

**Fallacies in Statistically-Based Claims about Massive
Election Fraud in 2020:
A Compendium***

Bernard Grofman

University of California, Irvine

bgrofman@uci.edu

<https://orcid.org/0000-0002-2801-3351>

Jonathan Cervas

Carnegie Mellon University

cervas@cmu.edu

<https://orcid.org/0000-0001-9686-6308>

9/3/22 1:36 PM

*This research was partially supported by the Peltason Chair of Democracy Studies, University of California, Irvine. The views expressed are solely those of the authors. The authors thank Sean Birch for his helpful assistance.

ABSTRACT

An immense amount has been written about alleged massive electoral fraud in the 2020 presidential election: In 2022, tens of millions of voters, and many Republican officials, still believe the 2020 election was stolen. Here, we examine claims of fraud in the presidential election of 2020 that are based on aggregate election data, and we further limit ourselves to claims in which the data itself is essentially undisputed. We show that the implications of that data for the presence of massive fraud have been largely or entirely misinterpreted through invalid statistical or logical reasoning. Our goal is not to provide new insights, since virtually all the points we make have been made by others, but rather to put together in one place a compendium of recent glaring misuses of statistical inference in a form that we believe will be useful as a teaching tool about the dangers of sloppy use of statistics. We discuss the fallacies in a non-technical way to make our critiques broadly accessible to non-specialist audiences.

I. Introduction

After the 2020 presidential election, the losing candidate, Donald Trump, claimed that he had been the victim of massive voter fraud that denied him the election. At the time, many Republican congress members, state attorneys general, and other Republican office holders endorsed this claim -- a claim which his supporters continue to reiterate. Tens of millions of voters, including a clear majority of Republicans and non-trivial numbers of independents and Democrats currently believe that there was massive fraud in 2020 (Gardner, 2021; Solender, 2020; UMass Amherst, 2021). Others believe that most elections are rigged.¹ In the 2022 primaries, some Republican candidates made the assertion of massive election fraud in 2020 a fundamental part of their election platform. Many of those candidates won their primary, and several will be able to affect future election administration in their states if they win in November (Medina et al., 2022). Thus, not only have claims about 2020 election fraud remained politically salient, but the persistence of voter beliefs in these claims, and the deep partisan divide about them leading to affective polarization, have frightening implications for the prospects of democratic breakdown in the U.S. (Grofman, 2022; Homans & Peterson, 2022; Leonhardt, 2022).²

The supposed evidence supporting massive election fraud comes in many forms, including personal affidavits alleging fraud in particular precincts (McClallen, 2021), to

¹ For fraud claims about earlier elections see e.g., Zorn (2017).

² For example, post-election audits that take place outside the regular legal framework for election certification have been demanded in contexts where they make no sense (Bump, 2022a).

videos allegedly showing direct evidence of vote tampering by poll workers (Gray, 2020), to how-to-videos showing the supposed ease for manipulating the record of votes produced by voting machines or mail ballots, to claims about a conspiracy by a particular voting machine vendor (Sganga, 2022), to assertions that more voters voted than were on the jurisdiction's electoral roll (Ayyadurai, 2020; Swenson, 2020). Despite the implausibility of a massive multi-state conspiracy, the volume and variety of claims make them like a hydra-headed monster almost impossible to successfully rebut all of them to a given voter's satisfaction. Moreover, the fact that many of these claims about massive fraud in 2020 (including most of those we discuss in this essay) are plausible on their face, even though fallacious, make them harder to refute. Also, we have the mesmerizing power of repetition. The claim of massive fraud in 2020 is stated again and again in conservative media sources and by former President Trump and his allies.³ Particular factual claims are often repeated even after clear contrary evidence has been presented (Hsu & Thompson, 2022).

There are many reasons that can be offered about why beliefs about massive election fraud in 2020 persists (see e.g., (Berlinski et al., 2021; Douglas et al., 2019; Edsall, 2022; Holman & Lay, 2018)⁴. But exploring why voters believe what they do

³ We would also note that, in the four years before the election, the claim that the 2020 election would be stolen by the Democrats was also repeatedly asserted by President Trump and his supporters, thus preparing the way for the post-election claims of fraud.

⁴ One obvious key factor is the level of present-day polarization, in which partisan identities shape beliefs. Strongly embedded partisan identities mean that the public polarizes on the fraud claim based on partisanship. Relatedly, we have a siloization

(Bump, 2022b) is not the purpose of this essay.⁵ Moreover, in this essay we will not discuss the vast bulk of claims about fraud in 2020, namely those that rest on contested facts. Our concern here is a narrowly focused one. We deal solely with claims about fraud that are grounded, at least in part, on indisputable facts about statistical features of the

of communication channels along partisan and ideological lines. The fraud claims are echoed as indisputable by a multiplicity of sources that voters trust. Relatedly, Lenz (2012), among others, have shown that the public changes its policy views to match the politicians they support (e.g., attitudes toward Russia among Republicans track Trump's changing views about Putin). Also, sources supporting the claim of Trump's having won the 2020 election denigrate the reliability of the mainstream media who refute the fraud claim and insist that the mainstream media are simply partisan mouthpieces for the Democrats.

⁵ We would note, though, that media choice is not the only factor. It was impossible not to be constantly reminded about the fraud claims regardless of which media sources one followed. Virtually every claim made in the conservative press alleging election fraud was given considerable space in the mainstream media, followed by a rebuttal that often went on longer than the statement of the claim. For example, claims that a particular election had more voters than were registered were subsequently rebutted by showing that the list of eligible voters used for the comparison was incomplete (Reality Check team, 2020)). And there were follow-up pieces in the mainstream press repeating multiple claims of fraud and their rebuttals, sometimes drawing on social science analysis such as the discussion of fallacious assertions about voting machine vulnerability to fraud in Eggers et al. (2021). We might also note that, in the *Washington Post*, President Trump had previously been given tens of thousands of Pinocchio points based on other kinds of claims he made found by the *Post* to be untruthful, with points added every time the claim was repeated (Kessler, 2020). This constant drumbeat about Trump's factual misstatements and deceitfulness by a source that conservative voters did not trust may simply have reinforced the view among these voters that Trump was telling truths that the mainstream press was determined to hide. Relevant in psychological terms is the aphorism that "the lady doth protest too much," in which repeated assertions that a given allegation is untrue become taken as support for a belief that it probably is true.

2020 presidential election, and comparisons of its outcomes to those of previous presidential elections.⁶

Our goal is not to provide new insights about these statistical fallacies. Rather, our goal is put together in one place a useful compendium of many of the most glaring recent misuses of statistical reasoning as applied to understanding elections,⁷ and to do so in a way that is readily accessible to non-technical readers⁸ We believe strongly that a discussion of statistical fallacies based on real-world examples should be part of any statistics or public policy curriculum.

We begin our inventory of fallacies with (a) arithmetic fallacies of a simple sort, such as drawing conclusions from unweighted averages where use of weighted average was required, cherry-picking the data to emphasize only those facts that lead to the desired conclusion, and confusing percentages and percentage point changes. Then we discuss (b) improper use of statistical significance, and then turn to (c) inaccurate

⁶ Similarly, we do not wish to argue about what state or federal courts should or should not have decided about election law in the cases brought before them in 2020.

⁷ We do, however, believe that the framework of classifying election-related statistical fallacies that we offer may prove useful to other scholars.

⁸ We would call particular attention to the excellent survey and analysis of 2020 presidential election fallacies by Eggers et al. (2021). We began writing our own essay before we were familiar with this article in its on-line form, but we have learned a lot from it and cite to it herein. Their detailed rebuttal of fallacies is more technical than ours, and we cover some fallacies that they do not. There are also various journalistic surveys of election fraud claims (Alba & Frenkel, 2021; Corasaniti et al., 2020; Feldman, 2020), but these primarily include claims that do not fall within the scope of this essay

probabilistic reasoning, such as improperly using as an indicia of fraud having voters with the same name and date of birth. Then we discuss (d) syllogistic arguments based on cross-election statistical comparisons that are either fallacious in form, or that have at least one premise that is indubitably false, and thus which give rise either to invalid or unfounded conclusions. Finally, we briefly consider similar types of statistical errors in (e) syllogistic arguments based on within-election comparisons. Though we certainly do not claim that we have identified every fallacious claim about 2020 that fits within our classification scheme, we believe we have a typology that includes the most common errors involving analysis of aggregate election data.

II. Fallacious Statistically Based Claims about Massive Fraud in the 2020 Presidential Election

A. Arithmetic Fallacies

Failing to weight units. Failing to recognize that large changes in one direction in low population subsets or low population demographic subgroups can be compensated for by small changes in the other direction in high population subsets or demographic units is a common mistake. There were various instances of this type of error in discussions of the 2020 election. We first focus on the most common instance.

It was observed that Trump won more counties in 2020 than he did in 2016, with the implication being that he must have done better in 2020 than in 2016 in terms of the popular vote (Swenson, 2021). But, of course, that is nonsense, since he could have done better in the remaining counties. These counties, though fewer in number, had more

voters in them. Indeed, Biden received over three million votes in Los Angeles County, alone. In fact, Biden netted an additional 609,000 more votes in 2020 than Clinton did in 2016, just in this one county! Out of the over 3,000 counties in the United States, the top 150 contained half of the total votes cast. Biden won 125 of those 150 (83.3%).

Errors in using unweighted averages when it is appropriate to use a weighted average can be plotted in several ways. Consider a choropleth map of election results by county (**Figure 1**). As Chief Justice Earl Warren famously quipped in *Reynolds v. Sims* (1964), “legislators represent people, not trees or acres”. Trump certainly won more acres, but he did not win more voters. If you look at map of U.S. counties showing those won by President Trump in red and those by Hillary Clinton (or Joe Biden) in blue, you will see a sea of red and only a relative handful of pockets of blues. But those pockets (mostly big cities) have lots of voters in them. When counties are instead represented on a map in proportion to their votes, such as with a cartogram -- a map that has been resized so the units’ area is equal to its population weight – it become clear that the 2020 election was close, but certainly not an overwhelmingly “red” county shown by a county-level election map.⁹ Similarly, in the context of unequally sized units, a “bubble map” can be especially useful (see **Figure 2**).

