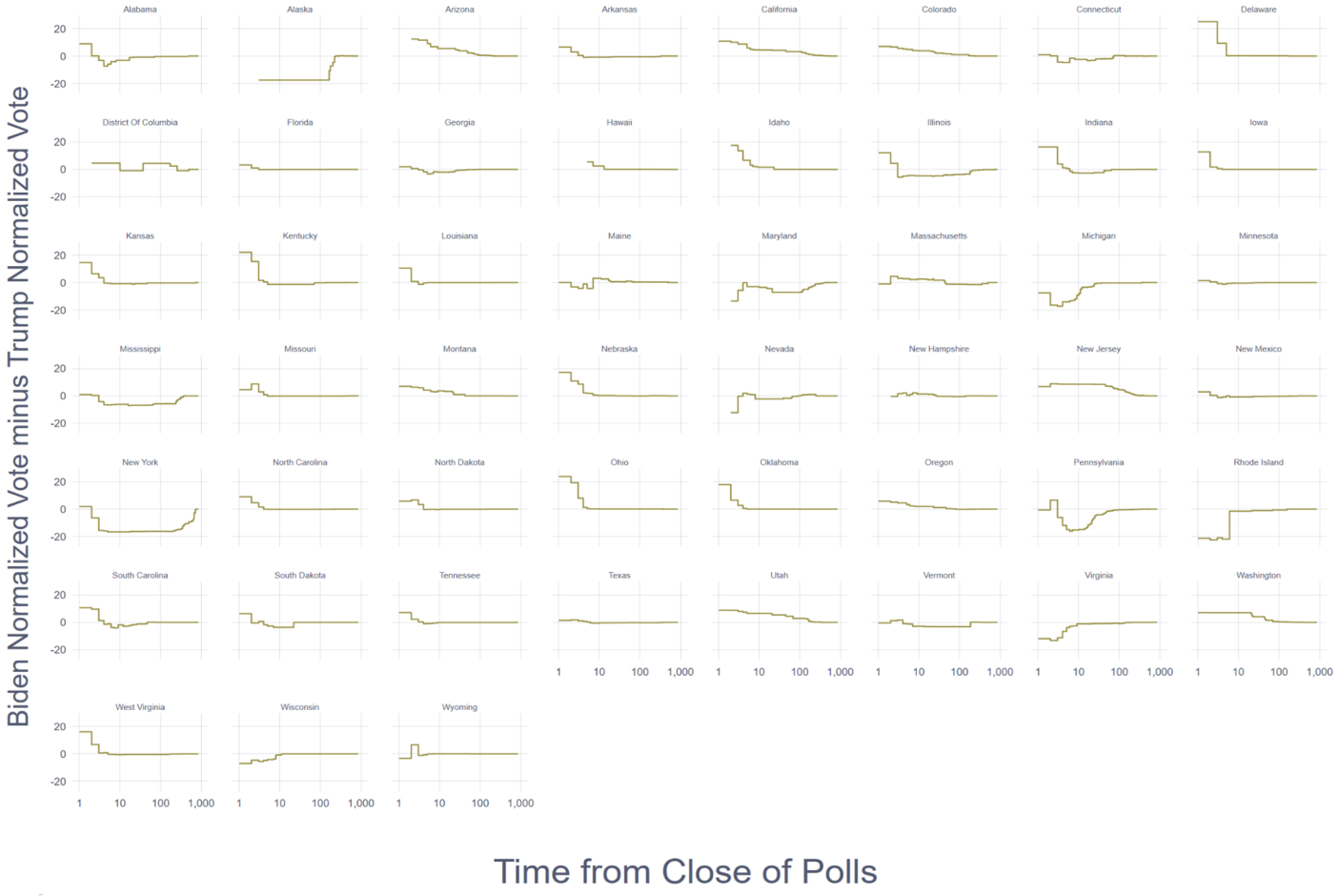
Although for most of the period when votes were being reported, Biden's votes were being reported more slowly than Trump's. However, that was not the case for the first two hours after polls closed. Among counties that could report results within an hour of polls closing, 61.6 percent of Biden's eventual votes were reported, compared to 55.5 percent of Trump's, for a difference of 6.1 points. Biden's votes continued to be counted more quickly than Trump's in the second and third hours, but by hour four, counties tended to report more of Trump's votes than Biden's. Applying a series of t-tests to all the hourly means displayed in Figure 8, Trump's voting pace within counties continued to exceed Biden's until hour 736 (30 days). At that point,

there was no statistically significant difference between the two paces, although the sign of the mean continued to be negative.

This time path is entirely consistent with counties that could preprocess their absentee ballots reporting absentee ballot results first, followed by Election Day ballots. If that is the explanation for the pattern in the first couple of hours, then this "preprocessing effect" was overtaken by the deluge of Election Day vote reports at about the three-hour mark.

As before, we are interested not only in the national pattern but also in the state patterns. Figure 9 extends the analysis just reported by graphing average measures of relative reporting pace for each state. For most states—35 of 43 that reported votes in the first hour—counties were further along in reporting votes for Biden in the first hour than Trump's votes. And, in most states, the reporting of Trump votes rapidly caught up.

Figure 9. Average difference in reporting pace of Biden and Trump votes, by state



However, close perusal of the graphs in Figure 9 reveals several trajectories. A typical pattern, evident in eight states,[15] saw Biden's vote report outpacing Trump's for several hours, before the two reached parity. Each of these states, except Montana, mailed ballots to all registered voters. Two notable battleground states about which there was considerable controversy, Michigan and Pennsylvania, saw Trump's votes reported at a proportionately faster rate than Biden's for a full day before the reporting pace came into alignment.

Therefore, the results for Hypothesis 2 are mixed. Viewed nationally, immediately after polls closed, counties generally reported Biden votes disproportionately faster than Trump's. This pattern quickly reversed in keeping with the hypothesis. However, when we break down the analysis by state, we see numerous states for which counties were more likely to report out Biden votes quicker than Trump's for an extended period. This was particularly true for states that mailed ballots to all registered voters. Again, this is likely to have been a product of preprocessing. If Biden voters sent back their ballots more quickly than Trump voters, then the first batches of processed mailed ballots would have reflected a greater proportion of votes for Biden. Once again, the details of election administration and voters' behavior complicate a simple story.

The Speed of Counting Mail Ballots: Evidence from Georgia

The two hypotheses that have remained unaddressed pertain to the speed of counting and reporting votes by different modes. Hypothesis 3 expects that mail ballots will take longer to count and report; Hypothesis 4 expects that Election Day votes will be the most quickly counted and reported, with early in-person votes somewhere in between.

[15] Arizona, California, Colorado, Montana, New Jersey, Oregon, Utah, and Washington.

Unfortunately, although the NEP separately reports absentee votes, the nature of those reports makes them unreliable to test these hypotheses for two reasons. First, the NEP report of absentee ballots generally combines results from mail ballots and early in-person votes. This makes a clean test of the mail-ballot-vs.-Election-Day votes hypothesis impossible.

Second, upon close inspection of the NEP data, it is clear that the absentee ballot report was not always updated when new votes were reported, even when those new votes include absentee votes. This is obvious from a frequent pattern in the dataset, whereby the total votes reported in a county increased, but the separate count of absentee ballots did not change. Later, the absentee ballot total would increase, but the value of the total number of votes would remain unchanged. In some instances, the absentee ballot accounting caught up quickly; in other cases, the absentee ballot report remained un-updated for days. In a few instances, counties that clearly had absentee ballots, such as Broward County, Florida, were never credited with having any absentee ballots at all.

Therefore, we are unable to test Hypotheses 3 and 4 using the NEP data. Instead, we look at the issue using data from Georgia, where we have access to reliable election reports by voting mode. Although we cannot generalize to other states, we can provide a framework for conducting this analysis when similar data are available.

The data we use were scraped from the Georgia Secretary of State's website. Unlike the NEP data, which is a report of what states were reporting and thus subject to translation errors, the Georgia data is precisely what the state was reporting and was available to the public in real-time. Our comparison of NEP reports with the Secretary of State's reports indicates that the total number of ballots reported by the former source typically followed quickly on the latter's heels, within a matter of minutes and even seconds. Furthermore, in addition to providing real-time

reports by mode—Election Day, early in-person, mail, and provisional—the state source also did so at the precinct level. We can thus directly test issues of intra-county heterogeneity that the NEP data did not allow us to explore.

