The Extraordinarily Low Rate of Dead People Voting by Mail: Evidence from Washington State Administrative Data*

Jennifer Wu[†], Chenoa Yorgason[†], Hanna Folsz[†], Cassandra Handan-Nader[‡], Andrew Myers[§], Tobias Nowacki[‡], Daniel M. Thompson[¶], Jesse Yoder[‡], and Andrew B. Hall[‖]

Democracy & Polarization Lab, Stanford University

November 19, 2020

Abstract

A common concern about vote-by-mail in the United States is that mail-in ballots are sent to dead people, stolen by bad actors, and counted as fraudulent votes. Studying Washington state's vote-by-mail program, we link counted ballots and administrative death records to estimate the rate at which dead people's mail-in ballots are improperly counted as valid votes, using birth dates from online obituaries to address false positives. Among roughly 4.5 million distinct voters in Washington state (2011-2018), we estimate that there are 14 deceased individuals whose ballots might have been cast suspiciously long after their death, representing 0.0003% of voters. Even these few cases may reflect two individuals with the same name and birth date, or clerical errors, rather than fraud. After exploring the robustness of our findings to weaker conditions for name-matching, we conclude that counting dead people's ballots as votes seems extraordinarily rare in Washington's universal vote-by-mail system.

*For helpful information, the authors particularly thank Stuart Holmes and the Washington Secretary of State's Office. For advice, the authors thank Charles Stewart. For research assistance, the authors thank Sarah Raza.

[†]Ph.D. Student, Department of Political Science, Stanford University

[‡]Ph.D. Candidate, Department of Political Science, Stanford University

[§]Predoctoral Research Fellow, SIEPR, Stanford University

[¶]Assistant Professor, Department of Political Science, UCLA

[‖]Professor, Department of Political Science and Graduate School of Business (by courtesy), Stanford University; Senior Fellow, SIEPR, Stanford University. andrewbhall@stanford.edu.

1 Introduction

One of the most common concerns raised about vote-by-mail in the United States, which has become highly salient during the COVID-19 pandemic, is that ballots sent to dead people could be mailed back and counted as valid votes. In the weeks after the 2020 U.S. election, President Trump has repeatedly claimed that dead people voted in the election,[1] echoing a long-running concern among certain conservative groups.[2] Claims like these about the security of vote-by-mail are important to assess because they call into question the legitimacy of the American electoral system and the 2020 election in particular. Attitudes towards the expansion of vote-by-mail are mixed and have polarized along partisan lines (Lockhart et al. 2020), and voter confidence in vote-by-mail is generally lower than in-person voting (Bryant 2020), which makes evaluating its security especially relevant.[3] Because who votes and who dies are both matters of public record in America, we can evaluate the general claim that dead people's mail-in ballots are regularly voted fraudulently in elections with data.

To do so, we link administrative data on deaths and voter turnout in the state of Washington, one of the most prominent states to administer elections entirely by mail. Among roughly 4.5 million distinct voters in Washington state between 2011 and 2018, when we focus on cases where records match on full name including middle name, we estimate that there are 14 deceased individuals whose ballots were cast suspiciously long after their deaths, representing 0.0003% of voters. Even these few cases may reflect two individuals with the same name and birth date, or clerical errors, rather than fraud. If we relax requirements for matching middle names to accommodate people who may not have middle names, we estimate that there are an additional 43 cases of potential fraud, but these are more likely to be false positives. On the whole, the results suggest that it is extremely rare for dead people's ballots to be counted as votes in Washington's universal vote-by-mail system.

[1] See for example: `https://twitter.com/realDonaldTrump/status/1328830381429288962?s=20`.

[2] See for example: `https://www.heritage.org/election-integrity/commentary/potential-fraud-why-mail-elections-should-be-dead-letter`

[3] On the effects of universal vote-by-mail, see for example Gerber, Huber, and Hill (2013) and Thompson et al. (2020).

Our findings should not be surprising to people who follow election administration closely. Indeed, the state of Washington carries out a better version of our analysis on a regular basis, checking voter rolls against death records using additional data like the last four digits of individuals' social security numbers that we do not have access to. The point of our analysis is to provide an independent verification of the state of Washington's process.

