Statistical Fallacies in the Claims about Massive Election Fraud in 2020: A Compendium

Journal:	*Statistics and Public Policy*
Manuscript ID	USPP-2021-0016
Manuscript Type:	Paper
Date Submitted by the Author:	09-Jun-2021
Complete List of Authors:	grofman, bernard; University of California Irvine, political science Cervas, Jonathan; Carnegie Mellon University, strategy and politics
Keywords:	fraud, elections, 2020 presidential contest
Abstract:	The amount that has been uttered/written about the assertions of massive electoral fraud in the 2020 presidential election is immense. Here, we do not discuss the factual lies and misinformation spread by proponents of the theory of massive fraud. Rather, we limit ourselves to statistical features of the election whose occurrence is essentially undisputed but whose implications for the presence of massive fraud are completely 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 useful compendium of recent glaring misuses of statistics.

Statistical Fallacies in the 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

*This research was partially supported by the Peltason Chair of Democracy Studies, University of California, Irvine

ABSTRACT

The amount that has been uttered/written about the assertions of massive electoral fraud in the 2020 presidential election is immense. Here, we do not discuss the factual lies and misinformation spread by proponents of the theory of massive fraud. Rather, we limit ourselves to statistical features of the election whose occurrence is essentially undisputed but whose implications for the presence of massive fraud are completely 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 useful compendium of recent glaring misuses of statistics.

I. Introduction

Our concern here is with claims of massive election fraud in the 2020 presidential election that are grounded, at least in part, on indisputable facts about statistical features of the election. We leave to fact checkers, election officials, and other academics the rebuttal of the vast mass of near infinitely repeated lies or half-truths, and occasional just plain misunderstandings about the ways in which the 2020 election was conducted and its results tabulated. And we leave claims that particular election practices violated that state's own constitution where they belong, namely in litigation in state courts that President Trump's attorneys lost with complete (and predictable(regularity. Instead we focus on (a) statistical fallacies of a simple sort[1] such as confusing percentages and percentage point changes, or failing to recognize that large changes in one direction in low population units of political geography or low population demographic subgroups can be compensated for by small changes in the other direction in high population units of political geography or demographic groupings,[2] and cherry-picking the data to emphasize only those facts that lead to the desired conclusion, and (b) syllogistic arguments that are either fallacious in form, or that have at least one premise that is indubitably false even while

[1] There has been at least one apparently sophisticated attempt to prove election fraud via statistical analysis by using comparisons in neighboring jurisdictions in different states that operated under different election tallying procedures: Lott, John R. "A Simple Test for the Extent of Voter Fraud with Absentee Ballots in the 2020 Presidential Election: Georgia and Pennsylvania Data," Unpublished Manuscript. https://papers.ssrn.com/sol3/papers.cfm?abstract id=3756988, 2021. But flaws in the econometric design in that study that completely vitiate its claims are convincingly demonstrated in Eggers, Andrew C., Haritz Garro and Justin Grimmer. "Comment on 'A Simple Test for the Extent of Voter Fraud with Absentee Ballots in the 2020 Presidential Election'." Unpublished Manuscript, January 4, 2021. Professor Lott has redone some parts of his analysis and claims to still prove fraud, but we are skeptical.

[2] For example, to state the obvious, the presidential candidate who wins the most counties do not necessarily win the election. Americans are also familiar with how the Electoral College can lead to "inversions," such that even winning more votes does not necessarily lead to electoral victory (Cervas and Grofman 2019).

another premise may be quite factual, and thus which give rise to invalid or unfounded conclusions. Our goal is not to provide new insights, since the various rebuttals we give have been given by others in various scattered sources, but rather to put together in one place a useful compendium of the most glaring recent misuses of election statistics.

We begin our analytic review with claims that have as their general form: "The only way these election results could have happened is if there was massive fraud."[3]

We start with two of the most nonsensical of such claims.[4] The first is based on comparisons between what was found in 2020 and what was found in 2016. The second is based on an internal comparison of ballots in 2020 cast that are tallied early and those that are tallied later. Each is based on observations found in 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). [5]

First, 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 improbably degree. Well, he's certainly right about that fact. But what that shows

[3] The examples we give are intended to be illustrative, and not at all exhaustive.

[4] Choosing which of the claims is the silliest is not at all an easy call.