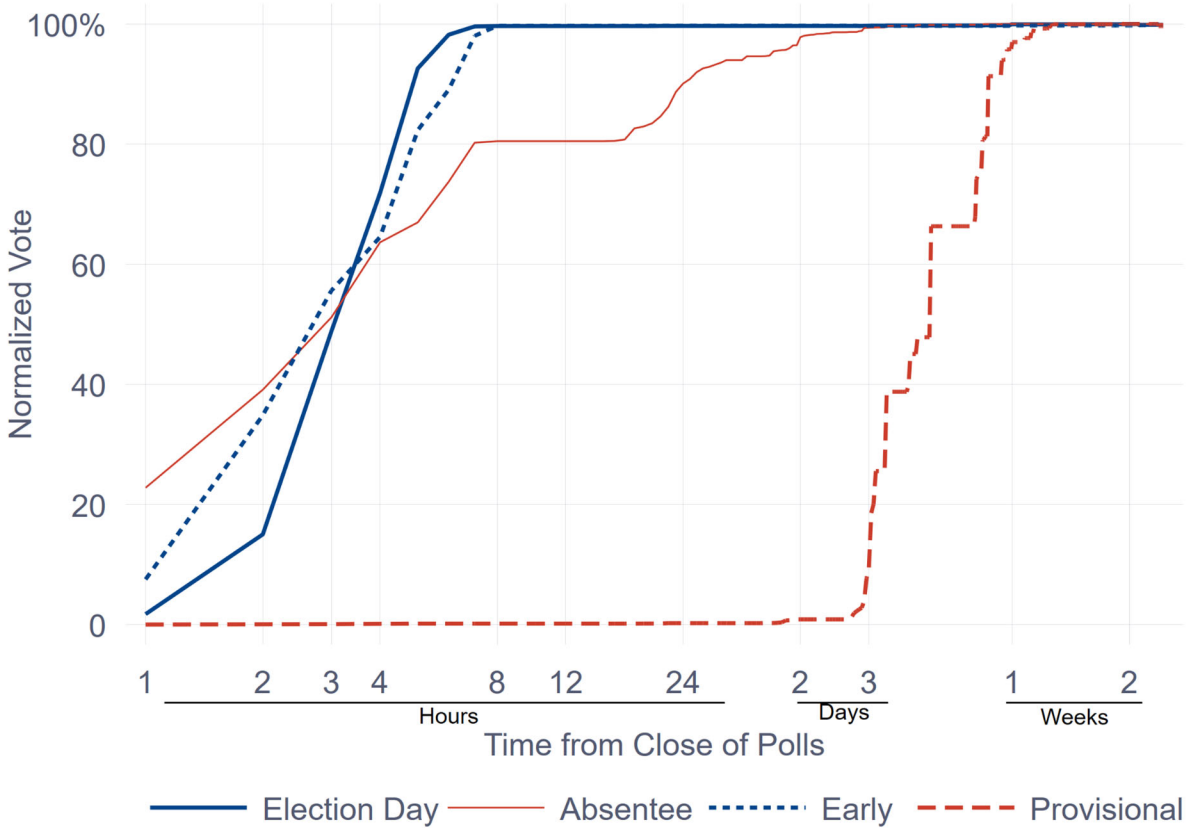
To help put the Georgia analysis in the context of the prior national analysis, we revisit Hypotheses 1 and 2, but only for the state of Georgia. Details of this analysis are in Appendix G. Hypothesis 1 (counties that supported Biden took longer to count votes than counties that supported Trump) is confirmed for Georgia. However, unlike the national pattern, where Biden counties were ahead in the first hour, in Georgia, Trump's counties were ahead of the vote throughout. Hypothesis 2 (counties reported votes for Trump faster than votes for Biden) meets mixed results in Georgia as it did nationwide. In Georgia, votes for Biden outpaced votes for Trump within counties for the first two hours after polls closed. (Nationwide, votes for Biden outpaced votes for Trump for the first three hours.) Therefore, on these two basic patterns, Georgia was similar to the nation as a whole, except that Trump votes started being reported in Georgia a little earlier than the national average.

The one seeming anomaly in testing Hypotheses 1 and 2 with the Georgia data is the fact that Trump's normalized vote was reported more quickly than Biden's from Hour 1. Yet, Biden's average normalized vote *within counties* was ahead of Trump for the first two hours. As shown in Appendix G, this anomaly is resolved by noting that a disproportionate number of absentee ballots in virtually every county was reported in the first couple of hours. However, it is still the case that counties Trump won reported their results more quickly than counties Biden won. Consequently, most Georgia counties experienced a blue shift on election night.

The question now turns to the issue of the pace of counting and reporting *by mode.* To test Hypotheses 3 (Election Day votes should be counted completely before mail votes) and 4

(early votes should be completed before mail ballots but after Election Day ballots), we calculated the normalized vote in Georgia on an hourly basis separately for each voting mode, showing the results graphed in Figure 10. The first pattern to note is that absentee ballots constituted a disproportionate share of election returns in the first hour. This lends credence to our prior speculation that Biden out-performed his eventual vote share in the first hour because of the release of mail ballot tallies, presumably among those that had been preprocessed. However, the rate of change in the various normalized vote series shows that the subsequent counting and reporting of new mail ballots was slower than either early in-person or Election Day votes from Hour 2 onward.

Figure 10. Normalized vote in Georgia, by voting mode.



This tortoise-and-hare dynamic resulted in Election Day votes reaching near-completion first, followed close behind by early voting totals, and lastly by mail ballots. Ninety-nine point five percent of all Election Day votes were counted by the seven-hour mark, compared to the eight-hour mark for early votes and 80 hours for mail ballots. (Ninety-nine and five-tenths percent of provisional ballots were counted by 210 hours—almost nine days—after polls closed.)

Therefore, as stated, Hypotheses 3 and 4 are confirmed. However, we emphasize that they are confirmed because they were stated in terms of completing the count. At the start of the reported count, mail ballots outpaced Election Day votes. However, the slower incremental counting of mail ballots, compared to ballots from the two in-person modes, is what led to the dynamic experienced by viewers of network news shows on election night, whereby once midnight arrived, the vote tally had slowed to a crawl because of the remaining absentee ballots left to be counted and reported.

Discussion

The 2020 general election will go down as one of the most contentious in U.S. history. Part of that contentiousness was because of patterns in the reported election returns on election night and beyond. However, our results suggest that the election returns were reported at a predictable pace, given what is known about administrative practices in tallying ballots and the preferences across partisans in the voting modes they use (at least in 2020).

Prior research that established the blue shift has shown non-random trends in the vote totals that start soon after the polls close until all the votes are counted and reported. However, that research has had to rely on newspaper accounts that picked up the vote-count story on the Thursday following Election Day, at a point where 99 percent of the vote has already been reported. Here, we have been able to start the story at the very beginning, as soon as the polls

closed, and at a much finer level of detail. In doing so, we have shown that the blue shift occurs over a more extended period but that it is more variable than previously revealed. It is more variable along the time dimension, to the degree that the very first reported votes, at least in 2020, were actually "bluer" than the final count, and that the trends previously documented by Foley and Stewart took a few hours to reveal themselves. It is more variable geographically, as well.

Viewed narrowly from the perspective of the hypotheses we explored in this paper, the highly granular data provided by the NEP and the State of Georgia confirm several expectations. We have summarized the results of the hypothesis tests in Table 2. "Bluer" counties reported election results more slowly than "red" counties; within counties, Trump votes were ahead of Biden votes. Election Day reporting finished up faster than the reporting of early votes, which was much faster than absentee ballots. The sheer number of mail ballots states had to handle slowed down the overall vote count, as did prohibiting the preprocessing of mail ballots and allowing absentee ballots to dribble in after Election Day.

Table 2. Summary of Hypotheses Tested

No.	Hypothesis	Data	Result	Exceptions
1	Counties that Biden won took longer to report results than counties that Trump won.	NEP	✓	Counties Biden won reported results more quickly in the first 2 hours
		GA SOS	✓	No
2	Counties reported votes for Trump faster than they reported votes for Biden.	NEP	✓	Counties reported Biden votes more quickly in the first hour
		GA SOS	✓	
3	Completing the count of mail ballots should take longer than Election Day votes.	GA SOS	✓	Counties initially reported mail ballot results more quickly than Election Day or early votes
4	Completing the count of early in-person ballots will take longer than Election Day votes but less time than mail votes.	GA SOS	✓	Early votes were initially reported more quickly than Election Day votes
5	The more absentee ballots cast in a state, the slower the reporting of election results overall.	NEP	✓	No
6	Allowing local jurisdictions to preprocess mail ballots will speed up reporting the vote count.	NEP	✓	No
7	Allowing mail ballots received after Election Day to be counted will slow the reporting of the vote count.	NEP	✓	No

Yet, most of these findings have important exceptions. The most important is that the hypotheses about the speed of counting and the relative balance of Biden and Trump votes are only confirmed once the very first tranches of mail ballots were reported. However, these exceptions prove the rule, in that they hold because of how preprocessing of mail ballots led to an initial surge in reported votes for Biden.

Our results, therefore, provide a starting point for creating a baseline to assess future shifts in post-election night results. Assuming a return to the everyday non-pandemic politics

before 2020, research by Li et al. (2020) in predicting who makes use of provisional ballots and where they vote will be of great importance in predicting the minute shifts following election night. Insofar as alternatives to in-person voting continue, our results suggest that the differential pace of tabulating results across counties will continue to be a concern.

Nonetheless, it is possible to avoid the types of delays and protracted partisan shifts that occurred in several states because of limitations about when mail ballots could be processed and tabulated. The two policy choices that sped up the count and condensed the period in which partisan trends appeared were (1) requiring mail ballots to be returned by Election Day and (2) allowing counties to begin processing early and mail ballots before Election Day. Florida, which has long been the whipping boy of election administration in the U.S., showed the nation what is possible when both early and mail ballots are tabulated on an ongoing basis as they arrive: nearly all its votes were reported within two hours of polls closing.