Our work adds to the large literature on voter fraud in American elections by quantifying the amount of voter fraud related to dead people's ballots specifically in the context of universal vote-by-mail, where concerns about this fraud have become particularly salient. In studying this form of voter fraud, we build directly on Hood III and Gillespie (2012), a study which combines automated and manual matching methods to quantify the rate of deceased voters' ballots being improperly counted in the 2006 general election in Georgia (not a universal vote-by-mail state), finding essentially zero cases of this form of fraud.[4] By directly linking administrative data to detect fraud, our study is also related to Goel et al. (2020), which performs a similar analysis to quantify rates of double voting, again finding minuscule rates. Beyond these studies, a much broader literature relies on other forms of data, like reported instances of fraud (e.g. Minnite 2010; Alvarez, Hall, and Hyde 2009; Levitt 2007) or suspicious statistical patterns in aggregate data (e.g Cottrell, Herron, and Westwood 2018; Alvarez, Hall, and Hyde 2009; Mebane 2008), again concluding that various detectible forms of voter fraud seem very rare.[5]

While our analyses suggest that fraud related to dead people's ballots and vote-by-mail is incredibly rare in Washington, our results do not directly speak to rates of fraud in other

[4]Our focus on a longer time period and on assessing a fast-moving debate relevant to the 2020 election comes at the cost of some depth; while Hood III and Gillespie (2012) presents a remarkably deep audit of suspicious cases, making public records requests and ruling out nearly all specific suspicious cases as false positives, we only rule out false positives based on publicly available online data. It is reassuring, then, that our broader analysis of universal vote-by-mail in Washington arrives at a similar conclusion to their deeper analysis for Georgia.

[5]There is also a recent Washington Post/ ERIC (Electronic Registration Information Center) analysis which relies on reported instances of fraud. See `https://www.washingtonpost.com/politics/minuscule-number-of-potentially-fraudulent-ballots-in-states-with-universal-mail-voting-undercuts-trump-claims-about-election-risks/2020/06/08/1e78aa26-a5c5-11ea-bb20-ebf0921f3bbd_story.html`.

states. Washington has spent years developing and honing its process for vote-by-mail. States that expanded voting by mail in 2020 without time to develop the same rigorous processes as Washington could see higher rates of fraud, including higher rates of dead people's ballots being counted. Still, fraud of this form is unlikely to be widespread in any context due to the precautions states take, which we discuss in the context of Washington below.

The purpose of this paper is not to endorse or reject universal vote-by-mail as a policy. There are many considerations in supporting or expanding the use of universal vote-by-mail that go beyond the particular kind of fraud we evaluate. As people continue to debate how America should administer its elections, the particular claim that dead people's mail-in ballots are fraudulently cast and counted as votes at high rates is likely to persist. The purpose of this paper is to evaluate this specific claim with data.

2 Why We Study Washington State

We focus on Washington state because it employs universal vote-by-mail. Every registered voter is mailed a ballot which can be mailed back with pre-paid postage, dropped off at one of many drop box locations, or returned in person to County Elections Offices. Since this is the specific policy that President Trump and others have suggested leads to widespread fraud related to deceased voters' ballots, it makes sense to focus on a state that employs this policy.

While a number of other states also employ universal vote-by-mail, Washington is ideal for our purposes because of the data it offers. Unlike many other states, Washington has made statewide voter file snapshots (voter rolls which include information about every registered voter at a particular point in time, such as unique Voter ID, name, county, and date of birth), as well as statewide voter histories (lists of Voter IDs who have cast a verified (non-rejected) vote, including local elections) since 2011, publicly available to researchers. The

Secretary of State maintains nearly monthly voter file snapshots, which enables us to have a nearly-perfect portrait of who has voted in each election within our period of study.

3 Ballot Security in Washington

To ensure election security, Washington takes a number of steps. Together, these steps likely make it difficult to fraudulently cast a dead person's ballot.