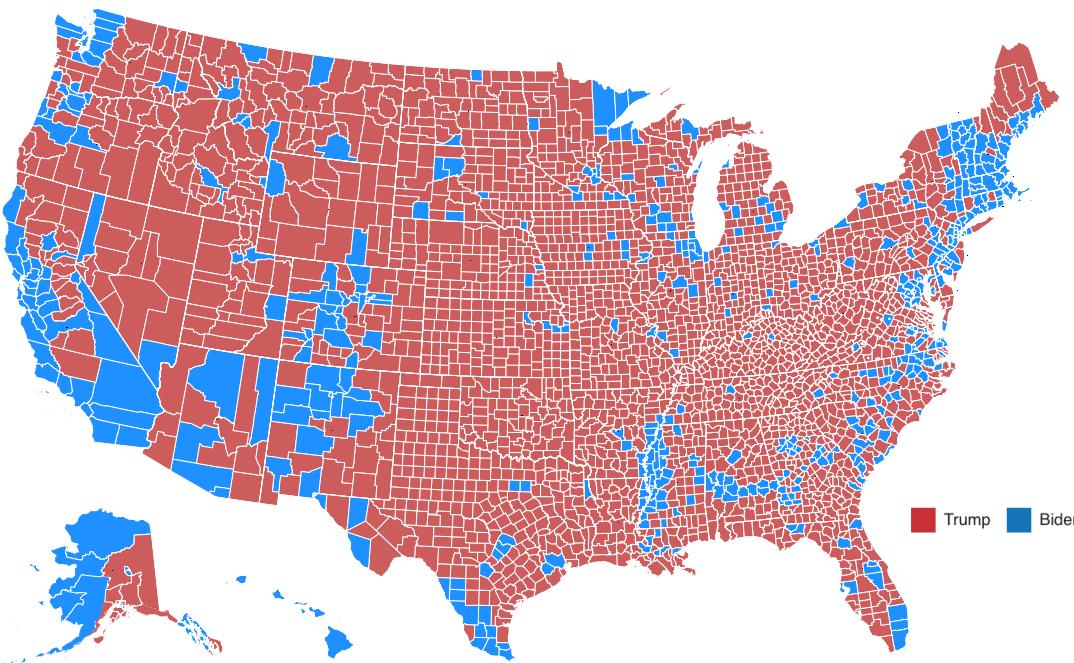
<<Figures 1 and 2 about here>>

⁹ For a nice overview of political graphics of different kinds, including a cartogram of the 2020 presidential election at the county level, see Bliss & Patino (2020). See also Cartographic Views of the 2020 US Presidential Election - Worldmapper (2020).

But it is still virtually impossible to visually sum-total results from a cartogram or bubble map to determine an election winner, especially when the number of units (say counties) is large, though the bubble size gradations on a bubble map make this task easier than the color variations on a cartogram that usually have a limited number of victory margin categories.¹⁰ Moreover, to further confound simple calculations, it is helpful to remember that the county with the most Republican votes anywhere in the USA was Los Angeles County, California. Trump received 1,145,530 votes there, although that number was dwarfed by Biden's 3,028,885. In fact, 8 of the 10 highest county vote totals for President Trump in 2020 come in states won by Biden.

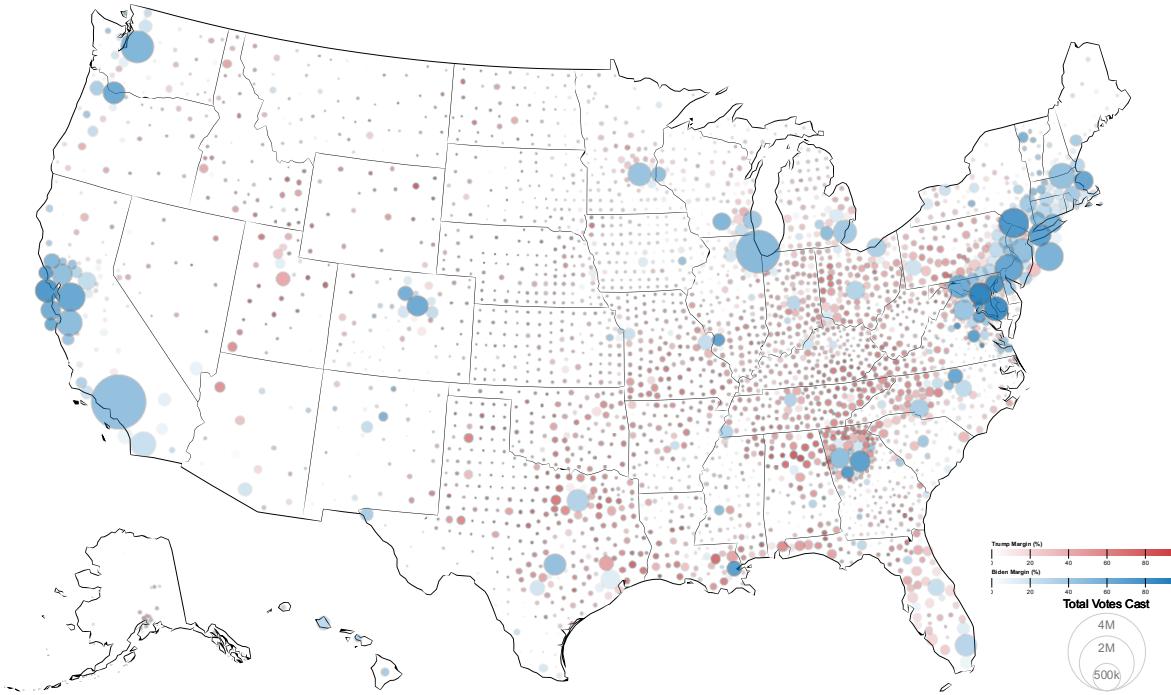
¹⁰ Though, note that the circle size in our map is scaled using the square root of the vote total, so the circle size itself is not a linear scale.

Figure 1 - Choropleth Plot, 2020 Presidential Election by county



Note: This choropleth map shows the 2020 Presidential election results by county. Each county is shown as geographically-sized using the Albers projection. This map depicts an election in which an overwhelming number of counties are colored red, indicating Trump received more votes than Biden. It does not indicate the number of votes each county was won by. Biden won 556 counties, while Trump won 2,595 (Data is estimated, as no official national election results are compiled. Brookings reports the differential to be 2,588 versus 551 (Frey, 2021). Dave Leip's Atlas of Presidential Elections, a widely reputed accumulator of election results, reports totals that do not match the certified, official federal elections results produced by the FEC).

Figure 2 - Bubble Plot, 2020 Presidential Election by county



Note: This bubble map shows a vote-weighted representation of the 2020 election. It describes two simultaneous variables, the winner of the county, and the number of votes the county was won by. Counties colored in blue were won by Joe Biden, and those colored in red were won by Donald Trump. The size of the circle indicates by what margin each candidate was victorious. Biden of course won the popular vote by over 7,000,000. Circles are not precisely positioned geographically but are adjusted to prevent overlap. Data sourced from the New York Times.

Failure to weight units also lead some commentators to emphasize Trump's gains among minority voters. Trump did show a substantial percentage point gain among minorities as compared to 2016. He won an estimated 12% of the Black vote in 2020, compared to 8% in 2016. Likewise, he won an estimated 32% of the Latino vote in 2020, an increase of 4 percentage points from 2016. But President Trump also won 1 percentage point more of the White vote in 2020 than he did 2016.¹¹

It might seem that these facts imply that Biden must have done worse in popular vote share than Clinton did in 2016. But there is one critical data point that suggests otherwise. While Trump did 1 percentage point better among White voters, Biden did 4 percentage points better among this same demographic! In 2020, 67% of the electorate identified as White in the exit polls, so increasing the margin among that category of voters will swamp decreases among other demographics. The actual combination of percentage point losses among minority voters and percentage point gains among non-minority voters gave rise to an election in which Biden won 81,268,924 votes (51.3%) of the popular vote (i.e., total vote) in 2020 compared to Clinton's vote in 2016 of 65,853,514 (48.2%).¹² Moreover, the increased total turnout further increased Biden's margin of

¹¹ Data extracted from exit polls: <https://www.cnn.com/election/2016/results/exit-polls>; <https://www.cnn.com/election/2020/exit-polls/president/national-results>

¹² Clinton won a plurality of the vote in 2016, not a majority. Several third-party candidates combined had vote totals surpassing the margin between Trump and Clinton.

<https://www.fec.gov/resources/cms-content/documents/2020presgeresults.pdf>
<https://www.fec.gov/resources/cms-content/documents/federalelections2016.pdf>

victory over Trump. Trump lost to Clinton by 2,868,686 million votes in 2016; he lost to Biden by 7,052,770 million votes.¹³

Of course, what really matters for Electoral College outcomes is *where* these changes occur. Since presidential elections are 51 separate state-wide elections (with the

¹³ There are other types of statistical confusions related to the Electoral College that we simply mention in passing because they are not directly linked to issues of fraud. Many voters have trouble understanding how the Electoral College works. It can most simply be thought of an example of weighted voting, where the weights are the number of electors each state is allocated; namely the sum of the number of representatives the state has in the U.S. House plus two, the number of U.S. Senators from the state. A common fallacy is to believe that a reversal of the popular vote outcome in the Electoral College occurs simply because of the two-seat “federal bonus” favoring small states. But Donald Trump would have won the Electoral College in 2016 even without the two-seat bonus (Cervas & Grofman, 2019). The vote in an Electoral College with only 438 members (538 minus 100 for the senate bonus) in 2016 would have been 248/438 (56.6%) as compared to the actual EC, 306/538 (56.9%); while Biden’s percentage an Electoral College of 438 members in 2020 would only one have gone up from 256/438 (58.4%) from the actual result of 306/538 (56.9%).

Because the Electoral College is a state-based weighted voting game (except for the two states where seats are allocated at the congressional district level as well), once a unit is won by a candidate, the margin of victory is irrelevant for Electoral College seat share but is relevant in terms of the popular vote share. Thus, we can have a candidate winning some states by very close margins in such a fashion that there is a discrepancy between the popular vote winner and Electoral College winner. Of the four instances post-Civil War examples of such a discrepancy, two of the four, 1888 and 2016, would still have happened without the bonus. But, in 2000, without the bonus, Al Gore would have been elected president, and in 1876 there also would have been no inversion had the two-seat bonus been eliminated. (Cervas & Grofman, 2019; Table 2, pp. 1328-29). Of course, we must be careful in asserting counterfactuals. For example, had there been no two-seat bonus in 2000, Gore and Bush might have deployed campaign resources differently.

minor complication of two states awarding votes by congressional districts), winning many more votes in states where the plurality is already secured creates “wasted” votes, since they do not contribute to winning any new electors. Taken all together, Biden was the benefactor of changes from 2016 that propelled him to the presidency, including narrow victories in a number of states won very narrowly by President Trump in 2016.

Two-party vote share versus shares of total votes cast. Reading the paragraphs above, the reader may understandably have gotten confused. President Trump’s percentage of support among non-minority voters increased in 2020 from what it had been in 2016. But Trump also made improvements among minority voters. So, how is it possible for President Trump not to have done better in 2020 than in 2016??!! The answer rests on a key difference between 2016 and 2020, namely the proportion of votes going to someone other than one of the major party candidates. In 2016, third-party candidates Jill Stein, Gary Jacobson, and others received a significant number of popular votes (but no electoral votes) totaling almost 8 million votes (almost 5% of the electorate), but in 2020, no candidate besides Trump or Biden received more than 1.25% of the national vote, with minor party candidates totaling less than 2% of the electorate collectively raw vote totals.