Our results also provide an innocent explanation to patterns of election reporting that were used to stir up controversy through claims of vote fraud. The differences that arose on post-election-night reporting speeds were driven by the fact that smaller and more rural counties, which favored Trump, could report their ballots before the counties with hundreds of precincts and hundreds of thousands of voters. The results of the 2020 election are surprising only to those who ignore the basics of where voters live.

Our results further speak to how poorly the pattern of election night results fit the assumptions of parallel vote trend (PVT) analyses; those seeking to use the technique in the U.S. should look elsewhere. Partisan time trends in the reported vote are readily explained by variation in election laws across states and counting demands within states. Absent credible physical evidence that bad actors, either election administrators or political operatives, have

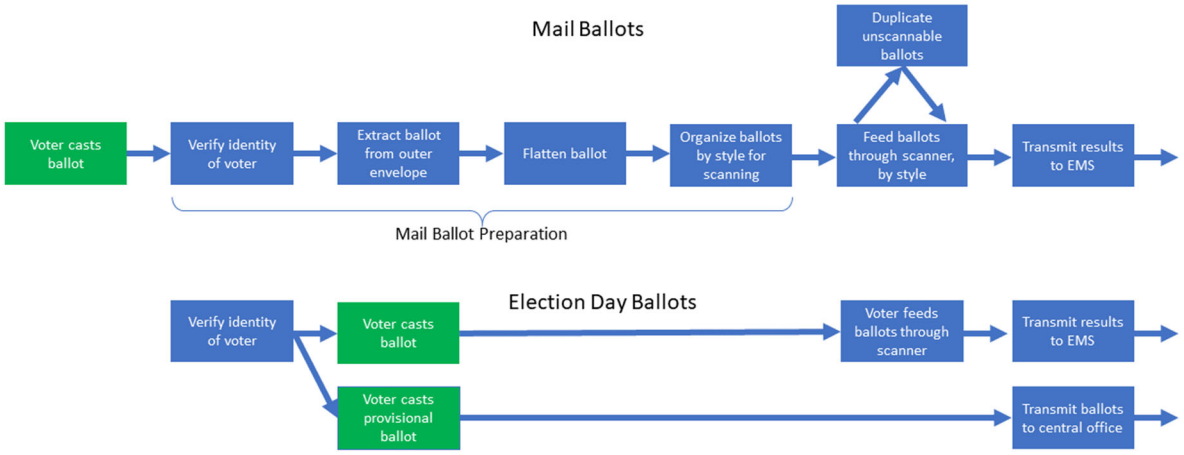
tampered with the collection and counting of ballots, we have to assume that partisan time trends in the reported vote count are because of policy choices and the administrative practices that flow from them.

Appendix A

Comparison of Absentee and Election Day Ballot Counting Workflows

In Figure A1, we have summarized the workflow of tabulating mail ballots and compared it to tabulating precinct-counted Election Day votes. Starting with Election Day votes, the voter arrives at the polling places and verifies their identity. They are then given a ballot, which is filled out and scanned by the voter in the process of depositing it into the ballot box. Once the polls close, the memory card is removed from the scanner, taken to the central election office, where the card contents are read and included in the jurisdiction's overall vote tally. Provisional ballots are held at the precinct and then transmitted to the central office for adjudication and potential counting.

Figure A1: Comparison of Workflow of Processing Mail Ballots and Precinct-Counted Election Day Ballots.



Mail balloting is more administratively involved once the voter has cast and returned the ballot (Kerevel and Atkeson, 2012; Harper et al., 2020). The voter's identity must be established before the outer envelope is opened and the ballot is extracted from the envelope. Once removed from the envelope, the ballot must be flattened and otherwise prepared to be fed through a scanner, which might be a high-speed "batch" scanner or a precinct scanner fed one ballot at a time. Unless the ballots have been encoded with the ballot style and the scanner can read multiple ballot styles, the ballots must be separated into batches according to ballot style. If they do not need to be divided into batches according to style, they still need to be placed in stacks in preparation for scanning. As the ballots are scanned, some may be unscannable — they may be torn or smudged, for instance. These ballots are "duplicated" by hand onto blank ballots and rescanned. At regular intervals, the memory cards of the scanners are removed, and the contents are read into the election night reporting system for accumulation into a jurisdiction-wide total.

Note that the process for Election Day ballots occurs entirely on Election Day, except for the adjudication of provisional ballots. Also, note that once the central election night reporting unit reads in the memory card from the precinct, virtually all that precinct's Election Day votes

are accounted for. The only exception is to account for provisional ballots added within a few days and any mistakes that might need to be corrected.

In contrast, the process of mail ballots is more involved and does not necessarily happen on Election Day and may not happen on a single day at all. The four middle steps are shown in Figure 1, which can be considered ballot preparation, do not have Election Day voting analogs. States have exercised discretion in deciding how much ballot preparation can occur before Election Day. They have also exercised discretion in determining whether mail ballots can be scanned before Election Day, in addition to whether mail ballots will be accepted for counting if they arrive after Election Day.

The decision to allow processing, and especially counting, of mail ballots before Election Day can speed things up, even to the point where mail-ballot results might be reported more quickly than Election Day votes. Allowing ballots to arrive after Election Day can delay the final resolution of the count, although in 2020, the number of ballots that arrived after Election Day in states that allowed it was minor.

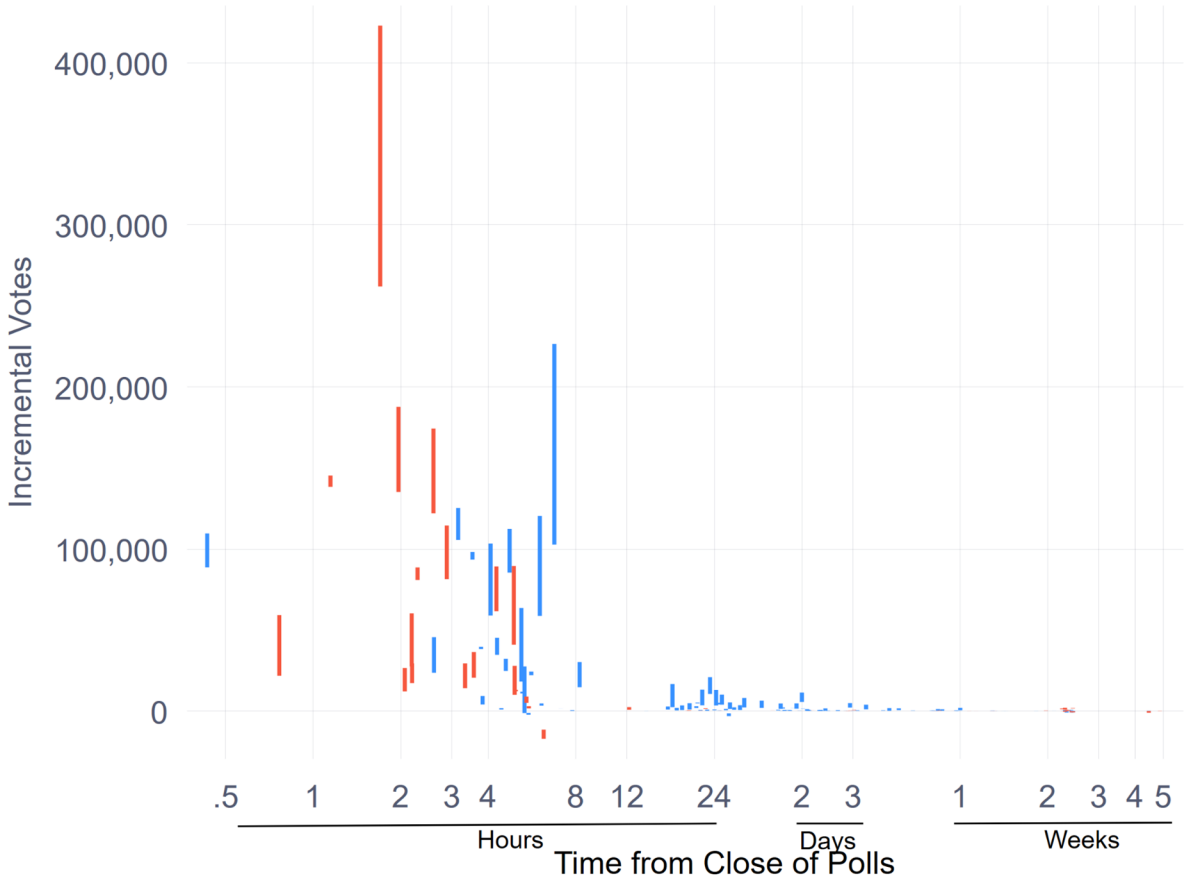
Appendix B

NEP Data Reports Using Georgia as an Example

As an example of what the data look like, Figure B1 shows when unique election return reports from Georgia were issued by the *New York Times.* Because the most rapid release of results occurred within hours of the polls closing — over half of all votes were reported within three hours of the polls closing — but the *Times* continued to update results for another five weeks, we have displayed the time dimension on the *x*-axis using a log scale.