First, ballots are assigned unique barcodes which allows voters to track their ballots online. This also allows the state to confirm that the returned ballot corresponds to a specific entry in the voter registration database, and is intended as one of a number of countermeasures to prevent people mailing in fake ballots, as they cannot duplicate these unique barcodes. As a result, the first step a fraudulent actor would have to take to vote in the name of a dead person is to obtain their actual ballot. Doing this at any meaningful scale would require knowing when specific ballots have been mailed and where they have been mailed to. Concerns around this type of fraud often focus on cases in which ballots are mailed to the wrong place, or are left somewhere where anyone might pick them up. Someone intending to commit fraud might be able to wait for random opportunities like these as another means for obtaining ballots, but they are unlikely to know in advance when or where such an opportunity might occur.

Like other states, after receiving a returned mail-in ballot, Washington compares the signature on the ballot envelope to the voter's signature in a government database in order to validate the identity of the voter. To successfully cast a dead person's ballot, a fraudulent actor would therefore need to forge the signature well enough to circumvent this process. While there is no doubt that signature verification is an imperfect filter, it is a real barrier, and ballots are regularly thrown out due to signature issues.[6]

[6]See for example `https://www-cdn.law.stanford.edu/wp-content/uploads/2020/04/SLS_Signature_Verification_Report-5-15-20-FINAL.pdf`.

Finally, the Elections division of the Secretary of State frequently purges newly-non-eligible voters such as felons, the dead, and individuals who have moved outside of the state. In the case of the recent registrants who have died, the Elections division uses Department of Health death records to match on name, date of birth, and the last four digits of one's Social Security number, which ensures a high confidence match that the purged voter is indeed recently deceased.[7] Furthermore, the state participates in the Electronic Registration Information Center, a consortium of 30 states that share voter file information in order to eliminate extraneous voter records by identifying cross-state movers, in-state voter updates and duplicates, and the deceased.

Therefore, after obtaining a dead person's ballot and forging their signature successfully, a fraudulent actor would then need to hope that the state has accidentally missed the death record of the individual associated with the ballot—otherwise, when the ballot is received, it will be flagged as belonging to a deceased individual and will not be counted.[8] In addition, an audit process will begin, and if it is determined that the ballot was cast fraudulently, a criminal investigation could follow. If a person is found guilty of fraudulently casting a ballot in this manner, it is a class C felony in the state of Washington punishable by up to 5 years in prison.

Given these countermeasures, it would seem difficult to carry off this type of fraud at the scale required to alter election outcomes meaningfully. Finding a large enough number of ballots, forging the signatures, and evading the validation countermeasures seem like daunting challenges to a would-be fraudster. Given the large felony penalty if a person is caught, and the dim prospects for changing an election outcome in this way, it is perhaps not surprising that existing research concludes this kind of fraud is rare.

[7]Some older voter records lack the last four SSN digits; for these, the Elections division examines possible matches based off of the decedent's name and date of birth.

[8]In Washington, a ballot cast by an individual who subsequently dies in the period between voting and election day is considered a valid vote.

4 Using Death Records and the Voter File

To assess the rate at which dead people's ballots are counted in Washington elections, we gathered official death records from 1990 onwards from the Department of Health Death Index in the Washington State Digital Archives. For each death record, we have a unique reference number, first name, middle initial, last name, date of death, county of residence, and age at death.

We then use the complete voter file and vote history files from the Washington Secretary of State. The voter files contain records of people who voted from 2011 to 2018, with information including a unique state voter ID, first name, middle name, last name, and date of birth. The vote history files contain the state voter ID, county, and election for each counted ballot.

We focus on data from 2011 through 2018. Although we do have access to data from 2006 up through 2011 as well, we do not use this data for two related reasons. First, the state of Washington did not have statewide universal vote-by-mail until 2011. Second, in communications with the Washington Secretary of State's office, we were made aware of potential issues in the data for the period prior to full adoption of universal vote-by-mail. Consistent with the idea that the high-quality data starts in 2011, we found that the numeric ID in the voter file meant to uniquely link voters to their voter histories—critical for our analysis—are not fully unique until 2011. We are therefore unable to distinguish genuine potential fraud cases from database error in this earlier period. As such, we have removed this period from the analysis, and we report counts of potential fraud and their rate among all voters using only 2011-2018 data.