Clinton won 37% of the non-Hispanic White vote in 2016 and Trump won 57% of the non-Hispanic White vote (with 6% going to other candidates), which was 71% of the electorate. This amounts to 35,903,019 votes for Clinton and 55,310,056 votes for Trump. In 2020, Biden won 41% of the non-Hispanic White vote compared to Trump’s 58% (Trump gained 1-percentage point, and Biden gained 4-percentage points over Clinton), but the share of the electorate that was non-Hispanic White was now just 67%. Biden in

2020 received 43,507,921 non-Hispanic White votes and Trump 61,547,790. That is, the Democratic 2020 candidate had an increase of 7,604,902 non-Hispanic White votes compared to the Republican candidate's 6,237,734. So, on balance, in terms of raw votes 2020, Biden made more of a gain among non-Hispanic white voters over Clinton's 2016 performance than Trump did in 2020 relative to his 2016 performance among that group.¹⁴

Cherry-picking the data. The most primitive form of failing to weigh the data properly is cherry-picking the data to emphasize only those facts that lead to the desired conclusion. In presenting only some facts, a claimant can appear to have been honest while suppressing pertinent information that otherwise would prove their claims either false or

¹⁴ Total turnout rose in 2020 as compared to 2016 among both minority and non-minority voters. The proportion of the electorate that was minority increased in 2020 as compared to 2016 both in raw numbers and in proportion of the total. In 2016, approximately 97,035,186 voters were non-Hispanic White, compared to in 2020, where approximately 106,116,880 were non-Hispanic White (a difference of 9 million). By comparison, the electorate was composed of 39,634,090 minority voters in 2016, and 52,266,523 minority voters in 2020 (a difference of 12.6 million). In 2016, roughly 29% of the electorate was minority, in 2020 that percentage rose to 33%. However, Trump's gain from minority voters in both absolute number of voters received and proportion of the group supporting him was compensated for by loss in support among the larger group of non-minority voters. Moreover, even though President Trump made percentage point gains among minority voters in 2020 compared to 2016, as well as gains in raw vote totals, nonetheless, in terms of total votes from minority voters, Biden also did better than Clinton since there were more minority voters in 2020 than in 2016. To understand what is happening we need to look at changes in each group's share of the electorate, which is a function of the change in the size of the voting populations and the change in vote propensities among those populations.

incomplete. Because the data being cited is accurate, cherry-picking can prove a persuasive tool.¹⁵

A standard way to cherry pick election data to show the potential for fraud is to focus on racial or demographic groups and to highlight the situations where Biden did worse than Clinton. For example, Biden won fewer counties than Clinton. You can also get even more specific, e.g., Trump won a greater share of the vote in the city of Philadelphia in 2020 than he did in 2016; you can find other big cities where Biden underperformed Clinton. In these cherry-picked examples, Biden did less well against Trump than did the previous Democratic candidate and, since Trump beat Clinton, absent fraud, the implicit (and often explicit) conclusion is Trump must also have beaten Biden in terms of overall total vote nationally. But, of course, while there were some urban areas where support for Biden was lower than for Hillary Clinton, there were also urban areas where support for Biden was higher than for Hillary Clinton. Similarly, as discussed above, it is true that in terms of percentage voting for the Democratic candidate, racial minority support for Biden was (marginally) lower than for Hillary Clinton, and Trump did better among rural voters in 2020 than he did in 2016. However, support among categories of white voters, namely the college-educated and those living in suburbs, was higher for Biden than for Hillary Clinton. And there are more voters living in suburban residences than in rural areas.

¹⁵ Cherry picking is thus a strategy useful to those who know that the full evidence is against them.

It is, of course, the combination of all the subgroup patterns of voting and the (changes) in their sizes across elections that matters for the total popular vote. To reiterate the obvious, looking only at some subsets of voters, or only some geographic areas, e.g., only rural counties, is misleading and can lead to ridiculous claims that the candidate who received more votes did not actually receive more votes.¹⁶

B. Misinterpreting Statistical Significance

In statistics, warnings about misinterpreting statistical significance are legion. One point frequently made is that statistical significance can only be interpreted in the context of the specific null hypothesis being tested. Another is a reminder that p-values are very strongly linked to sample sizes. A third point is a warning not to confuse statistical significance with substantive importance. A further note of caution is that high statistical significance in regressions does not imply a causal relationship between variables.

Our first example of misuse of statistical significance falls into the second and third categories. Our second example involves the misinterpretation of causality. Both are found in the expert witness testimony of Dr. Charles Cicchetti in the lawsuit brought by Texas Attorney General Ken Paxton challenging election results in Georgia, Michigan, Pennsylvania, and Wisconsin (*Texas v. Pennsylvania*, 592 U.S. ___, 2020).¹⁷ Dr. Charles Cicchetti's calculations were picked up and widely spread on the internet.

¹⁶ Moreover, in understanding Electoral College outcomes, we need also to look at the geographic location of each candidate's support.

¹⁷ As we noted earlier, we are providing a compendium; this and other claims of Dr. Cicchetti have already been rebutted elsewhere (see e.g., <http://web.archive.org/web/20220416221931/https://reason.com/volokh/2020/12/09/>

First, in comparisons between what was found in 2020 and what was found in 2016, Dr. Cicchetti noted that he could demonstrate beyond any possibility of error that the vote share distribution for Joe Biden in 2020 differed from that of Hillary Clinton to a statistically improbable degree. He is certainly right about that fact.

But what that shows about election fraud is — exactly nothing! It does show that the vote share distribution for Donald Trump in 2016 was not the same as in 2020. But we could have picked ANY two adjacent presidential election years and showed that the vote shares of the Democratic candidates in those two elections at the level of states or counties or precincts differed from one another. With a large enough sample size, such comparisons are virtually certain to generate statistically significant differences.

Second, in a comparison of partisan voting patterns in ballots in 2020 among ballots that are tallied early in the counting process and those that are tallied later, Dr. Cicchetti pointed out that, in some states, President Trump's share of the vote declined relative to those first reported as polls closed on election night as more ballots were tabulated, while in other states he found the reverse pattern. He found the difference between the early vote share for Trump and later vote share for share to be statistically

more-on-statistical-stupidity-at-scotus/
and <http://web.archive.org/web/20220416221815/https://statmodeling.stat.columbia.edu/2020/12/08/the-p-value-is-4-76x10%E2%88%92264-1-in-a-quadrillion/>. That report has been devastatingly critiqued in the expert witness report of Gary King in the same case. Claims about election fraud in *Texas v. Pennsylvania*, including Dr. Cicchetti's report, can be found here:
https://www.supremecourt.gov/DocketPDF/22/22O155/163048/20201208132827887_TX-v-State-ExpedMot%202020-12-07%20FINAL.pdf

significant beyond any reasonable doubt, though of course, the directionality of the difference was not uniform. And about those facts and the statistical significance of the observed differences he is again quite correct. But what that shows about election fraud is — once again, exactly nothing! Indeed, for those familiar with elections, it shows a pattern that was predicted in advance (Foley & Stewart III, 2020).¹⁸

While once it was believed that mail-in (absentee) ballots were disproportionately Republican, for the last two decades or so the evidence has shifted, suggesting that mail-in ballots are more Democratic-leaning than is the case for ballots cast in person.¹⁹ In some states in 2020, mail-in ballots were tallied as they came in; thus, they were the first ballot results to be reported; in other states, the tallying of mail-in ballots did not begin until after the polls were closed; and in select states, they were the last to be tallied. Thus, we would expect that there would be differences across states that would be in part predictable by when mail-in ballots were tallied. This could account for why, in some states, the early vote was more Democratic than the later vote, while the reverse was true

¹⁸ In the hearings by the House Select Committee investigating January 6, this discrepancy between the patterns in early and late votes was referred to as a “red mirage.”

¹⁹ Perhaps because states with more Democrats allow for more wide-spread use of mail-in balloting, e.g., eight states conduct general elections entirely by mail. They include California, Colorado, Hawaii, Nevada, Oregon, Utah, Vermont, and Washington. See (*Voting Outside the Polling Place: Absentee, All-Mail and Other Voting at Home Options*, 2022). All of these, except Utah, voted their electors to both Clinton and Biden.

in other states. Thirty-seven states allow election officials to begin processing mail-in ballots as they arrive prior to election day. Another ten states allow processing to begin on election day, but prior to polls closing. Only Maryland does not permit counting mail-in ballots until after polls close.²⁰ Battleground states Michigan, Wisconsin, and Pennsylvania are among the ten states that do not allow counting mail-in ballots until election day.

But there is also another reason why late ballot tallies and early ballot tallies might be expected to be different in their partisan outcomes. Ballots tallies come in from counties and then get tallied statewide. When ballots get reported depends upon the efficiency of the county officials doing the tallying. But it is often easier to complete in-person voting tallies in small rural counties where there are few polling stations than in counties with large cities. But rural counties differ from urban counties in their partisan propensities. So, it is essentially guaranteed that there will be differences in partisan vote share as county reports come in at different times on election night. In the end, only the final count of all ballots is valid.

In sum, when the null hypothesis is nonsensical, there is nothing at all surprising about it being rejected at a high level of statistical significance, especially when the sample size is high. Indeed, what would be surprising is if silly null hypotheses were NOT rejected.

²⁰ Two states, Connecticut and Ohio, do not specify when counting may begin. For more information, see Voting Outside the Polling Place: Absentee, All-Mail and Other Voting at Home Options (2022; Table 16).

C. Meretricious Probabilistic Reasoning

Voters with the Same Name and Date and Year of Birth. It has been noted that Joe Frazier voted in Pittsburgh in 2020 even though he had died in 2011. This fact was taken as evidence of fraud (Sadeghi, 2020). But of course, not all Joe Fraziers' are the famous (and now dead) boxer. Similarly, instances, where individuals with the same name and same birthday, and/or same birth year were found on voter rolls, have been taken as evidence of fraud.

It should be noted that some states record a default birth date (or a code for missing data) if none is provided, particularly if no date was provided when this information was collected on paper. This will, of course, if not corrected for, increase the likelihood of the appearance of voters who appear on the voter rolls multiple times. But the likelihood of occurrence of voters appearing on voting lists who have identical names and even identical birth dates, and even identical years of birth is dramatically underestimated by virtually everybody, except for those who have truly understood the famous "Birthday Paradox." The Birthday Paradox in its basic form is about how the number of people in a group affects the probability that some pair of them have the same birthday.