Figure B1 shows the release of new election results over this period, including the total number of new votes reported for Biden and Trump. Each vertical line indicates a new election result report. Blue lines indicate reports with more votes for Biden than for Trump; red lines indicate the opposite. Finally, when the line is blue (red), the top of the line is the number of new votes reported for Biden (Trump), while the bottom of the line is the number of new votes reported for Trump (Biden). For instance, the very first report, which was at 7:26:30 p.m. on election night (26 minutes — 0.43 hours — after the polls closed), gave Biden 109,587 votes and Trump 88,626. Thus, the vertical bar is blue. The second report, at 7:46:10 p.m., gave Biden an additional 21,774 votes and Trump 59,213. Thus, the second vertical line is red.

Figure B1. Example of election return reporting data for Georgia.

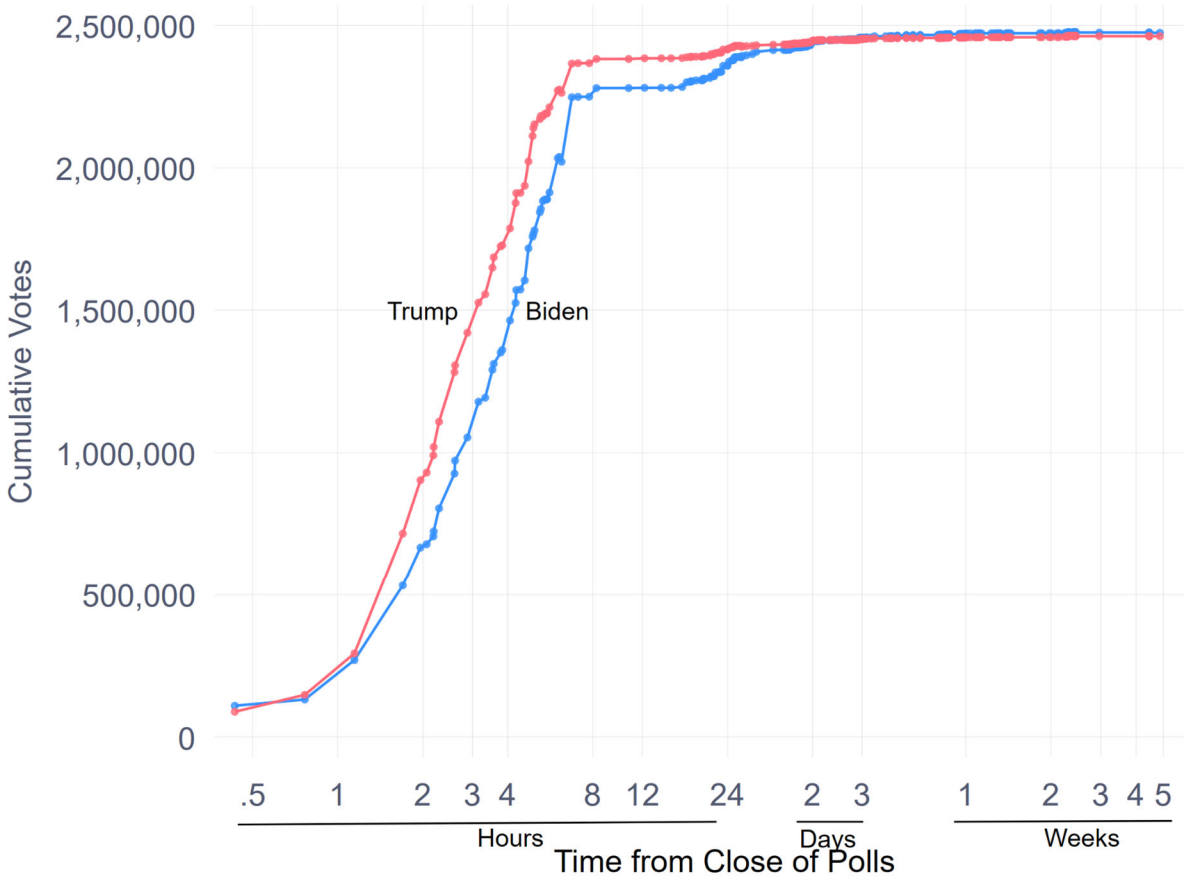


Data source: NEP via the New York Times

Note: Vertical bars indicate moments when election results were reported. A red bar indicates that more votes were reported for Trump than for Biden at that moment; a blue bar indicates more votes were reported for Biden. For red bars, the top of the bar indicates votes reported for Trump; the bottom indicates votes reported for Biden. The opposite is true for blue bars.

Figure B2 shows the same data differently, but accumulating the votes for Biden and Trump as they are reporting. Here we see that even though the very first report put Biden in the lead, starting with the second report, Trump had accumulated more votes than Biden and remained in the lead until the report issued by the *New York Times* at 4:47:31 a.m. on Friday morning, November 6. From that point forward, Biden led Trump in the state.

Figure B2. Cumulative votes for Trump and Biden in Georgia.



Data source: NEP via the New York Times.

Appendix C

Figures 1 and 2 with Actual Reporting Times

Figures 1 and 2 in the text record normalized votes reported for each state, presenting snapshots at hourly intervals. Figures C1 and C2 show the same data, displaying the actual reporting times. The same hourly national aggregation is also displayed.

Figure C1. Normalized votes reported for each state, as votes were reported.

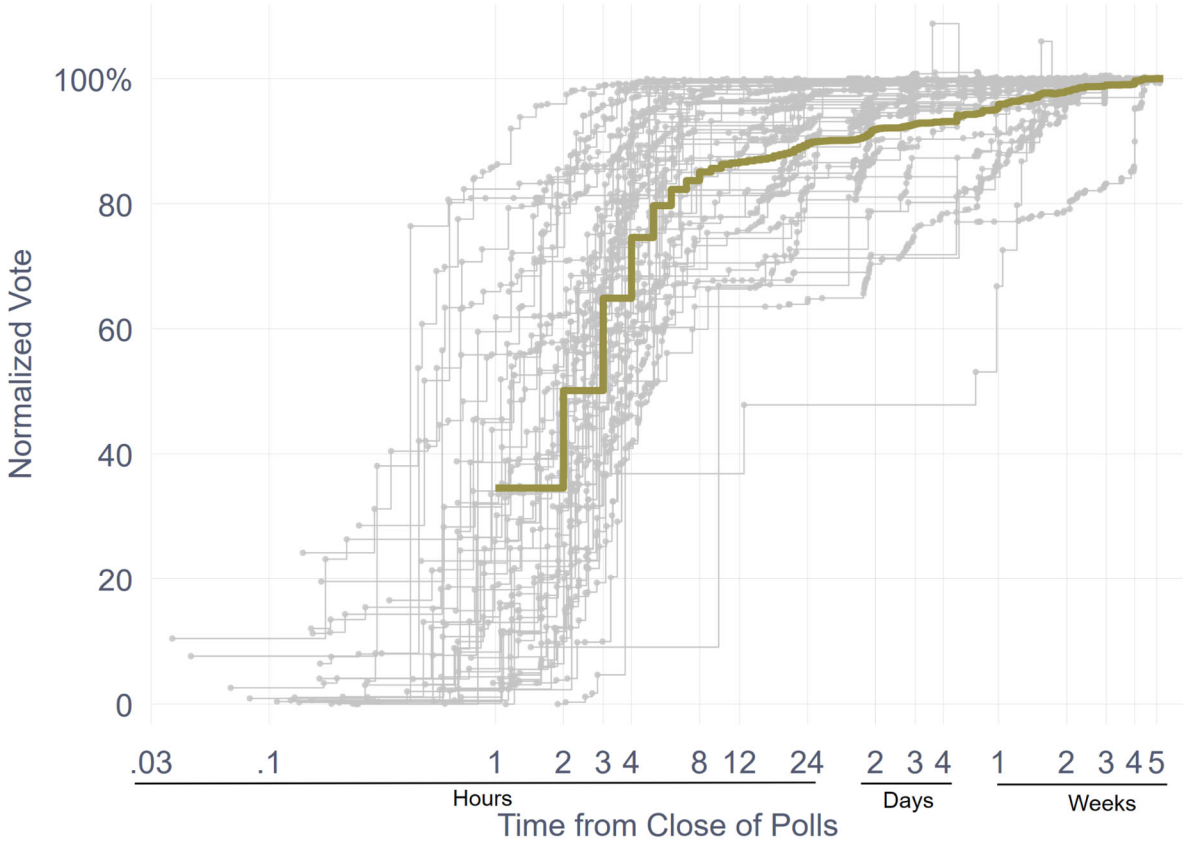
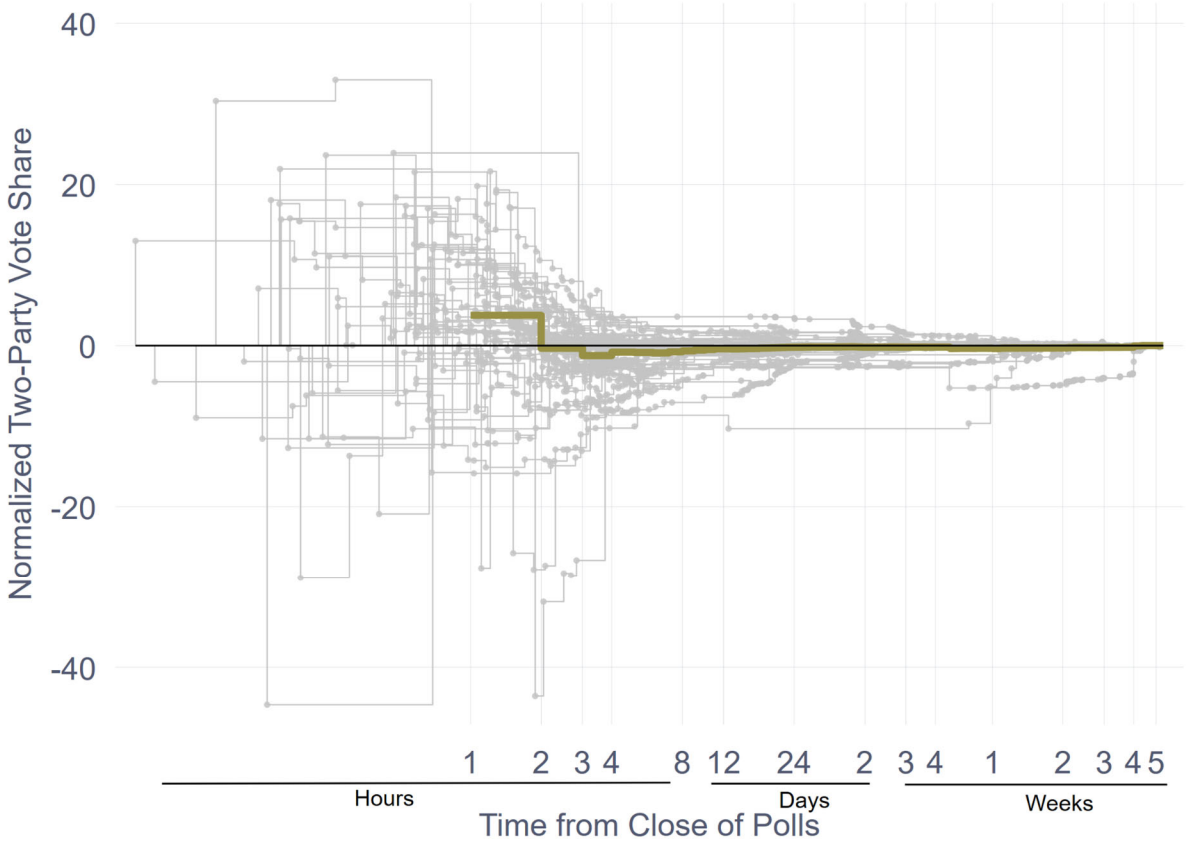