We use all federal races in our analysis, comprising all statewide primary and general elections during the time period we study.

5 Main Evidence: Minimal Fraud in Washington

We begin by presenting our most credible evidence on the rate of fraud related to deceased voters' ballots in Washington state.

We start by defining a "name match" as any death record that links to a counted ballot in the voter file under the following conditions:

1. Reflects a death that occurred more than 90 days prior to the election;
2. The death record and the voter record match exactly on first name, middle name, and last name;
3. The death record and the voter record match exactly on age (in years);
4. The death record and the voter record match exactly on county of residence;
5. The death record and the voter record match exactly on gender.

We restrict to deaths occurring more than 90 days prior to a given counted vote because, in the state of Washington, a ballot mailed in by a living voter who then dies prior to the election is a valid vote. Because voters can receive their mail-in ballots up to 90 days before the election,[9] any link we find between a deceased voter and a counted vote within 90 days of the election is likely to be legitimate.

After matching on full name, age, county, and gender, as Table 1 shows, we are left with 907 total name matches, out of roughly 4.5 million voters. Most of these are not fraud. Within a large county, a non-trivial number of people share the same name, age, and gender. As such, the vast majority of these possible cases actually reflect two different people, one of whom died, the other of whom cast a perfectly valid ballot. The state of Washington is able to rule out many of these cases because they have access to additional data, like dates of birth and the last four digits of Social Security numbers, that are not present in the public version of the death records.

To overcome this issue, we next collect data on dates of birth for these possible links, using online records. We conducted a manual search using FindAGrave.com, FamilySearch.org,

[9]Confirmed in personal correspondence with the Washington Secretary of State's office.

Table 1 – Finding Potential Cases of Voter Fraud Related to the Casting of Dead People's Mail-in Ballots, Washington State, 2011–2018

	All Voters		Name Matches		Plausible Cases
# Cases	4,550,505	$\longrightarrow$	907	$\longrightarrow$	14 [11, 210]
Rate	–		0.000199		0.000003
Variables Used To Link		Last Name First Name Middle Init County Age Gender		DOB	

Unit of observation is a distinct voter. Manski bounds in square brackets.

and other online sources to look for obituaries that provide a date of birth for the deceased. When we find that a voter with a counted vote in the voter file shares the same date of birth as the one we find through this search process, we count that as a positive match. If we find that the two individuals have different dates of birth, we count that as a confirmed negative match. When we cannot find a date of birth for the death record, we leave this as unconfirmed.

After this process, we find 11 confirmed matches of potential fraud. We rule out 697 of the cases. This leaves us with 199 cases we cannot rule out. To produce a single estimate of the number of potentially fraudulent cases for this group, we use the rate of confirmed matches from the cases we are able to rule on decisively. This rate is 0.016, i.e., 11 confirmed matches divided by the 11 confirmed matches plus the 697 confirmed non-matches. Multiplying this rate by 199 gives us an estimated 3 additional plausible cases for the unconfirmed set, leading us to estimate a total of 14 plausible cases. This constitutes a rate of this form of fraud of roughly 0.0003%.

These estimated 14 cases, including the 11 confirmed matches, are still not necessarily cases of fraud—they may indicate clerical errors, or cases in which two individuals shared the same name and birth date—but they are the most plausible cases that exist in the data.

Obviously, they constitute a tiny fraction of all voters in our sample, far too small to affect any major election outcome.

Next, we perform a Manski bounding exercise using the uncertain cases, by imagining that they are all false positives or false negatives. These bounds are given in square brackets in the table. If we assume that all of the unconfirmed cases are actual matches, we arrive at 210 matches from 2011 through 2018. This is clearly a large overestimate of the total number of matches, but it is still a very small rate of possible fraud, a rate of roughly 0.0005%.

6 Looking for Additional Cases of Fraud

Our baseline estimates reveal extraordinarily low rates of potential fraud related to deceased individuals' ballots in the state of Washington. However, we could be missing additional cases of potential fraud if the record linking procedure we used above is overly conservative. For instance, there could be cases in which someone's full name differs in the two databases due to differences in middle names, such as if one record includes only a middle initial, or if one record has no middle name while the other does. Misspellings or alternative spellings of the first name are another potential source of false negatives.