When first confronted with the question of how the number of people in a group affects the probability that some pair of them have the same birthday, students in statistics classes (or at least those in the undergraduate statistics classes taught by one of the present authors) think that, since any given birthday has a probability of only 1/365 (or 366 if you take into account leap years), it would take around 183 people before you would get a probability of 50% that two people in the group had the same birthday. The

Birthday Paradox is called a paradox because it only takes a group of size 23 for the probability of two people in the group to have the same birthday to exceed 50%.

The probability that two people in a group of size n do NOT share the same birthday can be written as $1 * \left(1 - \frac{1}{365}\right) * \left(1 - \frac{2}{365}\right) * \left(1 - \frac{3}{365}\right) * \dots * \left(1 - \frac{n}{365}\right)$. This product goes down much faster than one might think, and thus the probability that at least two people in the group share the same birthday, which is one minus this product, goes up much faster than one might think.²¹ Indeed, with only 75 people, the probability of a birthday match rises to 99.95%. In a similar way one can calculate the expected proportion of a group of size n who share a birthday with at least one other person in the group. The difficulty in appreciating how the increasing number of possible pairs that could share a birthday (and other attributes) increases in a non-linear way with increasing n is relevant to claims made in 2020 (and earlier) that find examples of people with the same name and same date of birth on the voting rolls was evidence of “double-voting” fraud.²²

²¹ This product can be rewritten in terms of combination or permutations, which gives a somewhat more intuitive way to see how the probability changes with n . See https://en.wikipedia.org/wiki/Birthday_problem#Calculating_the_probability. A program that calculates the probability of the birthday paradox for any given n is found at <https://www.statisticshowto.com/same-birthday-odds/>

²² Matching procedures are sometimes used to identify duplicate voter registration entries. States are required to maintain lists of eligible voters, which are complicated by people migrating between counties and states, but also within neighborhoods and municipalities. But matching procedures sometimes lead to “false positives”, where it appears two entries are the same people, but they are in fact not. See more in (Hasen, 2020), and (Weiser et al., 2006).

While, for any given name, it is obviously harder to find someone born on the same day and in the same year as it is merely to find people with the same name and the same birthday (but not the same year of birth). The logic of figuring out the probability of such a match happening is the same as the simple product formula given above, but now the divisor is no longer 365 since we need to take into account the years in which a person might have been born (though we may reasonably assume that everyone who votes was born at least 18 years ago). Moreover, as a further complication, we must consider the degree of heterogeneity in the distribution of names.²³ But if we take name, birthday, and birth year as mutually independent factors,²⁴ then we can simply multiply probabilities.²⁵ And we can further simplify by assuming a uniform distribution across the first two factors

²³ Calculating this probability from the name distribution in the population requires calculating a self-weighted average based on name and surname. Names vary greatly in their frequency. The U.S. Bureau of the Census publishes a list of surname frequencies based on national values, but the distribution of names (and their degree of heterogeneity) will vary with the racial/ethnic composition of the political unit (Grofman & Garcia, 2014, 2015).

²⁴ Birth year and name are not fully independent of one another. Changing patterns of immigration affect the relative surname shares in different generations, and some rather dramatic changes over time in the popularity of first names mean that first name probabilities are not independent of the year of birth. Also, first and last names are far from independent since both are linked to ethnicity. Indeed, we could, in principle, estimate birth decade probabilities just by looking at the prevalence of first and last names of those born during the decade (c.f., Grofman & Garcia, 2014, 2015)

²⁵ Note that the probability of finding such a match from among a set of n voters must not be confused with the probability of any specific name + birthday + birthyear combination being repeated. For obvious reasons the likelihood of a given combination being repeated will depend, *ceteris paribus*, on how common is the name.

and assess the likelihood of two randomly chosen individuals bearing the same name from the name distribution in empirical data.²⁶

Of course, multiplying probabilities for three different factors gives us low probability values, but not as low as one might think. For example, if a randomly chosen person has a 0.000074 percent chance of sharing both a first and a last name with the next randomly chosen person, which was the estimate from the McDonald & Levitt (2008; p. 119, fn. 26) study,²⁷ a total electorate of only 21,071 is enough to bring the probability of finding two people with the same name and the same birthday and the same birth year above 50%. And an n of 57,314 brings that probability to 99.5%. Moreover, as McDonald & Levitt (2008) demonstrate with their detailed analyses of New Jersey electoral rolls, there are other reasons why we see individuals shown with identical data identifiers, including flaws in the data such as the same individual simply being entered twice or individuals with missing data birth data (assigned a particular missing data code) being treated as having identical records.

²⁶ The most sophisticated study of double voting of which we are aware (McDonald & Levitt, 2008) investigates complaints of election fraud in New Jersey in 2004 that were based on apparent observance of thousands of instances of double voting, includes a comparison of the assumption of uniform distribution of year of birth and name distribution to actual data and show that assuming a uniform distribution over a 64 year interval and over the names in the data set tends to underestimate the prevalence of birth year matching by about $12\% \cong (487 - 433)/433$, see p. 119). The more there are birth bulges (as in the post-WWII baby boom), the more likely it is that two randomly chosen individuals will share the same year of birth, and some names are much more prevalent than others.

²⁷ For further compatibility with the McDonald & Levitt (2008) study, in the model, presented below, we also took our birth year time period as a 64-year span.

Benford's Law. There have been many attempts to use statistical tools to detect election fraud. One of these involve looking at suspicious data, e.g., vote tallies that disproportionately end in 0's or 5's since, in "inventing" data, there can be a human tendency to call these numbers to mind more often than would be expected from a purely uniform distribution ("Russian Elections Once Again Had a Suspiciously Neat Result," 2021). While it makes no sense to look at the first digits of election returns since these will be obviously contingent on the mean size of the units,²⁸ in investigating fraud, the frequency of digits other than the last or first digit has also been investigated. Benford's Law, which is a hypothesized frequency distribution of k^{th} digits, is an example of one tool that has been used to assess *prima facie* evidence of voter fraud. The political scientist Walter Mebane applied this tool to elections in multiple countries (Mebane & Kalinin, 2009). In the 2020 presidential election, analyses based on Benford's Law were provided in videos and tweets as evidence that, in various locales, elections had been rigged.²⁹ We will make no attempt to repeat the logic that leads to Benford's Law (see the discussion of the supposed Law in Wikipedia and references therein);³⁰ we simply note that almost all of those who have investigated it empirically is dubious about its application to elections.³¹

²⁸ Mebane (2020) notes that "It is widely understood that the first digits of precinct vote counts are not useful for trying to diagnose election frauds."

²⁹ See e.g., (Jenny, 2020).

³⁰ Wikipedia entry [Benford's Law]:
https://en.wikipedia.org/wiki/Benford%27s_law#cite_note-37

³¹ See e.g., Deckert et al. (2011). Even Mebane, who has been a repeated applicant of the Law as a fraud detector in elections in multiple countries, has reiterated that

According to (Mebane, 2011), *Benford's Law* implies that the mean value of the second digit of a distribution of votes at the precinct or other unit-level should equal 4.187. But a little reflection will show that the frequency of even second digits will be linked both to the size of units and their variance.

Imagine, for example, that the mean Democratic votes cast in a set of precincts is 1,650 voters, with a variance of 200. The second digit will have disproportionately many more 4's, 5's, 6's, 7's and 8's than it has other numbers. If we change either the mean or the standard deviation of that distribution, then the expected mean value of the second digit will also change. Thus, there cannot be some fixed value that is the expectation of the mean value of the 2nd digit in all distributions. And the same point applies to whichever digit we focus on, or whether we look at all of them since the scale of units will affect which digits are most likely to deviate from Benford's Law expectation. But perhaps even more importantly, it is quite possible for Benford's Law to work well for some candidates and badly for others since different candidates will have different means and variances in their vote distribution across units..

The next two examples of meretricious probabilistic reasoning are more about human psychology than they are about statistics, *per se*.

Tip of the iceberg fallacy. No one who studies elections thinks that the level of election fraud in 2020 was zero. In an election with over 150 million voters, there will be some who

while its violation might be taken as suggesting an anomaly, a violation of the supposed Law does not prove fraud. Fraud would need to be directly investigated (Mebane, 2020).

voted who were not entitled to do so, some who voted twice, some whose ballots were lost, some whose mail ballots were stolen and cast by others, and some who voted while dead (perhaps having died after casting a ballot). And there will be some election officials who (at least initially) reported inaccurate results. But while fraud in any given state may have cases in the tens, or perhaps even hundreds, this level is not enough to change presidential election outcomes (see below).

One fallacy of probabilistic inference is to infer from the indubitable fact of some electoral wrongdoing in most states that electoral fraud is massive in both magnitude and geographic spread. The careful studies of fraud, such as the one of the 2020 election undertaken by the Ohio Secretary of State, reveals fraud at a minuscule level of fraud: 27 cases out of 6 million or so ballots cast in Ohio (McDade, 2022). However, findings such as those of the Ohio Secretary of State, rather than being taken as evidence that the fraud level was trivial, can instead be interpreted as supporting evidence for a high probability of massive fraud.³² The implicit and quite wrong-headed probabilistic argument is that any examples of fraud that are found should be presumed to be “only the tip of the iceberg.”³³

³² There is also a belief in conservative internet circles that, if there is fraud, it will be fraud committed by Democrat officials and found among the kinds of voters (such as racial minorities) who are most likely to vote Democratic. But is it is amusing to note that one of the first documented examples of actual fraud, so-called “voting the graveyard,” was committed by a Republican “in an attempt to further President Trump’s campaign” (Vella, 2020).

³³ As Gelman, 2021) points out, some people say things like “who’s to say” when they hear claims that are patently implausible but respond in a way suggesting that they believe that claim might be true. Here, possibility is confounded with probability. A common aphorism also relates to this confusion, e.g., “Where there is smoke there

Straw man fallacy. On the other hand, the claim that there was no massive fraud can be rhetorically equated to the claim that there was no fraud, and the latter claim rebutted as a straw man, as if successfully rebutting the claim of no fraud demonstrated the truth of the claim of massive fraud. Straw man arguments can be used to perpetuate a big lie, as has been the case following the 2020 presidential election.

Now let us turn to a set of other claims that have as their general form: “The only way these election results could have happened is if there was massive fraud.”

D. Logically Invalid Arguments with a True Premise involving Historical Election Results Comparisons

Spoiled ballots. An example the empirically accurate observation that the spoiled ballot rate of mail-in ballots in 2020 was much lower (in some states remarkably lower) than in 2016. This fact is taken to be evidence of mail-ballot fraud by Trump supporters. Presumably the theory is that the lowered spoilage rate came about either because more ballots were illegally cast in 2020 than in 2016 and that those stealing unfilled-in or incomplete mail ballots and submitting them knew how to properly fill out the forms, or that ballot workers in polling stations where the mail ballots were likely to be Democratic ones deliberately allowed improperly filled out mail ballots to pass muster in a way that they had not in 2016.

is fire.” Here, the size of the fire remains unspecified. In the context of election fraud, confusing possibility with probability is likely to be more prevalent among those who see the world in conspiratorial terms. Gelman quotes one person who accepts claims such as Obama is a Muslim as saying: “We see what they want us to see, I mean anything could be anything.”