Figure C2. Normalized two-party vote share in each state, as votes were reported.



Data source: NEP via the New York Times.

Appendix D

Alternative Specifications of Table 1 Regressions

Table D1. Dummy variable indicating states allowing late-returned absentee ballots to be counted. (Robust standard errors. Alaska omitted)

	4 Hours	8 Hours	12 Hours	24 Hours	48 Hours
Mail pct.	-0.127 (0.075)	-0.194*** (0.047)	-0.183*** (0.050)	-0.145** (0.045)	-0.107** (0.037)
No preprocessing	-12.9** (4.7)	-8.84** (2.88)	-7.72** (2.87)	-2.94 (2.13)	-1.46 (1.64)
Return deadline	0.174 (4.32)	-6.98* (2.78)	-7.74** (2.82)	-6.47** (2.26)	-5.88** (1.81)
Intercept	87.2*** (5.24)	102.9*** (2.37)	103.1*** (2.42)	103.5*** (1.78)	103.2*** (1.42)
N	49	50	50	50	50
RMSE	15.2	9.31	9.42	7.44	5.84
R^2	.155	.380	.364	.340	.355

***$p < .001$ **$p < .01$ * $p < .05$

Table D2. Inclusion of Alaska in Regressions. (Robust standard errors.)

	4 Hours	8 Hours	12 Hours	24 Hours	48 Hours
Mail pct.	-0.281 (0.155)	-0.186* (0.079)	-0.122* (0.053)	-0.102* (0.048)	-0.054 (0.043)
No preprocessing	-16.2** (5.84)	-13.5* (5.33)	-7.02* (3.06)	-4.44 (2.32)	-2.04 (1.91)
Return deadline	-0.231 (0.675)	-1.68** (0.52)	-1.48*** (0.404)	-1.55*** (0.33)	-1.47*** (0.33)
Intercept	89.0*** (6.38)	103.3*** (4.66)	100.2*** (2.67)	102.2*** (2.25)	101.4*** (1.97)
N	51	51	51	51	51
RMSE	20.2	14.0	10.8	8.6	7.7
R^2	.190	.393	.386	.483	.480

Table D3. Quantile regression. (Robust standard errors, Alaska omitted)

	4 Hours	8 Hours	12 Hours	24 Hours	48 Hours
Mail pct.	-0.290 (0.155)	-0.128 (0.080)	-0.130* (0.054)	-0.083 (0.056)	-0.044 (0.031)
No preprocessing	-16.2 (8.2)	-9.81* (4.24)	-7.60* (2.88)	-3.21 (2.96)	-0.535 (1.661)
Return deadline	0.028 (0.873)	-1.35** (0.45)	-1.36*** (0.306)	-1.50*** (0.314)	-1.27*** (0.18)
Intercept	93.7*** (8.55)	101.0*** (4.42)	101.4*** (2.99)	101.4*** (3.08)	100.6*** (1.73)
N	50	50	50	50	50
Pseudo R^2	.141	.280	.349	.322	.342

Appendix E.

Figure 6 with County Data

Figure E4. Counties won by Biden. (Data for counties won by Biden)

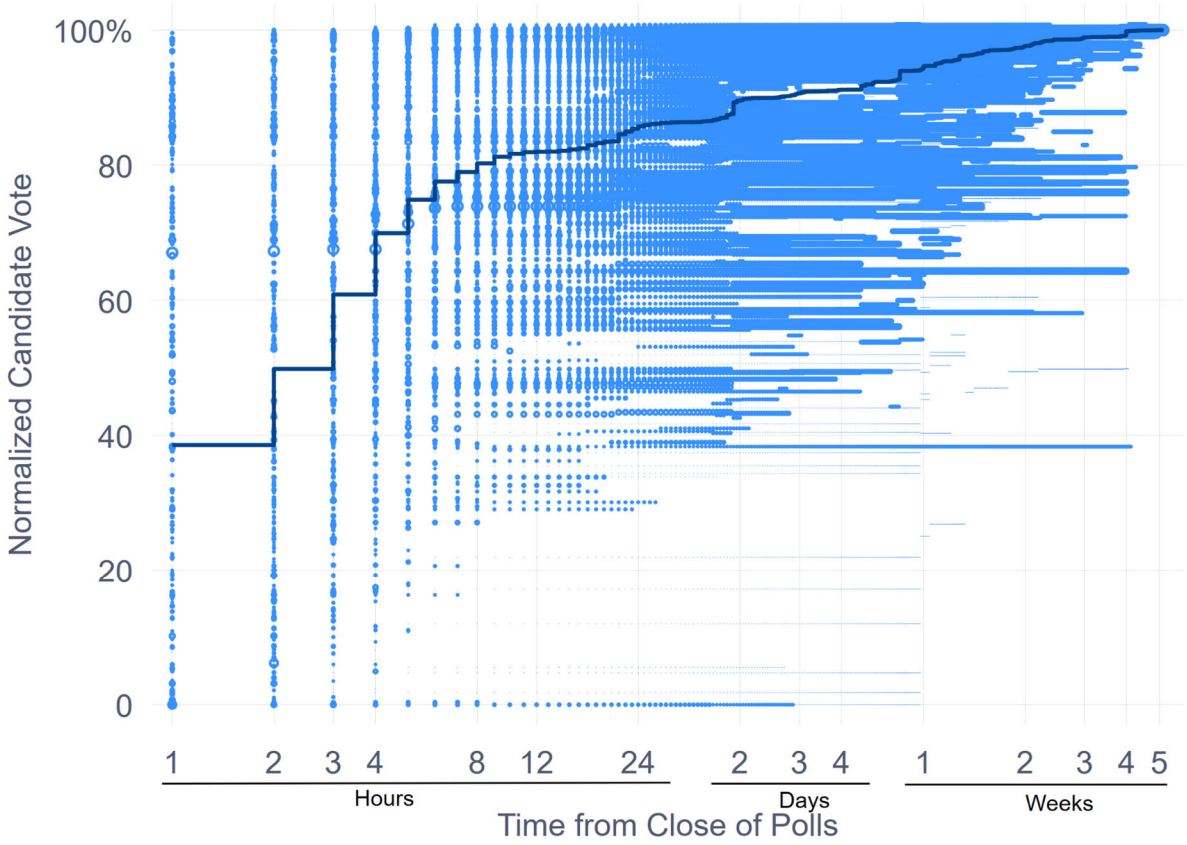
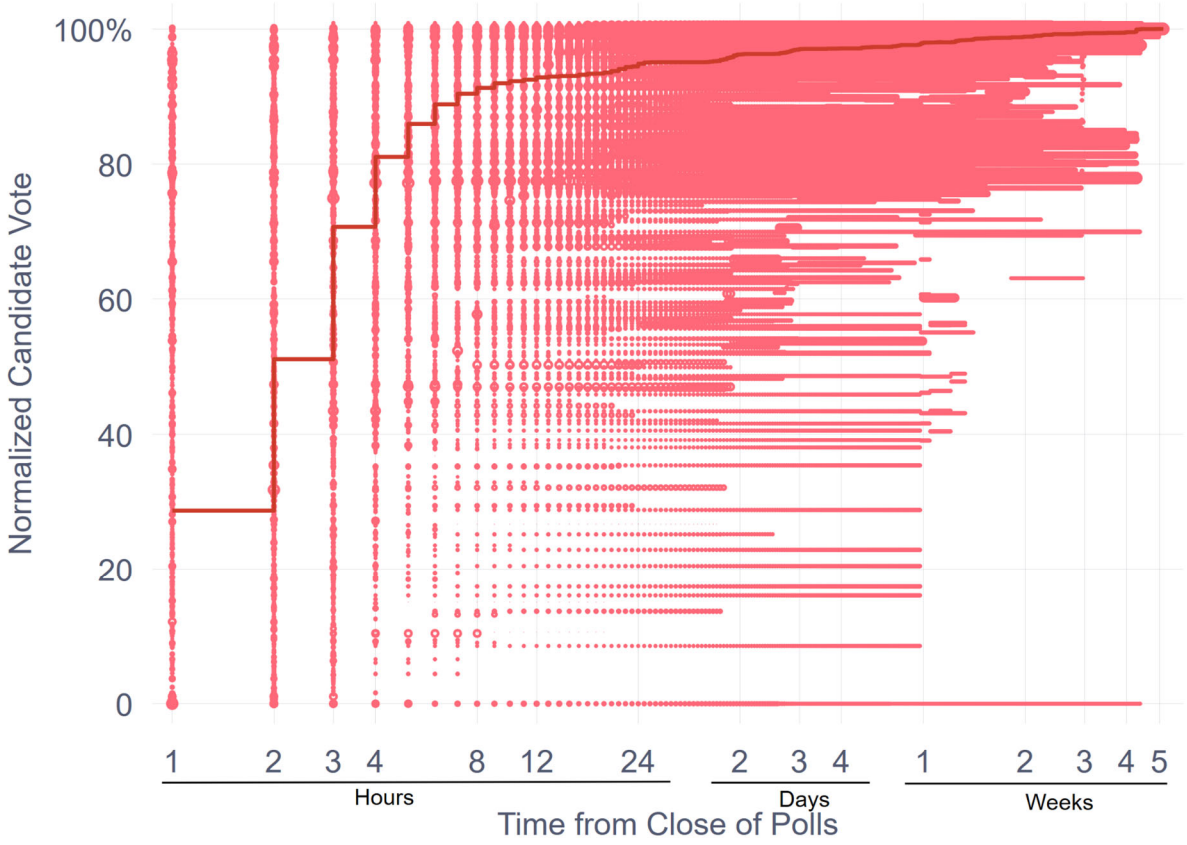