[5] While these two claims have already been rebutted elsewhere (https://reason.com/volokh/2020/12/09/more-on-statistical-stupidity-at-scotus/ and https://statmodeling.stat.columbia.edu/2020/12/08/the-p-value-is-4-76x10%E2%88%92264-1-in-a-quadrillion/, they are so entertainingly stupid that we could not resist sharing them. These two arguments are not the only bizarre aspects of Dr. Cicchetti's expert witness report. 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

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. Indeed, President Trump, now running as an incumbent, had a lower share of the two-party vote in 2020 than he did in 2016.

Dr. Cicchetti also found 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. 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 significant beyond any reasonable doubt. And about that fact he is quite correct. In states where the late-tallied ballots were disproportionately mail-in ballots, they were disproportionately Democratic. In states where the late-tallied ballots were disproportionately in person ballots, they were disproportionately Republican. Dr. Cicchetti apparently took the fact that the mail-in votes were more Democratic than in-person votes to be evidence that the mail-in ballot reflected pro-Biden vote fraud. The best way to characterize this argument is:

> If there were fraud that disproportionately affected mail-in votes and tilted them in a pro-Biden direction (A),
>
> then there would be a difference between mail-in votes and in-person votes, with Biden doing better in the latter (B).
>
> There was such a difference in the predicted direction (B).
>
> Therefore, there was fraud (A).

It is easy to see that this argument is simply an example of the well-known logical fallacy of *affirming the consequent.*

A difference in support among voters who vote in person and voters who vote by mail proves nothing about fraud. There is no reason to think that the mail ballots and the in-person ballots should have identical distributions of Trump votes. The likelihood of casting a ballot by mail varies across areas within a state, with Democratic leaning areas (cities and more populous

counties) more likely to cast mail ballots Consequently, it was expected well in advance of the final tabulations in 2020 that, in most states, mail-in ballots would disproportionally tend Democratic (Foley 2013).. This pattern was also found in 2016, and moreover, was predicted in advance of the 2020 election (Foley and Stewart 2020). Similarly, comparing the vote distribution in 2016 to that in 2020 is not an appropriate hypothesis test. That they are different only proves one thing, that we should not expect the same outcome; not fraud!

Other Cross-Election Comparisons

Other arguments 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. 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. Moreover, only five incumbents have received at least 90% of their primary vote.[6] Shurk also observes that "no incumbent in over 100 years who has gained votes in his reelection bid has lost his quest for reelection.[7] But of course, it is not total votes but share of the votes and, for the Electoral College, distribution of the vote, that matters.

It would be equally valid to claim that incumbents whose overall presidential approval rating was as low as that of President Trump were very likely to lose re-election. Therefore,

[6] 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

[7] This claim fails to account for well-known political science theory that suggests that competitive elections increase voter turnout (Downs 1957), and that modern elections are increasingly competitive (Lee 2016).

Trump almost certainly *must* have lost. Or a similar argument could show why neither Trump nor Hillary Clinton could possibly have been elected President in 2016 since both were so disliked. Unfortunately for this type of argument, there is no guarantee that what was true in the past will be true in the future. And it is helpful to remember the fundamental rule of two-party elections that, no matter how unpopular you are, the other person might be even more unpopular. And, of course, all 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.

Presidential Coattails

A related fallacious argument is 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, so Biden must not have won the election, and thus 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 is built, that presidential coattails are inevitable, is false. By coattails, we are referring to increasing the number of members in the U.S. House of Representatives that share the incoming president's party (Campbell 1986). Actually, negative coattails are not uncommon, and in contemporary politics, have become more likely.[8] Since 1868, there have been thirteen

[8] See Morris Fiorina (2016), *Unstable Majorities*, and Frances Lee (2016), *Insecure Majorities*. Modern elections are very likely to result in divided government, and control of

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 (d) there are many more seats won by the previous president in the election four years earlier which have a sitting House member of the other party than there are seats lost by the previous president in the elections four years earlier which have a sitting House member of his party. 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%.[9] 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). 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 -- we would not expect to see Biden coattails in those remaining states.[10] Democratic gains in the House in the 2018 midterm were significant, and turnout was a level not seen before universal adult franchise (Jacobson 2019). There were 35 House constituencies carried by Trump in 2016 but with a Democratic House member elected in

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.