To see if there are additional cases of fraud we might be missing, we conduct an automated evaluation of a much larger pool of possible cases. Our expanded pool of possible cases includes all instances where the age, county, gender, and first and last name of a voter match a death record but the middle initials in both records do not match or are missing. We also loosen the match on first name to permit differences in spelling by defining a match for the first name as any case in which the Jaro-Winkler string distance between the first name in the two records is below 0.1. By loosening the match conditions in these ways, we significantly increase the likelihood of false positives, but it allows us to assess whether there many additional potential cases we've missed.

We conduct this automated evaluation by scraping FindAGrave.com and FamilySearch.org, the two sources we most often used to confirm or disconfirm a case manually.

Casting this wider net, we find a total of 25 cases where we verify matching birth dates, from among 11,165 possible cases based on our fuzzy name match along with exact matches on county, age, and gender. Because these rely on weaker name-matching conditions, the likelihood of these being cases of two different people with similar names and the same birth date is higher than in our previous analysis. But the fact that we find only 25 potential cases even with this potentially high rate of false positives is informative.

Of the 11,165 name matches under this procedure, we are unable to find date of birth information for 6,418 cases. Using the same technique as before to impute a rate of true matches for this group, we arrive at an estimate of 59 total plausible matches.[10] We suspect many of these may be false positives, but even if these were all fraudulent cases, it is a very small number of voters among 4.5 million individuals we study.

7 Conclusion

The COVID-19 pandemic has strained our election system, and in so doing, it has elevated concerns about the logistics of our elections. With the massive increase in voting by mail, and with a number of states implementing universal vote-by-mail, the claim that fraudulent actors steal dead people's ballots and vote with them has become salient. This particular claim is especially interesting because it is directly testable, because who votes and who dies are both matters of public record in America.

Using these public records, we have found that dead people's ballots are almost never voted fraudulently and subsequently counted as valid votes in the state of Washington. These

[10] Since our automated procedure for validating links leaves many more cases unconfirmed, we evaluate the sensitivity of our estimate of plausible matches to alternative imputation strategies in the appendix. Our estimates of the rate of plausible cases is similar after adjusting for a large number of potential observable differences between confirmed and unconfirmed potential cases.

results are likely to extend to other contexts where states take similar precautions to those taken in Washington.

References

Alvarez, R Michael, Thad E Hall, and Susan D Hyde. 2009. *Election Fraud: Detecting and Deterring Electoral Manipulation*. Brookings Institution Press.

Bryant, Lisa A. 2020. "Seeing Is Believing: An Experiment on Absentee Ballots and Voter Confidence: Part of Special Symposium on Election Sciences." *American Politics Research* 48(6).

Cottrell, David, Michael C Herron, and Sean J Westwood. 2018. "An Exploration of Donald Trump's Allegations of Massive Voter Fraud in the 2016 General Election." *Electoral Studies* 51: 123–142.

Gerber, Alan S., Gregory A. Huber, and Seth A. Hill. 2013. "Identifying the Effect of All-Mail Elections on Turnout: Staggered Reform in the Evergreen State." *Political Science Research and Methods* 1(1): 91–116.

Goel, Sharad, Marc Meredith, Michael Morse, David Rothschild, and Houshmand Shirani-Mehr. 2020. "One Person, One Vote: Estimating the Prevalence of Double Voting in US Presidential Elections." *American Political Science Review* 114(2): 456–469.

Hood III, MV, and William Gillespie. 2012. "They Just Do Not Vote Like They Used To: A Methodology to Empirically Assess Election Fraud." *Social Science Quarterly* 93(1): 76–94.

Levitt, Justin. 2007. "The Truth About Voter Fraud." Working Paper. `https://www.brennancenter.org/sites/default/files/legacy/The%20Truth%20About%20Voter%20Fraud.pdf`.

Lockhart, Mackenzie, Seth J Hill, Jennifer Merolla, Mindy Romero, and Thad Kousser. 2020. "America's electorate is increasingly polarized along partisan lines about voting by mail during the COVID-19 crisis." *Proceedings of the National Academy of Sciences of the United States of America* 117(40): 24640–24642.