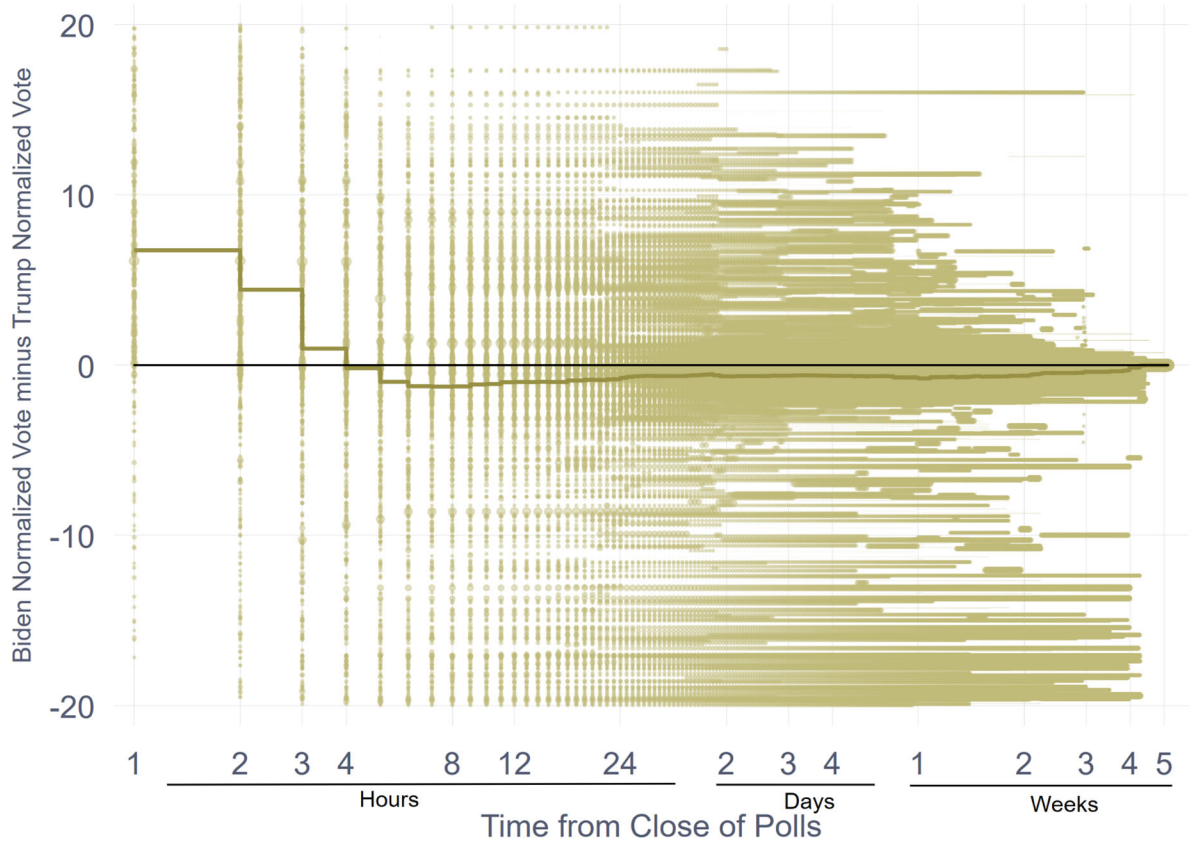
Figure E2. Counties won by Trump. (Data for counties won by Trump shown in red)



Appendix F.

Figure 8 with County Data

Figure F1. Average difference in reporting pace of Biden and Trump votes, by time since polls closed.



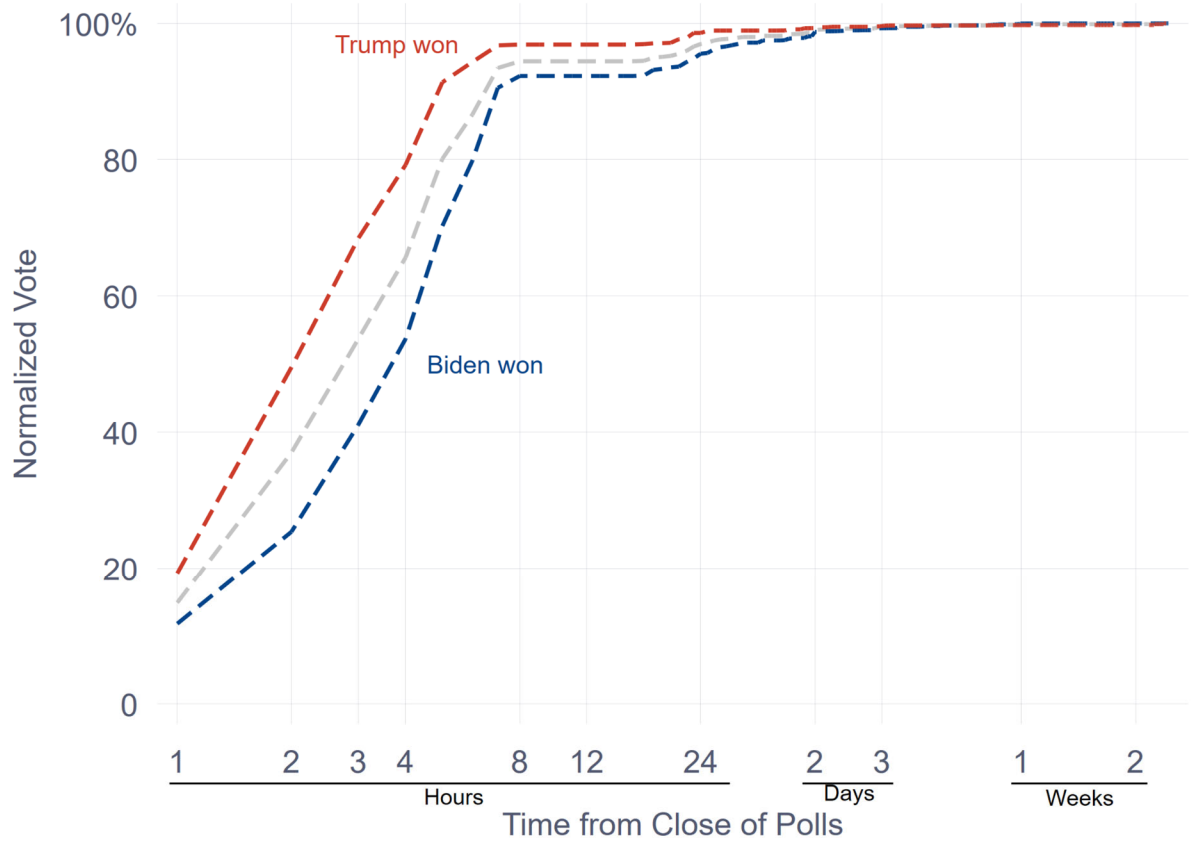
Note: The line connects hourly weighted averages. (Weights are the total number of votes reported by the county.) Data tokens are proportional in size to the number of votes reported by the county. Only data points within ± 20 are displayed. For Hour 1, this includes 89 percent of all observations. Starting at hour 21, it includes 95 percent, at hour 662 (3 weeks), 99 percent.