[9] For more details on Biden's overperformance compared to U.S. House Democratic candidates, see William A. Galston, "Why did House Democrats underperform compared to Joe Biden?", December 21, 2020, *Brookings*, https://www.brookings.edu/blog/fixgov/2020/12/21/why-did-house-democrats-underperform-compared-to-joe-biden/.

[10] House Democrats nationally underperformed their 2018 performance, which accounts for the eleven net seats gained by the Republicans. (As of this writing on February 7, 2021, one seats has yet to be called, though the Republican candidate leads by 12 votes). Relative to the 115th Congress (2016-2018), the 177th Congress (2021-2023) has 28 more Democrats.

2018,[11] and only 5 House constituencies lost by Trump in 2016 but with a Republican House member elected in 2018.[12] Thus, Democrats in 2020 had many more vulnerable House seats than did the Republicans. 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.[13]

Bellwether counties

Another similar argument 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 actually a valid argument -- another example of *denying the consequent*. However, even though "not B" is empirically accurate, once again, the premise on which the argument is built, namely that bellwethers predict elections, is false. Many decades ago, Tufte. Tufte (1974, chapter 3; 1975) 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

[11] An increase of 22 from 2016.

[12] A decrease by 19 from 2016.

[13] Figure omitted for space reasons.

track record at the next election than the non-bellwether counties. And yet, belief in bellwether units of geography, more particularly in the existence of bellwether counties, refuses to die.

Spoiled Ballots

A further example of fallacious argument based on comparisons of past and present election results is based on the empirically accurate observation that the spoiled ballot rate of mail-in ballots was much lower (sometime more than an order of magnitude lower) than in 2016. This fact is taken to be evidence of mail-ballot fraud by Trump supporters. But it is not.

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 once again have the fallacy of *affirming the consequent*.

There are good reasons why ballot spoilage was lower in 2020 than in 2016, 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.[14] 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

[14] "Better Know a Ballot", Accessed January 29, 2021, *A Late Show with Stephen Colbert,* https://www.betterknowaballot.com/.

themselves aired on television channels and as ads on streaming services.[15] The Democratic National Committee also spent millions of dollars on television ads with information about returning mail-in ballots.[16] There was also ample coverage in newspapers about properly filing out and mailing a ballot so that it would not be rejected.[17]

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.[18] These states are disproportionately Democratic; Trump won just 5 of the 18. But in our federal system, absent issues that raise genuine constitutional questions, states can and do differ in the details of their election administration.

The next fallacies involve some very basic statistical mistakes.

Weighted-unit statistical fallacy

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

[15] In Pennsylvania, this included pointing viewers to a website developed by the Department of State. Ads featured prominent actors and athletes from the state. https://www.votespa.com/Pages/default.aspx.

[16] "DNC Launches New Digital Ads in PA Reaching Vote-By-Mail Voters: "How to Return Your Ballot!"", September 24, 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/.

[17] Lai, Jonathan. 2020. "How 'Naked Ballots' in Pennsylvania Could Cost Joe Biden the Election." *The Philadelphia Inquirer*.

[18] "States That Permit Voters to Correct Signature Discrepancies", September 21, 2020, *National Conference of State Legislatures*, https://www.ncsl.org/research/elections-and-campaigns/vopp-table-15-states-that-permit-voters-to-correct-signature-discrepancies.aspx.

direction in high population subsets or demographic units is a common mistake. This type of error can be seen visually when comparing a choropleth map of election results with a cartogram -- a map that has been re-sized so the units' area is equal to its population weight.[19]

There were various instances of this types of error. For example, 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.[20] But, of course that is nonsense, since he could have done better in the remaining counties and these counties, though fewer in number, had more voters in them. Indeed, Biden received over three million votes in Los Angeles County, alone.[21] In fact, Biden net an additional 609,000 more votes in 2020 than Clinton did in 2016, just in this one county!

Cherry-picking Data

A closely related fallacy, cherry-picking the data to emphasize only those facts that lead to the desired conclusion, is a strategy employed often by those who know that the full evidence is against them. In employing only the partial story, a claimant can appear to have been honest while suppressing pertinent information that otherwise would prove their evidence either false, or incomplete.