We can write the argument as

If there is ballot fraud involving mail-in ballots (A),

then the spoilage rate among mail-in ballots will be lower than in the past
(B).

The spoilage rate among mail-in ballots was lower than in the past (B)

Therefore, there was ballot fraud involving mail-in ballots (A).

Here we have the fallacy of *affirming the consequent*.

There are good reasons why ballot spoilage was lower in 2020 than in 2016 that have nothing to do with fraud, namely much greater effort on the part of election administrators to inform voters of what they needed to do to cast a valid ballot. For instance, popular late night comedy Stephen Colbert created a rather sophisticated website aimed at informing those in all 50 states about the specifics for casting a ballot in each of those states (*Better Know a Ballot*, 2021). His “Better Know a Ballot” also aired many times in the months before the elections on his highly rated “The Late Show.” Ads developed by the states themselves aired on television channels and as ads on streaming services.³⁴ The Democratic National Committee also spent millions of dollars on television ads with information about returning mail-in ballots(*DNC Launches New Digital Ads in PA*

³⁴ In Pennsylvania, this included pointing viewers to a website developed by the Department of State. Ads featured prominent actors and athletes from the state.

Reaching Vote-By-Mail Voters: “How to Return Your Ballot!”, 2020). There was also ample coverage in newspapers about properly filing out and mailing a ballot so that it would not be rejected (Lai, 2020).

Moreover, in some states, there were greater efforts to ensure that those who submitted a mailed-in ballot with an envelope which had some correctable error that would prevent the ballot inside the still unopened envelope from being counted were informed of the error and given the opportunity to correct it. Eighteen states allowed voters to “cure” their ballots if there is a discrepancy (*States That Permit Voters to Correct Signature Discrepancies*, 2020). These states are disproportionately Democratic; Trump won just 5 of the 18. But in our federal system, absent genuine issues of potential voting rights violations, states can and do differ in the details of their election administration.

E. Logically Valid Arguments with a False Premise involving Historical Election Results Comparisons

Now we turn to claims about election fraud that fall into the category of valid arguments with one or more false premises.

Presidential coattails. One claim about election fraud was based on the observation that winning presidential candidates have coattails that aid members of their party in the House of Representatives to gain seats. The Democrats lost 12 seats in the House of Representatives in 2020, which violates this expectation, and so the claim goes that the implication is that there must have been massive multi-state fraud.

The structure of this argument is:

If a presidential candidate wins election (A),

then there will be a gain in the number of members of his party in the U.S. House of Representative (B).

There was no gain for the Democrats in the House in 2020 (not B)

Therefore, Biden must have lost the election (not A)

This is a valid argument. It is an example of *denying the consequent*. However, the premise on which it is built, that presidential coattails are inevitable, is false.

By *coattails*, we are referring to an increase in the number of members in the U.S. House of Representatives that share the incoming president's party (Campbell, 1986). Negative coattails are not uncommon, and in contemporary politics, have become more likely.³⁵ Since 1868, there have been thirteen elections where a president has had negative coattails (including 2016 and 2020). Negative coattails are more likely when (a) elections are close in popular vote (b) there is substantial partisan bias against the party of the presidential winner in the House, (c) a substantial portion of the votes for the winning presidential candidate are wasted in states that are won by large margins, and

³⁵ See Morris Fiorina (2017) "Unstable Majorities", and Frances Lee (2016) "Insecure Majorities". Modern elections are very likely to result in divided government, and control of any branch of government is often won or lost in the margins. Partisan bias, such as that introduced by malapportionment or gerrymandering, can also affect the ability to carry marginal House or Senate seats.

(d) the winning president's party picked up a significant number of seats in the previous midterm election. All four of these features are found in 2020.

Biden's share of the major party vote was only 52.27%; the estimated partisan bias in 2020 in the House of Representatives in 2020 was 2.7%. Congressional districts have become far less competitive in recent elections, leaving fewer chances for a president to provide coattails large enough to flip seats (Engstrom, 2020).³⁶ If we eliminate the states that gave the widest raw margin to Biden (California and New York and Massachusetts) from the calculations, Trump has a majority of the vote in the remaining states – hence, we would not expect to see Biden coattails in those remaining states.³⁷ Democratic gains in the House in the 2018 midterm were significant, and turnout was a level not seen before universal adult franchise (Jacobson, 2019). Moreover, up through 2016 there is a time trend of decreasing presidential coattails which, when projected onto 2020, would create an expectation of a negative coattail in the 2020 election.³⁸ But perhaps most importantly, there were 35 House constituencies carried by Trump in 2016 but with a Democratic House member elected in 2018,³⁹ and only 5 House constituencies lost by Trump in 2016

³⁶ For more details on Biden's overperformance compared to U.S. House Democratic candidates, see Galston (2020).

³⁷ House Democrats nationally underperformed their 2018 performance, which accounts for the fourteen net seats gained by the Republicans. Relative to the 115th Congress (2016-2018), the 117th Congress (2021-2023) has 28 more Democrats.

³⁸ Figure omitted for space reasons.

³⁹ An increase of 22 from 2016.

but with a Republican House member elected in 2018.⁴⁰ Thus, Democrats in 2020 had many more vulnerable House seats than did the Republicans.

Bellwether counties. An argument of a similar form is that Biden lost most of the counties that had been bellwether counties, and therefore since bellwether counties predict presidential elections, Biden must really have lost the election.

Again, we can write this argument as

If a presidential candidate wins the election (A)

then they can be expected to carry almost all the bellwether counties (B),

Biden lost almost all of the bellwether counties (not B)

Therefore, Biden must have lost the election (not A)

This, too, is a valid argument -- another example of *denying the consequent*. However, even though “not B” is empirically accurate, the premise on which the argument is built, namely that bellwethers predict elections, is false.

Many decades ago, the political scientist Edward Tufte (1974; chapter 3) wrote a devastating rebuttal to work on the power of bellwethers. Tufte showed that, over the period 1916-1968, there were no real state-level bellwethers and, most importantly, the U.S. counties identified as presidential bellwethers at time t had no better track record at

⁴⁰ A decrease by 19 from 2016.

the next presidential election than the non-bellwether counties. (Hopkins, 2017) showed the same result for much more recent data. Yet, belief in bellwether units of geography, more particularly in the existence of bellwether counties, refuses to die.

Grofman & Chen (2022) explain the predictive failures of bellwethers partly in terms of classic work of Deutsch & Madow (1961). A simple binomial model shows that by chance alone, in large groups, some individuals can appear to have repeated (predictive) success even though, for any given event, the probability of success of any actor is only 0.5. This is true when we are dealing with independent trials where past success tells us exactly nothing about future success.⁴¹ But Grofman and Chen also elaborate on the conditional probability model that generates the likelihood of bellwethers by showing that as partisan polarization increases, and presidential politics nationally is competitive, so that the Electoral College sometimes have Democrats winning and sometimes Republican winning (three each in the 21st century), the likelihood of bellwethers declines.

The intuition here -- which they examine at the level of counties -- is a very simple one: in order to be a bellwether county, a county must vote for the winner both when the winner is a Democrat and when the winner is a Republican, but *ceteris paribus*, the increasingly polarized patterns of voting have by now put virtually all counties firmly into one partisan camp or the other. In the 21st century, Grofman and Chen show that more than 70%+ of all counties vote consistently Republican, and another portion are

⁴¹ Relatedly, (Grofman & Chen, 2022) provide a conditional probability model to further explain change over time in the ability to use past elections to predict future elections.

dependably Democrat. These counties make great bellwethers only when their party is always winning, but awful bellwethers otherwise.⁴² Thus, bellwether performance has been falling. Moreover, if inter-election changes in vote propensities vary across types of voters (or geographic units) than bellwethers can perform particularly dismally in a subsequent election. As a graph in 538 shows, counties that previously served as bellwethers have been shifting rightward (Matsumoto, 2021). Thus, when the Republican candidate loses, they will not serve as bellwethers.

Other Cross-Election Comparisons. Other arguments with specious premises supposedly demonstrating that Biden could not have won in 2020 also make use of differences between the 2020 election and patterns found in previous elections. For example, Trump supporters such as Shurk (2020) noted that no incumbent who has won more than 75% of the primary vote has lost their election. Therefore, since Trump had won 94% of the primary vote, he must have won re-election.⁴³ Shurk also observes that “no incumbent in over 100 years who has gained votes in his reelection bid has lost his

⁴² Indeed, Eggers et al. (2021; Figure 2) show that only 2% of counties had a different party winner in 2020 than they did in 2016.

⁴³ This claim links to Twitter user David Chapman (@davidchapman141), a self-proclaimed “Author & Historian”. This “thread” is filled with statistics purportedly showing how Biden is the historic underdog going into the 2020 election., e.g., “Incumbents are 6/6 when facing re-election during civil unrest”. See more: <https://twitter.com/davidchapman141/status/1315440579485069314?s=20>

Internal links to this claim on Twitter say that the first primary was in 1912 and that Trump had received a higher percentage of the primary vote than Eisenhower, Nixon, Clinton, and Obama. And it is noted that only five incumbents have received at least 90% of their primary vote.

quest for reelection.”⁴⁴ On the other hand, we could just as easily claim that incumbents whose overall presidential approval rating among independents was as low as that of President Trump must lose re-election. Unfortunately for this type of argument, there is no guarantee that what was true in the past will be true in the future. And, of course, these historical comparisons are based on small n , especially when we limit our comparisons to those where there is an incumbent running for re-election.⁴⁵

F. Logically Valid Arguments with False Statistical Premises Using Comparisons Based on Features or Components of the Same Presidential Election

Matching design (within election split ticket voting versus straight ticket voting).

One matching comparison involves comparison of elections where there are two kinds of votes: straight ticket votes (where a single check casts a vote for all candidates of a given party for all offices) and ballots where this box is not checked. Ayyadurai (2020)⁴⁶ claims that support levels in straight ticket votes and split ticket votes should be unrelated, unless there is voter fraud. He uses data from four counties in Michigan where the straight ticket vote option was available. Ayyadurai asserts that a negative correlation indicates that the

⁴⁴ This claim fails to account for well-known political science theory that suggests that competitive elections increase voter turnout (Downs, 1957), and the empirical fact that modern elections are increasingly competitive at the national level (Lee, 2016) even though less competitive for other offices, such as the U.S. House (<https://alarm-redist.github.io/fifty-states/>).

⁴⁵ We might also argue that neither Trump nor Hillary Clinton could possibly have been elected President in 2016 since both were so disliked. Of course, no matter how unpopular one candidate may be, the other person might be even more unpopular. But also, relative likeability, is only one of the factors affecting vote choice.