Appendix G.

Testing Hypotheses 1 and 2 for Georgia

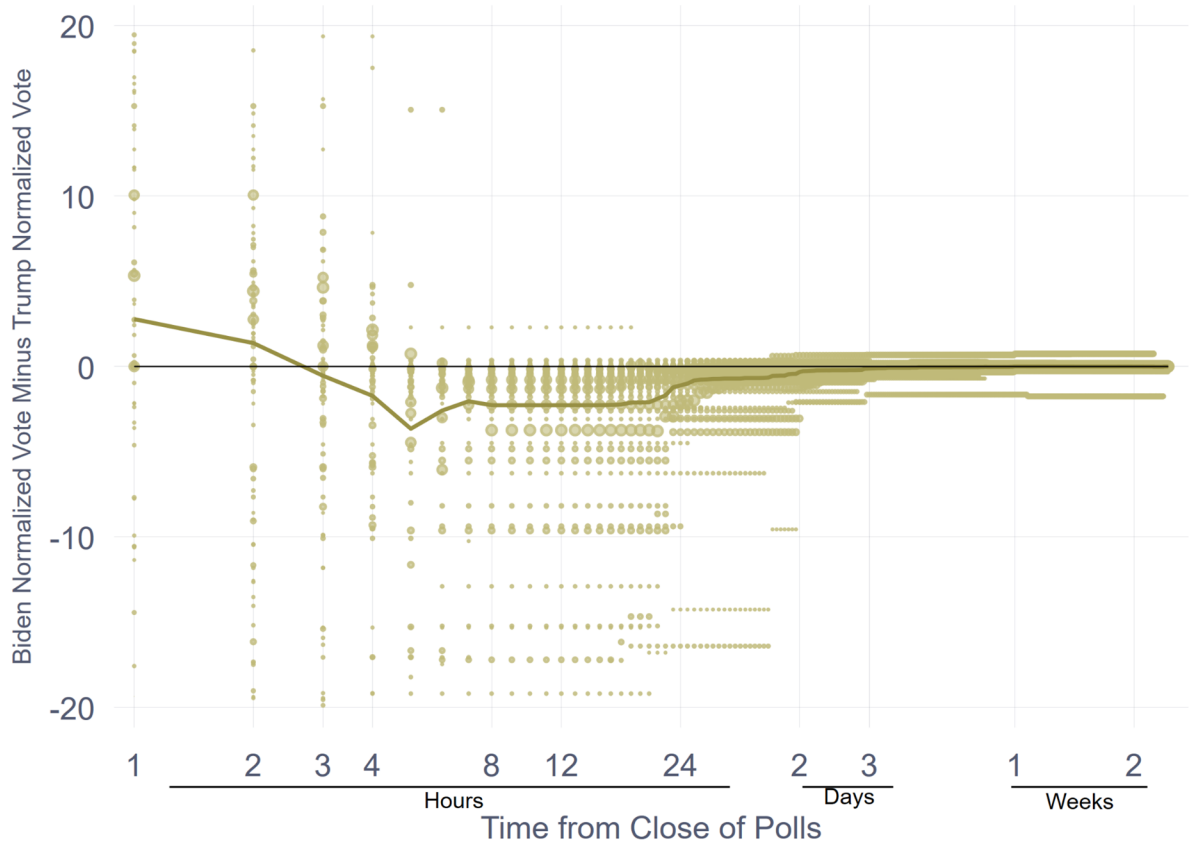
To help put the Georgia analysis in the context of the prior national analysis, we revisit Hypotheses 1 and 2, but only for the state of Georgia. We test Hypothesis 1 (that counties won by Trump reported more quickly than counties won by Biden) as we did with the NEP data. Figure G1 plots the hourly weighted averages of the normalized vote for counties won by Biden and Trump. The grey line shows the statewide average.

Figure G1. Percentage of Georgia County Votes Reported, by Time Since Polls Closed. Counties grouped by whether Biden or Trump won them.



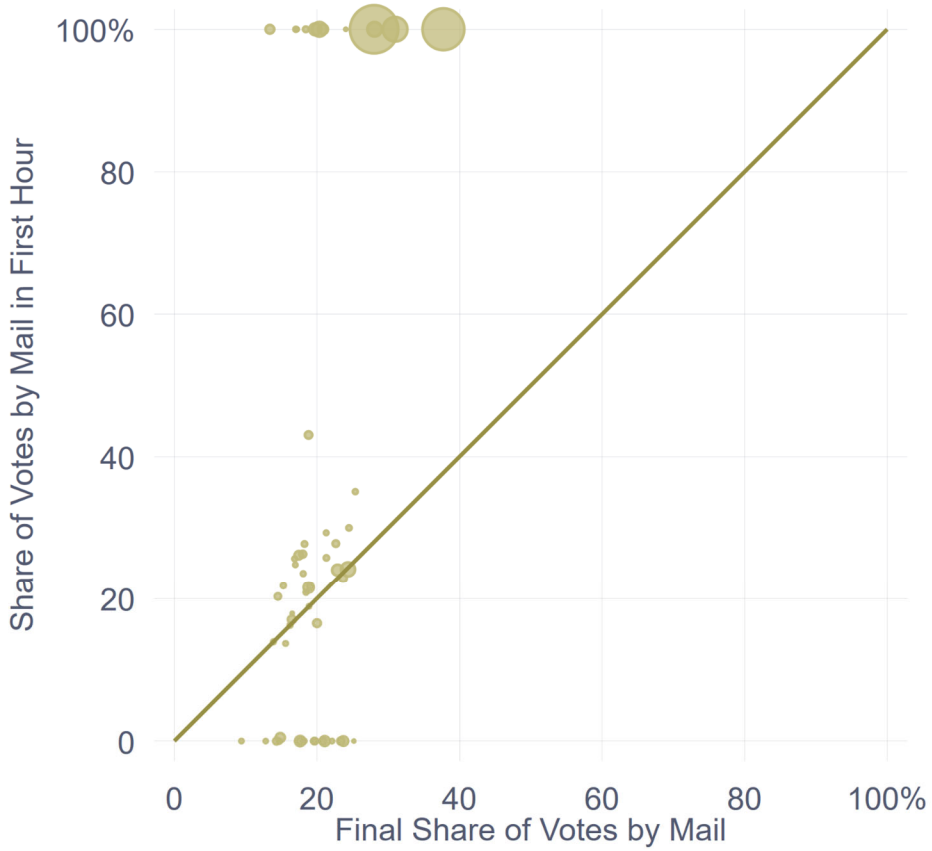
Hypothesis 2 is that *within counties*, Trump votes were reported more quickly than Biden votes. As before, we tested this hypothesis by subtracting the normalized Trump vote from the normalized Biden vote for each county in each hour.

Figure G2. Average Difference in Normalized Candidate Vote by Hour in Georgia.



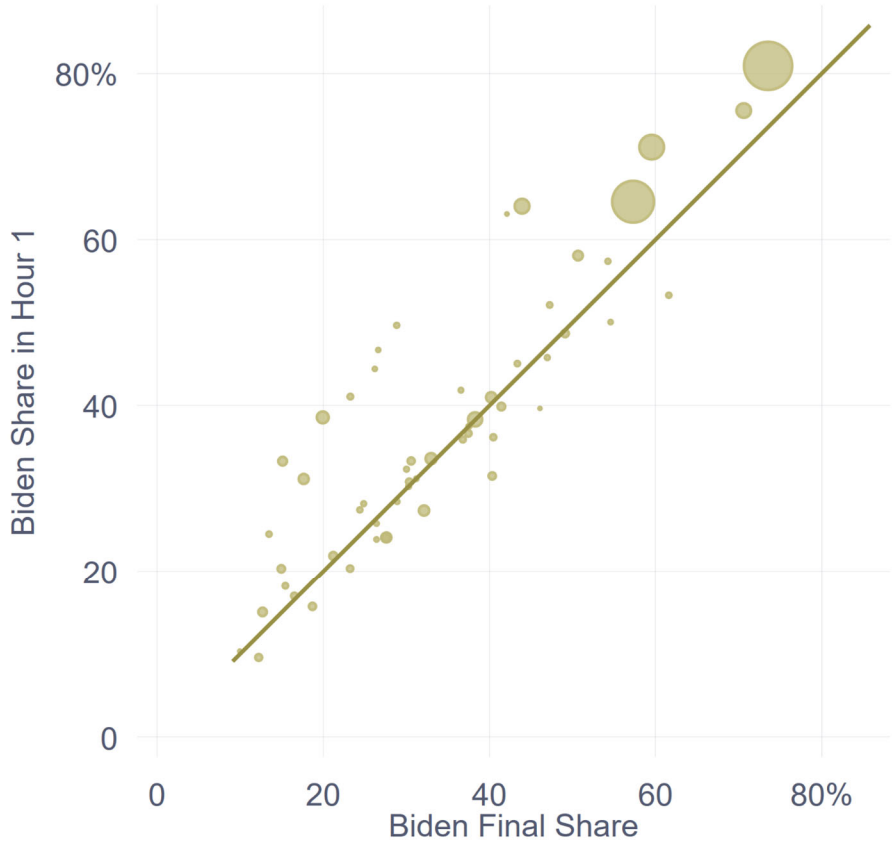
The finding that in Georgia, Trump's normalized vote outpaced Biden's from the start, and yet average Biden's normalized vote *within counties* outpaced Trump's may initially seem anomalous. The resolution to this seeming paradox starts with seeing that in 34 of the 55 counties that reported at least some votes in the first hour, a disproportionate share of those votes came from mail ballots compared to Election Day and early votes. This is illustrated in Figure G3, which plots the percentage of first-hour votes in each county attributable to mail ballots on the *y*-axis against the county's final share of mail ballots. Figure G3 makes clear that the counties that reported in Hour 1 fell into three categories, according to the fraction of mail ballots that were included in their count: (1) only mail ballots (12 counties), (2) no mail ballots (15 counties), and (3) a mix of mail and in-person ballots (28 counties). Among those reporting a mix of mail and in-person ballots, most (22 of 28) reported a greater share of mail ballots in the first hour and the final count. Among counting reporting only mail ballots, three were among Georgia's largest and most pro-Biden counties: Fulton, Cobb, and Chatham.