[19] For a nice explanation of displaying political graphics of different kinds, see Bliss, Laura & Marie Patino, "How to Spot Misleading Election Maps", November 3, 2020, *Bloomberg CityLab*, https://www.bloomberg.com/news/articles/2020-11-03/a-complete-guide-to-misleading-election-maps. See also "Cartographic Views of the 2020 US Presidential Election, November 27, 2020, *World Mapper*, https://worldmapper.org/us-presidential-election-2020/.

[20] See Swenson, Ali, "Winning more counties doesn't translate to an election win for Trump", December 21, 2021, *Associated Press*, https://apnews.com/article/fact-checking-afs:Content:9848943909.

[21] Out of the over 3,000 counties in the United States, the top 150 contained half of the total votes casted. Biden won 125 of those 150 (83.3%).

For example, it is true that racial minority support for Biden was lower than for Hillary Clinton. However, support among categories of white voters, namely the college educated and those living in suburbs, was higher for Biden than for Hillary Clinton. Similarly, there were some urban areas where support for Biden was lower than for Hillary Clinton but there were also urban areas where support for Biden was higher than for Hillary Clinton. Moreover, decreases (increases) in the *percentage* of votes from given group or area can be offset by increases (decreases) in the number of raw votes. It is the combination of all the subgroup patterns of voting and the pattern of their geographic location as it affects Electoral College outcomes that matters. Looking only at some subsets of voters, or only some geographic areas, is misleading and can lead to ridiculous claims that the candidate who received more votes did not actually receive more votes.

Percentages and percentage point changes

Relatedly, a similar percentage of increased support among minority voters and decreased support among white voters does not have the same consequences, since white voters make up a much larger percentage of the electorate. Thus, even though Trump made gains among minority voters, those were not enough to give him the election. Consider a very simple example. Had Trump received 5 votes of Hispanic voters and Clinton 95, with total turnout of 100, Trump will have received 5% of the Hispanic vote. If, in the next election Trump received 10 votes to Biden's 110, so that his vote tally from Hispanics doubles (+100%), he would have now have received 8.3% of the Hispanic vote. However, as can be readily observed, the net vote margin *increases* for Biden by 10 votes in the second election as compared to the first. This statistical fallacy is easy to miss since results are sometimes presented as raw votes and sometimes as percentages, and changes are sometimes given as percentage point changes, and sometimes as percentage changes.

I. Conclusions

While it is impossible to address all the misleading claims and specious arguments made in the blogosphere and by President Trump himself, we believe the compendium of statistical fallacies given above can be useful to those interested in the misuse of statistics. To paraphrase

Jeremy Bentham, as anyone with any sense knows, claims of massive fraud based on statistical features of the 2020 election are not just nonsense, but "nonsense on stilts."

II. References

Bentham, Jeremy, *Rights, Representation, and Reform: Nonsense upon Stilts and Other Writings on the French Revolution*, ed. Schofield, P., Pease-Watkin, C., and Blamires, C., Oxford, 2002 (*The Collected Works of Jeremy Bentham*), pp. 317–401.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. Harper & Row.

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

Fiorina, Morris P. 2017. *Unstable Majorities*. Stanford: Hoover Institution Press.

Foley, E. B. (2013). A Big Blue Shift: Measuring an Asymmetrically Increasing Margin of Litigation The Voting Wars: Elections and the Law from Registration to Inauguration. *Journal of Law & Politics*, *28*(4), 501–546. https://heinonline.org/HOL/P?h=hein.journals/jlp28&i=517

Foley, E. B., & Stewart III, C. (2020). Explaining the Blue Shift in Election Canvassing. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3547734

Jacobson, Gary C. 2019. "Extreme Referendum: Donald Trump and the 2018 Midterm Elections." *Political Science Quarterly* 134(1): 9–38. https://onlinelibrary.wiley.com/doi/abs/10.1002/polq.12866 (January 14, 2020).

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

Sturk, J.B., "5 More Ways Joe Biden Magically Outperformed Election Norms", The Federalist, November 23, 2020, https://thefederalist.com/2020/11/23/5-more-ways-joe-biden-magically-outperformed-election-norms/.

Tufte, Edward R. 1974. *Data Analysis for Politics and Policy*. ed. Robert A. Dahl. Prentice-Hall.