⁴⁶ Ayyadurai has a Ph.D. from MIT in biological engineering.

split ticket votes have been manipulated to reduce the apparent Trump share of the vote, i.e., that a negative correlation is a sign of voter fraud.

We can write this argument as

If the difference between straight ticket vote share for the candidate of a given party and split ticket vote share for the candidate of that party is negatively correlated (A),

then there is vote fraud in favor of the other party (B).

The difference between straight ticket vote Republican share and split ticket Republican vote share in four Michigan counties is negatively correlated with straight ticket Republican vote share (A).

Therefore, there must have been vote fraud favoring the Democrats in those counties (B).

This is, in principle, a valid argument. But the premise makes no sense.

The notion that the two-vote shares should be uncorrelated if there is nothing suspicious going on might appear on its face to be plausible, but some reflection reveals that it is completely misguided. Depending upon whether we take the difference as split ticket vote share minus straight-ticket vote share or as straight-ticket vote share minus split ticket vote share, it should be inevitable to get a non-zero correlation. In fact, a

negative correlation is guaranteed by the model used in Ayyadurai, since he subtracts x from y . Below we work out a toy example.

Let's take as an approximation that straight ticket voters have the same vote propensities as among split-ticket vote. Let's set s to the share of the Republican voters who cast a straight ticket vote and r_i to the Republican share of the vote cast by voters of the given type, with $i = 1$ straight ticket voters and $i = 2$, split ticket voters. $r_1 s$ is then correlated with $r_1 s - r_2(1 - s)$, or with $r_2(1 - s) - r_1 s$. Let us simplify further by positing that $r_1 = r_2$. Now, doing what Ayyadurai wishes us to require simply correlating r with either $2r - 1$ or $1 - 2r$, thus guaranteeing either a positive or a negative correlation depending upon which way we do the difference. But there is NO reason to expect is a zero correlation.

But, if the argument above did not convince you that Ayyadurai syllogism is misguided, just think about the failure of symmetry. Under our simplifying assumptions, the exact same argument he makes for Republicans goes through for Democrat votes as well, so that his argument leads us, on the one hand, to the conclusion that there was fraud favoring Democrats (when we look at Republican vote shares) and on the other hand, to the contradictory conclusion that there was fraud favoring Republicans (when we look at Democratic vote shares).⁴⁷

⁴⁷ This rebuttal to Ayyadurai (2020) is presented in (Eggers et al., 2021), acknowledging its previous elucidation by Kabir (2020) and Parker (2020), each of whom show empirical evidence of the party-independence of results. However, the Eggers, Garro, and Grimmer main rebuttal to the Ayyadurai (2020) claim uses the logic of latent variable analysis by demonstrating how regression to the mean effects lead to negatively sloped regression lines in the situation posited by Ayyadurai.

Comparison of differences involving samples with different means. There's a second type of fallacy involved with the examples shown in Ayyadurai. It is claimed that in Wayne County the pattern of Trump drop-off is different, and therefore the algorithm that transferred votes to Biden was not used. However, one need only look at the x-axis to see that in Wayne County the most Republican precinct has only about 30% Trump support. In the other counties, some precincts have 80% Trump support. In the range of the Wayne County plot, the pattern is like the other counties. One must be careful when presenting or consuming information from graphs that may mislead, even if done in ways that are not intentional. In this case, the data simply does not exist such that there are heavily Republican precincts in Wayne County to compare to the other counties.

Matching design (within-election comparisons of areas with and without fraud claims). Lott (2020) uses an apparently sophisticated attempt to prove election fraud via statistical analysis, which is now forthcoming in a peer-reviewed journal (Lott, n.d.). While Lott offers various forms of analysis, the most intriguing one, which appears to be quite plausible,⁴⁸ is a matched pairs analysis which compares the difference in support for Trump in adjacent precincts in a pair of counties in Georgia and pair of counties in Pennsylvania, such that one of each pair is in a county which voted for Biden and one is in a county which voted for Trump.⁴⁹ The counties are distinguished by their expected

⁴⁸ Indeed, we regard Lott's analysis as sufficiently plausible, on its face, to explain why, despite being either completely wrong or at best much overstated, it could nonetheless pass peer review.

⁴⁹ Lott (2020) also offers a similar kind of argument about why turnout differences in adjacent pro-Biden versus pro-Trump counties could also be indicators of fraud. But that argument also is fundamentally flawed because it assigns to fraud differences

partisan outcome. Lott takes a major urban county (which voted heavily for Biden) and treats it as one in which fraud might be present; the adjacent counties (which are either suburban or rural) are usually Republican dominated (and voted for Trump) are treated as a baseline in which no fraud would be expected. Lott attempts to explain the difference in absentee votes for Trump between adjacent pairs of precincts in the two types of counties by comparing them to the difference in in-person votes for Trump in the same pairs of precincts. The regression model includes a dummy variable for precincts located in the county which voted for Biden. He claims that a positive sign on this dummy variable indicated the presence of fraud. He obtains positive but not statistically significant results for both Georgia precinct pairs and Pennsylvania precinct pairs. He then combines the data for both states and finds a positive and statistically significant result on his dummy variable for the pooled data.

Eggers et al. (2021) provide a variety of arguments, both theoretical and empirical, to show that the econometric designs in the (Lott, 2020) study are so flawed that its conclusions are completely vitiated. That study has the advantage of making use of the data provided by Lott for replication purposes.⁵⁰ Here we simply make two points about conceptual flaws in the Lott analyses.

across units that have other very plausible non-fraudulent explanations. See Eggers et al. (2021).

⁵⁰ One point made by Eggers et al., (2021) is that the results for Democratic vs. Republican County comparisons depend upon the order in which we select the units; another point they make is that the exactly which precincts are paired can reverse

On the one hand, it is a huge leap to believe that a positive sign on the dummy variable for precincts in pro-Biden counties in his analyses can be taken as evidence of fraud. There can be many reasons that have nothing to do with fraud for why Trump vote shares among mail-in voters differ from in-person voters across precincts. For example, while we do expect the area on the border between the two paired counties are similar, in Pennsylvania where we are most familiar with the data, there are key differences, such as income levels, homeownership levels and tax rates between the two paired counties that can be expected to affect voting choices.⁵¹ These differences are unlikely to be captured by control variables such as age and gender, and can lead to differences in mean levels of Trump voting that can have major consequences for how the expected shape of in-person versus mail-in ballots across the two counties. Such matching/experimental design assumes that the units (and the individuals who make up that unit) are identical on every meaningful characteristic. The premise is the same as a laboratory experiment when a scientist wants to determine if a treatment has an effect. If the control group varies from the treated group, we cannot know whether the treatment has caused the effect, or if some other difference has a causal impact. Counties, which are the administrators of elections, vary on many aspects of election administration that

the sign of the results and that Lott's specification is one of those that results in a higher level of statistical significance for the dummy variable.

⁵¹ One of the present authors grew up in Pennsylvania in one of the precincts in Lott's data set and is familiar not just with the county in which that precinct is located but also familiar with the adjacent county whose precincts are used for comparison purposes.

make comparisons across counties potential fraught. For instance, the location and number of polling places might differ. The availability and location (accessibility) of drop-boxes for mail-in ballots might have an impact on the propensity of voters to utilize those resources. County administrators might also have different procedures for curing ballots.⁵² Most importantly, the likelihood that a person chooses to vote by mail is correlated highly with vote choice. Even in matched precincts in adjacent counties, there might be significant problems with an independence assumption.

More particularly, implicit in the Lott regression model is that in all precincts, both in counties where Trump wins and in counties where Trump loses, the share of the Trump vote that is cast in-person is independent (at least on average) of the level of Trump support in the precinct. That is, the share of the Trump vote that is cast in person (as opposed to mail) is uncorrelated with Trump's share of the overall vote. Since the total vote in a precinct, V_i , is the sum of the two types of ballots, P_i or M_i , and with $0 \leq P_i + M_i \leq 1$. We can posit that, on average, one (say, in-person, \hat{P}_i), is not correlated with \hat{V}_i (and therefore \hat{M}_i is not correlated with \hat{V}_i).

We show in Figure 3 that the assumption of independence is meritless.

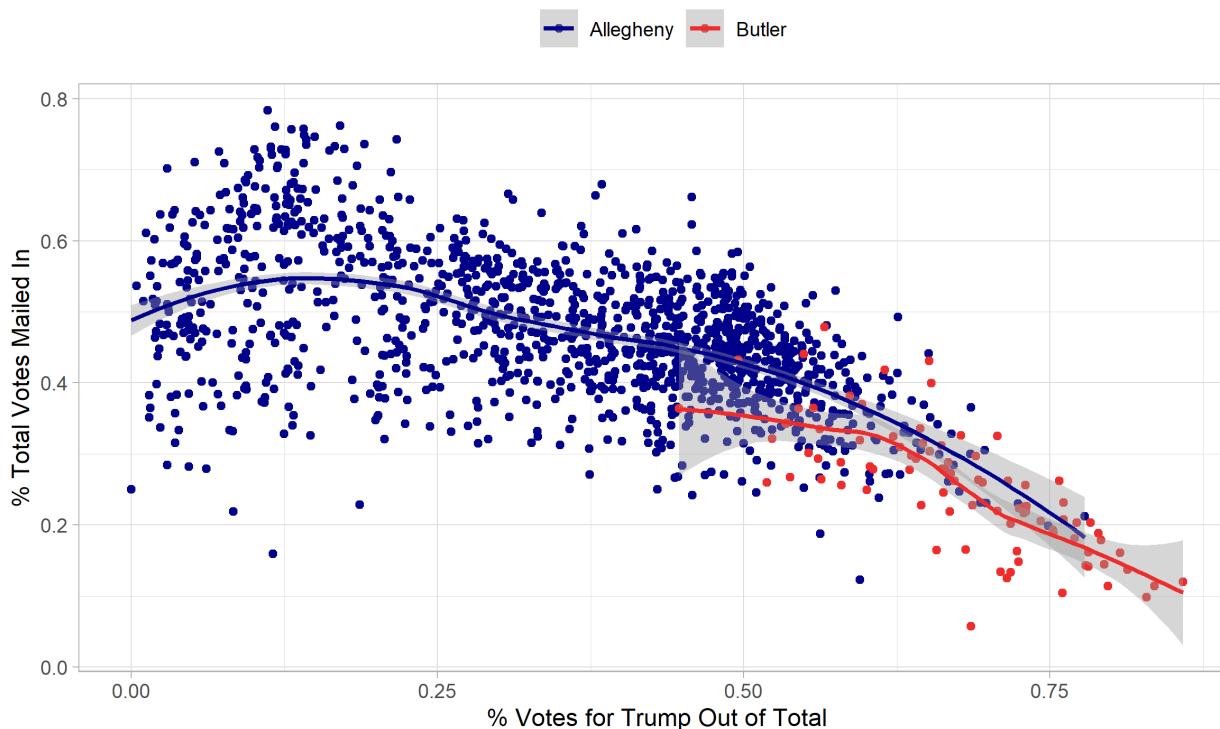
⁵² Counties following different procedures for allowing the correction of what would be invalid ballots in Pennsylvania was challenged and led to a lawsuit. White House Press Secretary Kayleigh McEnany said that the PA Secretary of State allowed ballot curing to “tip the scales of an election to functionally favor the Democrat Party.” But all counties were provided the opportunity to cure ballots, even if some refused to. For more information in the differential treatment of invalid ballots across PA counties, see (Farley, 2020).