Figure G3. Share of Mail Ballots in First-hour Georgia County Reports Plotted against Share of Mail Ballots in the Final Report.



Consequently, most counties reporting in the first hour (34 of 55) showed a two-party vote share that exceeded the county's final count, as illustrated in Figure G4.

Figure G4. Biden Two-Party Vote Share in Georgia Counties in the First Hour Compared to the Final Count.



References

Bjornlund, Eric. 2011. "Vote Count Verification: A User's Guide for Funders, Implementers, and Stakeholders." Democracy International, report prepared for the United States Agency of International Development.

Bjornlund, Eric. "Election 2020: America Votes." *Foreign Policy*, November 10, 2020. https://foreignpolicy.com/projects/america-votes-2020-election-live/ (accessed March 3, 2021).

Borneman, Walter R. 2008. *Polk: The Man Who Transformed the Presidency and America.* New York. Random House.

Bronner, Laura, Chadwick Matlin and Michael Tabb. "Red And Blue Shifts Happen. That's Not Evidence Of Fraud." *Five Thirty-Eight,* November 4, 2020. https://fivethirtyeight.com/videos/vote-counts-swing-from-one-party-to-another-thats-not-evidence-of-fraud/ (accessed March 2, 2020).

Burns, Alexander and Jonathan Martin. "As America Awaits a Winner, Trump Falsely Claims He Prevailed." *New York Times*, November 4, 2020. https://www.nytimes.com/2020/11/04/us/politics/election-trump-biden-recap.html.

Chervinsky, Lindsay. "History Teaches Us Election Delays Are Nothing New." *Governing*, October 28, 2020. https://www.governing.com/now/History-Teaches-Us-Election-Delays-Are-Nothing-New.html (accessed March 3, 2021).

Curiel, John A. and Emily Kohn, "North Carolina Mail Ballot Trends," Stanford-MIT Healthy Elections Project, November 2, 2020. https://healthyelections.org/sites/default/files/2020-11/nc_report_2020-110220.pdf (accessed March 15, 2021).

Curiel, John A. and Charles Stewart III. "Biden's win shows that suburbs are the new swing constituency." *Washington Post* the Monkey Cage Blog. November 24, 2020. https://www.washingtonpost.com/politics/2020/11/24/bidens-win-shows-that-suburbs-are-new-swing-constituency/ (accessed November 25, 2020).

Daxecker, Ursula E. 2012. "The Cost of Exposing Cheating: International Election Monitoring, Fraud, and Post-Election Violence in Africa." *Journal of Peace Research* 49(4): 503–516.

Foley, Edward B. 2013. "A Big Blue Shift: Measuring an Asymmetrically Increasing Margin of Litigation," *Journal of Law and Politics* 28: 501 - 546.

Foley, Edward B. 2019. "Preparing for a Disputed Presidential Election: An Exercise in Election Risk Assessment and Management." *Loyola University Chicago Law Journal* 51(2): 309-362.

Foley, Edward B. and Charles Stewart III. 2015. "Explaining the Blue Shift in Election Canvassing." MIT Political Science Department Research Paper No. 2015-21.URL: https://ssrn.com/abstract=2653456

Foley, Edward B. and Charles Stewart III. 2020. "Explaining the Blue Shift in Election Canvassing." *Journal of Political Institutions and Political Economy* 1(2): 239 - 265.

Garber, Larry and Glenn Cowan. 1993. "The Virtues of Parallel Vote Tabulation." *Journal of Democracy* 4(2): 95 - 107.

Graff, Garrett M. "A Day-By-Day Guide to What Could Happen If This Election Goes Bad." *Politico*, October 23, 2020. https://www.politico.com/news/magazine/2020/10/23/election-nightmares-guide-430915 (accessed March 8, 2020).

Harper, Tim, Rachel Orey, and Collier Fernekes. 2020. "Counting the Vote During the 2020 Election." Bipartisan Policy Center, August 2020. https://bipartisanpolicy.org/wp-content/uploads/2020/08/Counting-the-Vote-During-the-2020-Election-Bipartisan-Policy-Center-August-2020.pdf (accessed March 1, 2021).

Hyde, Susan. 2008. "How International Election Observers Detect and Deter Fraud," in *Election Fraud: Detecting and Deterring Electoral Manipulation*. Eds. R. Michael Alvarez, Thad E. Hall, and Susan D. Hyde. 201 - 215.

Idrobo, Nicolás, Dorothy Kronick and Francisco Rodríguez. 2020. "Do Shifts in Late-Counted Votes Signal Fraud? Evidence From Bolivia." Available at SSRN: https://ssrn.com/abstract=3621475

Chris Kahn and Jason Lang. 2020. Explainer: Red mirage, blue mirage—Beware of early U.S. election wins. *Reuters*, October 22, 2020, https://www.reuters.com/article/us-usa-election-mirage-explainer-idUSKBN2771CL

Kerevel, Yann and Lonna Rae Atkeson. 2012. "Counting the Ballots: A Comparison of Machine and Hand Counts in New Mexico." in, *Confirming Elections: Creating Confidence and Integrity through Election Auditing*, ed. R. Michael Alvarez, Lonna Rae Atkeson, and Thad E. Hall. New York: Palgrave Macmillan.

Li, Yimeng, Michelle Hyun, and R. Michael Alvarez. 2020. "Why Do Election Results Change After Election Day? The 'Blue Shift' in California Elections." *APSA Preprints*. DOI: 10.33774/apsa-2020-s43xx.

Mitchell, Robert. "A presidential election history lesson: Americans often waited days or weeks for the outcome." *Washington Post,* November 4, 2020. https://www.washingtonpost.com/history/2020/11/03/history-election-night-wait-results/ (accessed March 3, 2021).

Neiheisel, Jacob R. and Barry C. Burden. 2012. 'The Impact of Election Day Registration on Voter Turnout and Election Outcomes,' American Politics Research, 40(4): 636–664.

Pavia, Jose M. 2010. "Improving predictive accuracy of exit polls." *International Journal of Forecasting* 26(1): 38 - 56.

Pavia, Jose M. 2005. "Forecasts from non-random samples: The election night case." *Journal of the American Statistical Association* 100: 1113–1122.

Pavia, Jose M., B. Larraz, and J.M. Montero. 2008. "Election forecasts using spatio-temporal models. *Journal of the American Statistical Association* 103: 1050–1059.

Pavia, Jose M., and B. Larraz-Iribas. 2008. Quick counts from non-selected polling stations. *Journal of Applied Statistics* 35: 383–405.

Pettigrew, Stephen and Charles Stewart III. 2020. "Protecting the Perilous Path of Election Returns: From the Precinct to the News." *Ohio State Technology Law Journal* 16(2): 587 - 638.

Smidt, Hannah. 2016. "From a Perpetrator's Perspective: International Election Observers and Post-Electoral Violence." *Criminal Justice Policy Review* 53(2): 255–278.

Stewart, Charles III and Stephen Ansolabehere. 2015. "Waiting to vote." *Election Law Journal* 14(1):47–53.

Talev, Margaret. 2020. "Exclusive: Dem Group Warns of Apparent Trump Election Day Landslide." September 1, 2020, *Axios.com,* https://www.axios.com/bloomberg-group-trump-election-night-scenarios-a554e8f5-9702-437e-ae75-d2be478d42bb.html.

Wasserman, David. 2020. “Beware the ‘Blue Mirage’ and ‘Red Mirage’ on Election Night.” NBC News. November 3, 2020, https://www.nbcnews.com/politics/2020-election/beware-blue-mirage-red-mirage-election-night-n1245925.