Mebane, Walter. 2008. "Election Forensics: The Second-Digit Benford's Law Test and Recent American Presidential Elections." *Election Fraud: Detecting and Deterring Electoral Manipulation* pp. 162–181.

Minnite, Lorraine C. 2010. *The Myth of Voter Fraud*. Ithaca, NY: Cornell University Press.

Thompson, Daniel M., Jennifer A. Wu, Jesse Yoder, and Andrew B. Hall. 2020. "Universal Vote-by-mail Has No Impact on Partisan Turnout or Vote Share." *Proceedings of the National Academy of Sciences* 117(25): 14052–14056.

Online Appendix

Intended for online publication only.

Contents

S.1 Sensitivity of Potential Case Rate Calculation

Throughout the paper, we estimate the number of potential cases of fraud using the rate of positive cases as a share of cases we can confirm either way and multiplying this by the number of potential cases. This calculation assumes that the cases we cannot confirm have a similar number of positives as the cases we can confirm. We cannot directly confirm this assumption. Still, we can rule out that the cases we fail to confirm are clearly different from the cases we can confirm in ways that would make them much more likely to be positive if we could confirm them.

To assess the possibility that the unconfirmed cases are different in important, observable ways, we estimate the rate of positive cases after accounting for observable characteristics of the case that may relate to the likelihood that the case is positive. We use logistic regressions of a flag for positive cases on a set of covariates, and calculate the average predicted probability of a positive for each potential case including the cases we cannot confirm either way. In each regression, an observation is a potential link, meaning that voters can be linked to multiple death records and some are in this analysis.

Table S.1 reports our estimates. In column 1, we report the positivity rate using the simple approach we use in the paper. In our regression framework, this is equivalent to on constant-only regression—assuming that all cases have an equal probability of being positive regardless of their characteristics. In column 2, we report estimates after relaxing this assumption, instead calculating a probability that a case is positive for each match type. We categorize the matches into five categories: exact name match; first and last name match but middle initial is missing in both records; first and last name match but middle initial is missing in one record; first and last name match but middle initials are different; last name matches but first name is slightly different. This adjustment does not meaningfully change our estimate of the positivity rate.

Column 3 accounts for the population of the county in which the person lived and voted. Since we are more likely to find positive—probably false positive—cases in counties with many people, adjusting for the county population could change our expected positivity rate if the unconfirmed and confirmed cases came from different counties. In column four, we adjust for the commonness of the decedent's last name, suspecting that common last names also increase the rate of false positives. We find that both of these adjustments are not consequential.

In column 5, we adjust for the availability of Social Security Death Index (SSDI). People born prior to 1936 or who died after 2014 may not be listed in the SSDI. When we adjust for this, our positivity rate goes up modestly. While we cannot directly translate this estimate

Table S.1 – Sensitivity of Plausible Case Rate Calculation.

	Plausible Cases/Potential Cases				
	(1)	(2)	(3)	(4)	(5)
	0.0048	0.0047	0.0046	0.0047	0.0061
Controls					
Match Type Dummies	No	Yes	Yes	Yes	Yes
Log(Deaths in County)	No	No	Yes	Yes	Yes
Log(Last Name Freq in Death Records)	No	No	No	Yes	Yes
SSDI Records Availability Dummy	No	No	No	No	Yes

Each cell reports an estimate of the share of plausible links that would be potential links. Estimates are average predicted probabilities from logistic regressions. Each regression regresses a dummy variable for a potential case on covariates expected to predict potential cases. Regressions are estimated using cases where the scraper finds definitive evidence of a potential case or rules the case out. The share of plausible links is estimated by using the regression to extrapolate to the cases the automated searching algorithm cannot classify.

into a number of voters, we can approximate how these differences would change our main point estimate by inflating the rate we use for imputation by $\frac{0.0061}{0.0048}$ from columns 5 and 1. This would increase our point estimate from 53 to 68.

In total, Table S.1 tells us that our simple method of estimating the rate of plausible cases produces similar results as other methods that explicitly adjust for differences between the cases we can confirm and those we cannot.