If this assumption, which is critical to Lott's argument were to hold, then we can write the in-person vote share difference between the precincts (weighed by the population in the precincts) in a pro-Trump county and those in a pro-Biden county as $\hat{P} * (V_1 - V_2)$, where V is the Trump share of each county's vote. We can likewise write the mail-in vote difference between the precincts in the pro-Trump county and those in the pro-Biden county as $\hat{M} * (V_1 - V_2)$. Hence, we can form a regression with Trump mail-in vote share in County 1 minus Trump mail-in vote share in County 2 as the dependent and Trump in-person vote share in County 1 minus Trump in-person vote share in County 2 as the independent variable. But, in accord with the homogeneity assumption above, this leads us to the regression $y = [(\hat{P}/\hat{M})]x$.

In this regression, because of the posited linearity, we should find an intercept of zero. Lott interprets a non-zero intercept as evidence of fraud. But any non-uniformity in the posited linkages can produce a non-zero intercept. Below we demonstrate this non-uniformity for precincts in Allegheny and Butler Counties in Pennsylvania, two of the counties analyzed by Lott. Indeed, what we find is that the higher the Republican vote share in the precinct, the higher the proportion of voters who cast in-person votes. Moreover, the two counties have similar levels of in-person voting at similar precinct Trump support. We interpret this as a result of the more highly Republican areas having been more effectively cautioned by party leaders against the use of mail-in ballots.

Figure 3 – Trump mail-in vote by level of total support

Absentee % Vs Trump Vote; Allegheny & Butler, Pennsylvania
By Precinct, 2020 Presidential Election



Note: Each dot is one precinct. There are many more precincts in Allegheny County than Butler County.

On the other hand, even if we reject the various rebuttals of the appropriateness of Lott's matched pairs design and take as reasonable that the sign of the dummy variable tells us something about fraud, the evidence that the dummy variable he uses as an indicator of fraud is non-zero is not that strong. As noted earlier, Lott obtains statistically significant results for the dummy variable only for data pooled across two very different states. But that pooling can in principle create a fallacy akin to *Simpson's paradox*. If the two states differ, say, in their partisan propensities in the chosen counties, then an effect

(e.g., a positive sign on the dummy variable) that is due to differences between the states can be attributed to within-county differences in each state.⁵³

III. Conclusions

To paraphrase Jeremy Bentham, claims of massive fraud based on aggregate level statistical features of the 2020 election are not just nonsense, but “nonsense on stilts” (Bentham, 2002). While it is impossible to address all the misleading claims and specious arguments made on the internet or even by President Trump himself, we believe the compendium of statistical fallacies given above can be useful to those interested in the misuse of statistics. And, as noted earlier, this essay is deliberately written in a non-technical way to be comprehensible even to beginning students in statistics.

⁵³ Grofman (1989) looks at an analogous issue. Pooling data across nations he finds that minimum-winning cabinet coalitions are considerably more durable, on average than non-minimum-winning cabinet coalitions, i.e., it takes longer until the coalition is dissolved. And yet, within each country, he finds that minimum winning coalitions are no more durable than non-minimum winning coalitions. Rather, he finds that countries with more durable coalitions also have a higher-than-average proportion of minimal winning coalitions.

IV. References

- Alba, D., & Frenkel, S. (2021, January 7). Six False Claims About the November Election. *The New York Times*, A14.
<https://www.nytimes.com/2021/01/06/business/election-fraud.html>
- Ayyadurai, S. (2020, November 10). *Dr.SHIVA LIVE: MIT PhD Analysis of Michigan Votes Reveals Unfortunate Truth of U.S. Voting Systems*.
<https://www.youtube.com/watch?v=Ztu5Y5obWPk>
- Bentham, J. (2002). Rights, Representation, and Reform: Nonsense upon Stilts and Other Writings on the French Revolution. In P. Schofield, C. Pease-Watkin, & C. Blamires (Eds.), *The Collected Works of Jeremy Bentham* (1st ed., pp. 317–401). Oxford University Press.
- Berlinski, N., Doyle, M., Guess, A. M., Levy, G., Lyons, B., Montgomery, J. M., Nyhan, B., & Reifler, J. (2021). The Effects of Unsubstantiated Claims of Voter Fraud on Confidence in Elections. *Journal of Experimental Political Science*, 1–16. <https://doi.org/10.1017/XPS.2021.18>
- Better Know a Ballot.* (2021). A Late Show with Stephen Colbert.
<https://www.betterknowaballot.com/>

Bliss, L., & Patino, M. (2020). *How to Spot Misleading Election Maps*. Bloomberg CityLab. <https://www.bloomberg.com/news/articles/2020-11-03/a-complete-guide-to-misleading-election-maps>

Bump, P. (2022a, February 8). The bizarre voter-fraud hunt in a New Mexico county Trump won by 25 points - The Washington Post. *The Washington Post*. <https://www.washingtonpost.com/politics/2022/02/08/bizarre-voter-fraud-hunt-new-mexico-county-trump-won-by-25-points/>

Bump, P. (2022b, July 14). We have reached the apex of election-fraud debunking. *The Washington Post*. <https://www.washingtonpost.com/politics/2022/07/14/we-have-reached-apex-election-fraud-debunking/>

Campbell, J. E. (1986). Presidential Coattails and Midterm Losses in State Legislative Elections. *American Political Science Review*, 80(1), 45–63. <https://doi.org/10.2307/1957083>

Cartographic Views of the 2020 US Presidential Election - Worldmapper. (2020, November 27). World Mapper. <https://worldmapper.org/us-presidential-election-2020/>

Cervas, J. R., & Grofman, B. (2019). Are Presidential Inversions Inevitable? Comparing Eight Counterfactual Rules for Electing the U.S. President*. *Social Science Quarterly*, 100(4), 1322–1342. <https://doi.org/10.1111/ssqu.12634>

Corasaniti, N., Epstein, L., & Rutenberg, J. (2020, November 11). Election Officials Nationwide Find No Fraud. *The New York Times*, A1. <https://www.nytimes.com/2020/11/10/us/politics/voting-fraud.html>

Deckert, J., Myagkov, M., & Ordeshook, P. C. (2011). Benford's Law and the Detection of Election Fraud. *Political Analysis*, 19(3). https://www-jstor-org.cmu.idm.oclc.org/stable/23011436?searchText=Benford%27s+Law+and+the+Detection+of+Election+Fraud&searchUri=%2Faction%2FdoBasicSearch%3FQuery%3DBenford%2527s%2BLaw%2Band%2Bthe%2BDetection%2Bof%2BElection%2BFraud%26so%3Drel&ab_segments=0%2Fbasic_search_gsv2%2Fcontrol&refreqid=fastly-default%3A5a2d6bef5d671beae4e916217d563b45#metadata_info_tab_contents

Deutsch, K. W., & Madow, W. G. (1961). A note on the appearance of wisdom in large bureaucratic organizations. *Behavioral Science*, 6(1), 72–78. <https://doi.org/10.1002/bs.3830060108>

DNC Launches New Digital Ads in PA Reaching Vote-By-Mail Voters: "How to Return Your Ballot!". (2020). Democratic National Convention. <https://democrats.org/news/dnc-launches-new-digital-ads-in-pa-reaching-vote-by-mail-voters-how-to-return-your-ballot/>

Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., & Deravi, F. (2019). Understanding Conspiracy Theories. *Political Psychology*, 40(S1), 3–35. <https://doi.org/10.1111/POPS.12568>

Downs, A. (1957). *An Economic Theory of Democracy*. Harper & Row.

Edsall, T. B. (2022, June 22). Why Conspiracy Theories Flourish in Trump's America. *The New York Times*.
<https://www.nytimes.com/2022/06/22/opinion/trump-conspiracy-theories.html>

Eggers, A. C., Garro, H., & Grimmer, J. (2021). No evidence for systematic voter fraud: A guide to statistical claims about the 2020 election. *Proceedings of the National Academy of Sciences*, 118(45).
<https://doi.org/10.1073/pnas.2103619118>

Engstrom, E. J. (2020). Congressional Elections: Electoral Structure and Political Representation. In J. L. Carson & M. S. Lynch (Eds.), *New Directions in American Politics*. Taylor & Francis.
<https://books.google.com/books?id=-CnZDwAAQBAJ>

Farley, R. (2020, November 13). Ballot “Curing” in Pennsylvania - FactCheck.org. *FactCheck.Org*. <https://www.factcheck.org/2020/11/ballot-curing-in-pennsylvania/>

Feldman, M. (2020). *10 Voter Fraud Lies Debunked* Informed citizens are our. <https://www.brennancenter.org/our-work/research-reports/10-voter-fraud-lies-debunked>

Fiorina, M. P. (2017). *Unstable Majorities*. Hoover Institution Press.

Foley, E. B., & Stewart III, C. (2020). *Explaining the Blue Shift in Election Canvassing*.

Frey, W. H. (2021, January 21). Biden-won counties are home to 67 million more Americans than Trump-won counties. *Brookings*.
<https://www.brookings.edu/blog/the-avenue/2021/01/21/a-demographic-contrast-biden-won-551-counties-home-to-67-million-more-americans-than-trumps-2588-counties/>

Galston, W. A. (2020, December 21). Why did House Democrats underperform compared to Joe Biden? *Brookings*.
<https://www.brookings.edu/blog/fixgov/2020/12/21/why-did-house-democrats-underperform-compared-to-joe-biden/>.

Gardner, A. (2021, July 5). In ramp-up to 2022 midterms, Republican candidates center pitches on Trump's false election claims - The Washington Post. *The Washington Post*. https://www.washingtonpost.com/politics/republican-trump-2022-midterms-election-falsehoods/2021/07/04/3a43438c-d36f-11eb-ae54-515e2f63d37d_story.html

Gelman, A. (2021, August 21). Alex Jones and the fallacy of the one-sided bet. *Statistical Modeling, Causal Inference, and Social Science*.
<https://statmodeling.stat.columbia.edu/2021/08/21/alex-jones-and-the-fallacy-of-the-one-sided-bet/>

Gray, J. (2020, December 4). Georgia election officials show frame-by-frame what happened in Fulton surveillance video – WSB-TV Channel 2 - Atlanta. *WSB-TV 2 Atlanta*. <https://www.wsbtv.com/news/politics/georgia-election-officials-show-frame-by-frame-what-really-happened-fulton-surveillance-video/T5M3PYIBYFHFFOD3CIB2ULDVDE/>

Grofman, B. (1989). The Comparative Analysis of Coalition Formation and Duration: Distinguishing Between-Country and Within-Country Effects. In *Journal of Political Science* (Vol. 19, Issue 2). <https://about.jstor.org/terms>

Grofman, B. (2022). Prospects for Democratic Breakdown in the United States: Bringing the States Back In. *Perspectives on Politics*, 1–8. <https://doi.org/10.1017/S1537592721003285>

Grofman, B., & Chen, H. (2022). Understanding the Factors that Affect the Incidence of Bellwether Counties: A Conditional Probability Model. *Political Research Quarterly*. <https://doi.org/10.1177/10659129211057601>

Grofman, B., & Garcia, J. R. (2014). Using Spanish Surname to Estimate Hispanic Voting Population in Voting Rights Litigation: A Model of Context Effects Using Bayes' Theorem. *Election Law Journal: Rules, Politics, and Policy*, 13(3), 375–393. <https://doi.org/10.1089/elj.2013.0190>

Grofman, B., & Garcia, J. R. (2015). Using Spanish Surname Ratios to Estimate Proportion Hispanic in California Cities via Bayes Theorem*. *Social Science Quarterly*, 96(5), 1511–1527. <https://doi.org/10.1111/ssqu.12214>

Hasen, R. L. (2020). *Election Meltdown: Dirty Tricks, Distrust, and the Threat to American Democracy*. Yale University Press.

<https://www.amazon.com/Election-Meltdown-Distrust-American-Democracy/dp/0300248199>

Holman, M. R., & Lay, J. C. (2018). They See Dead People (Voting): Correcting Misperceptions about Voter Fraud in the 2016 U.S. Presidential Election.
<Https://Doi.Org/10.1080/15377857.2018.1478656>, 18(1–2), 31–68.
<https://doi.org/10.1080/15377857.2018.1478656>

Homans, C., & Peterson, M. (2022, July 24). How ‘Stop the Steal’ Captured the American Right - The New York Times. *The New York Times*.
<https://www.nytimes.com/2022/07/19/magazine/stop-the-steal.html>

Hopkins, D. A. (2017). *Red Fighting Blue*. Cambridge University Press.
<https://doi.org/10.1017/9781108123594>

Hsu, T., & Thompson, S. A. (2022, August 10). Hunting for Voter Fraud, Conspiracy Theorists Organize ‘Stakeouts’ - The New York Times. *The New York Times*. <https://www.nytimes.com/2022/08/10/technology/voter-drop-box-conspiracy-theory.html>

Jacobson, G. C. (2019). Extreme Referendum: Donald Trump and the 2018 Midterm Elections. *Political Science Quarterly*, 134(1), 9–38.
<https://doi.org/10.1002/polq.12866>

Jenny. (2020, November 7). *Joe Biden's votes violate Benford's Law*. 0–1.

<https://web.archive.org/web/20220417144342/https://gnews.org/534248/>

Kabir, N. (2020, November 14). The Fraud Of Dr. Shiva Ayyadurai: Oakland County, Michigan | by Naim Kabir | Medium. *Medium*. <https://naim-kabir.medium.com/the-fraud-of-dr-shiva-ayyadurai-oakland-county-michigan-1bc51bcebf1b>

Kessler, G. (2020, December 18). The biggest Pinocchios of 2020. *The Washington Post*.

<https://www.washingtonpost.com/politics/2020/12/18/biggest-pinocchios-2020>

Lai, J. (2020, September 21). How 'naked ballots' in Pennsylvania could cost Joe Biden the election. *The Philadelphia Inquirer*.

Lee, F. E. (2016). *Insecure Majorities: Congress and the Perpetual Campaign*. University of Chicago Press.

Lenz, G. S. (2012). *Follow the Leader? How Voters Respond to Politicians' Policies and Performance*. University of Chicago Press.
<https://press.uchicago.edu/ucp/books/book/chicago/F/bo11644533.html>

Leonhardt, D. (2022, July 19). Bigger than Trump: Election Conspiracy Theories Have Become Central to the Republican Party. *The New York Times*.
<https://www.nytimes.com/2022/07/19/briefing/stop-the-steal-anti-democracy.html>

Lott, J. R. (n.d.). Simple tests for the extent of vote fraud with absentee and provisional ballots in the 2020 US presidential election John. *Public Choice*.

Lott, J. R. (2020). *A Simple Test for the Extent of Voter Fraud with Absentee Ballots in the 2020 Presidential Election: Georgia and Pennsylvania Data*.
<http://dx.doi.org/10.2139/ssrn.3756988>

Matsumoto, R. (2021, February 1). Where Did All The Bellwether Counties Go? *FiveThirtyEight*. <https://fivethirtyeight.com/features/where-did-all-the-bellwether-counties-go/>

McClallen, S. (2021, June 17). Over 7,000 affidavits delivered to Michigan lawmakers claim election fraud | Michigan | thecentersquare.com. *The Center Square*. https://www.thecentersquare.com/michigan/over-7-000-affidavits-delivered-to-michigan-lawmakers-claim-election-fraud/article_78b6812c-cf98-11eb-868e-734c5e3a51de.html

McDade, A. (2022, February 1). Ohio Secretary of State Touts Security of Election Process After Referring 27 Fraud Cases. *Newsweek*. <https://www.newsweek.com/ohio-secretary-state-touts-security-election-process-after-referring-27-fraud-cases-1675215>

McDonald, M. P., & Levitt, J. (2008). Seeing Double Voting: An Extension of the Birthday Problem. *Election Law Journal: Rules, Politics, and Policy*, 7(2), 111–122. <https://doi.org/10.1089/elj.2008.7202>

Mebane, W. R. (2011). Comment on “Benford’s Law and the Detection of Election Fraud.” *Political Analysis*, 19(3), 269–272.
<https://doi.org/10.1093/pan/mpr024>

Mebane, W. R. (2020). *Inappropriate Applications of Benford’s Law Regularities to Some Data from the 2020 Presidential Election in the United States* *. 1045. https://github.com/cjph8914/2020_benfords

Mebane, W. R., & Kalinin, K. (2009). Comparative Election Fraud Detection. *APSA 2009 Toronto Meeting Paper*, 1045.
https://papers.ssrn.com/Sol3/papers.cfm?abstract_id=1450078

Medina, J., Epstein, R. J., & Corasaniti, N. (2022, August 4). In 4 Swing States, G.O.P. Election Deniers Could Oversee Voting - The New York Times. *The New York Times*. <https://www.nytimes.com/2022/08/03/us/politics/gop-election-deniers-trump-arizona-michigan.html>

Parker, M. (2020, November 14). *Do these scatter plots reveal fraudulent vote-switching in Michigan?* - YouTube. YouTube.
<https://www.youtube.com/watch?v=aokNwKx7gM8>

Reality Check team. (2020, November 23). *US election 2020: Fact-checking Trump team’s main fraud claims* - BBC News. BBC News.
<https://www.bbc.com/news/election-us-2020-55016029>

Russian elections once again had a suspiciously neat result. (2021, October 11). *The Economist*. <https://www.economist.com/graphic->

[detail/2021/10/11/russian-elections-once-again-had-a-suspiciously-neat-result](https://www.politico.com/article/2021/10/11/russian-elections-once-again-had-a-suspiciously-neat-result)

Sadeghi, M. (2020, November 14). Fact check : No evidence vote was cast in

Joe Frazier's name. USA Today.

<https://www.usatoday.com/story/news/factcheck/2020/11/14/fact-check-no-evidence-late-joe-frazier-voted-2020-election/6283956002/>

Sganga, N. (2022, June 3). U.S. finds no evidence flaws in Dominion voting

machines were ever exploited - CBS News. CBS News.

<https://www.cbsnews.com/news/dominion-voting-machines-2020-election-cisa/>

Shurk, J. B. (2020, November 23). 5 More Ways Joe Biden Magically

Outperformed Election Norms. The Federalist.

<https://thefederalist.com/2020/11/23/5-more-ways-joe-biden-magically-outperformed-election-norms/>

Solender, A. (2020, December 10). These Are The Republicans Who Have

Acknowledged Biden's Victory. Forbes.

<https://www.forbes.com/sites/andrewsolender/2020/12/10/these-are-the-republicans-who-have-acknowledged-bidens-victory/?sh=7e71eaea417b>

States That Permit Voters to Correct Signature Discrepancies. (2020). National Conference of State Legislatures. <https://www.ncsl.org/research/elections-voting/signature-discrepancies.aspx>

[and-campaigns/vopp-table-15-states-that-permit-voters-to-correct-signature-discrepancies.aspx](https://www.ncsl.org/research/elections-and-campaigns/vopp-table-15-states-that-permit-voters-to-correct-signature-discrepancies.aspx)

Swenson, A. (2020, December 29). There were not more votes than voters in Pennsylvania. *AP News.* <https://apnews.com/article/fact-checking-afs:Content:9887147615>

Swenson, A. (2021, December 21). Winning more counties doesn't translate to an election win for Trump. *Associated Press.*

Tufte, E. R. (1974). *Data Analysis for Politics and Policy* (R. A. Dahl, Ed.). Prentice-Hall.

UMass Amherst. (2021, December 38). *Presidential Election & Jan 6th Insurrection at the US Capitol.* UMass Amherst. <https://polisci.umass.edu/toplines-and-crosstabs-december-2021-national-poll-presidential-election-jan-6th-insurrection-us>

Vella, V. (2020, December 21). Delaware County man charged with registering dead relatives to vote in presidential election. *The Philadelphia Inquirer.* <https://www.inquirer.com/news/bruce-bartman-election-fraud-delaware-county-20201221.html>

Voting Outside the Polling Place: Absentee, All-Mail and other Voting at Home Options. (2022, March 15). National Conference of State Legislatures. <https://www.ncsl.org/research/elections-and-campaigns/absentee-and-early-voting.aspx>

Weiser, W. r., Levitt, J., & Munoz, A. (2006). *Making the List: Database Matching and Verification Processes for Voter Registration* | Brennan Center for Justice. <https://www.brennancenter.org/our-work/research-reports/making-list-database-matching-and-verification-processes-voter>

Zorn, E. (2017, January 5). Polls reveal sobering extent of nation's fact crisis -

Chicago Tribune. *Chicago Tribune.* <https://www.chicagotribune.com/columns/eric-zorn/ct-polling-ignorance-facts-trump-zorn-perspec-0106-md-20170105-column.html>