

## Columbia Journalism Investigations

### Projecting rejected absentee ballots in the 2020 general election

This November we can expect historic levels of voting by mail. Experts believe at least half of the electorate will vote absentee, millions for the first time. If the rate of rejected mail ballots from 2016 is about the same this time around, the CJI analysis found that more than one million voters are likely to have their ballots tossed out. This analysis was completed in collaboration with Merlin Heidemanns, a Political Science PhD student at Columbia University.

#### Data

We used the only source for national election administration data—the [Election Administration and Voting Survey \(EAVS\)](#)—from the 2016 general election. We rely on the 2016 data because it is most complete. While 2016 showed low levels of voter turnout among minority voters compared to 2008 and 2012, those years would have been less complete while also being less representative in terms of absentee voting.

The purpose of our analysis is to see—assuming that turnout, absentee ballot submission and rejection rates remain the same—how an increase in absentee requests would impact the number of ballots rejected. We focused our analysis only on civilian mail ballots (Section C of EAVS) using the EAVS Data Public Release Version 4.

#### Data Cleaning

Elections across the U.S. are run on the local-level. The county most commonly runs elections, but in a few states, they are administered on a hyper-local level: by parish, ward or municipality. Wisconsin for example runs elections by ward, and has over 1,800 election offices. For a few states, we chose to aggregate the data up to the county level which provided a more robust sample size and the ability to merge in demographic data. Our final dataset has 3,114 jurisdictions.

#### Missing Data

While EAVS 2016 is the most complete data set we have for a general election year, there is still missing data. 167 jurisdictions of the 3,114 total did not report rejected ballot data (C4b) across eight states. We have almost complete data in battleground states however, with only Arkansas reporting some missing data. Submitted ballots (C1b) and transmitted ballots (C1a) also had missing data, but it made up only 0.8% and 0.7% of all jurisdictions respectively.

The data was supplemented with additional data that was collected from the election-related state and local offices and merged into the EAVS data based on the associated FIPS code. If 85% of the jurisdictions did not report data on rejected ballots in a state, we contacted the jurisdictions and states to fill in the missingness. We based the 85% threshold on the [EPI methodology](#). In the following states, fewer than 85% of jurisdictions did not report data in 2016: Texas, Arkansas, New Mexico, Alabama, and Vermont.

- Alabama reported no rejected ballot data

- Vermont did not report any data at all. The state of Vermont provided a full dataset upon request.
- New Mexico reports that “data may not add up because it comes from different, unidentified sources”—we were unable to get clarification from the state despite multiple attempts to reach them.

After merging in data we received from Vermont, there was still a significant amount of missing data in Texas, Arkansas, New Mexico, and Alabama. We then called county officials to fill in missing data on the local level as much as possible.

If the states or jurisdictions were not able to fill in the missing data, we calculated rejected ballots by subtracting absentee ballots counted (C4a) from absentee ballots returned (C1b). This calculation is only an approximation of the number of rejected ballots. There may be instances where a problem is identified by an election official and a new ballot is issued or the voter decides to vote provisionally in-person on election day. In both cases this could overestimate the number of ballots rejected. For this reason we used this approach minimally, applying to only 1.5% of all observations.

When data was still missing and couldn't be calculated, we first imputed the state-level mean values. The state average was applied to 58 jurisdictions in Texas, 12 jurisdictions in Arkansas, 10 in New Mexico, and one in Hawaii and Maine. If the state had not reported any data (e.g. Alabama did not report the number of rejected absentee ballots), we imputed the national mean.

### **Late Ballots**

As reported in the EAVS 2016 report, absentee ballots that arrive late is the most common reason ballots are tossed. A few states however report all missing numbers for the “ballot not received on time/missed deadline” including Alabama, Connecticut, Hawaii, Illinois, Mississippi, and Rhode Island. We reached out to these states to confirm that they do in fact count a late ballot as rejected, and received confirmation from Alabama, Mississippi and Rhode Island that late ballots are considered rejected. In Illinois some jurisdictions do classify late ballots as rejected ballots while other jurisdictions do not—considering them instead as unreturned.

### **In-Person Early Voting**

In some jurisdictions, in-person early voting is included in mail ballot data (Section C data). One solution was recommended by election experts: if in-person early voting count (F1f) plus the number of absentee ballots counted (F1d) equals or is greater than the number of mail ballots counted (Section C), we could reasonably assume that early voting was included in our absentee ballot data. 74 jurisdictions in 12 states fit this criteria. In these cases, it was recommended to subtract in-person early voting count (F1f) from all absentee ballot data. In a majority of these jurisdictions, early vote center participation was 0 or resulted in negative numbers, thus we decided against applying this workaround.

## Exclusions

Two states presented data quality issues that led to their exclusion from the analysis. Connecticut and Hawaii both indicate that they transmitted fewer ballots to voters than were submitted. In Connecticut, military and overseas (UOCAVA) voters are counted among ballots that were submitted but not among those that were transmitted. Given the data quality problems and the lack of workarounds to correct them, we decided to exclude both of these states.

## Limitations

The analysis assumes that the reported data is accurate and that where it was applied rejected ballots if missing can be computed with the number of submitted and counted ballots. It assumes further that the sample of individuals who requested absentee ballots in 2016 is representative of the county level population as a whole such that submission and rejection rates can be extrapolated to the rest of the population in the county. This assumption is likely problematic given that white, richer, and older voters with lower rejection rates are more likely to vote absentee than minority populations that are poorer and younger. Notably, the latter are likely to vote absentee in larger numbers than before which may increase the average rejection rate making this a conservative estimate.

## Analysis

Our goal was to estimate the expected number of rejected ballots if more people vote by mail in 2020 than in the 2016 general election. Given the increase in mail balloting in the 2020 primaries, the number of mail ballot requests for November already, and the growing spread of the coronavirus, it's likely to be the [case](#).

In early July, we started reaching out to experts in the field such as Charles Stewart, MIT and Tammy Patrick, Democracy Fund. Back then estimates were ranging between 50-60% and even up to 70%. When we started our analysis, it was [uncertain](#) how many people would choose absentee balloting over voting in-person, so we decided to run two scenarios to explore a range of possibilities in the following weeks: we assumed half and 75% of the electorate voting by mail in every jurisdiction. In early October most of the experts we talked to estimated that between 40-50% or up to 60% may vote by-mail this November.

We know that that's not entirely realistic for all jurisdictions: not all jurisdictions will see a large increase in mail ballots, and the levels of mail balloting will vary based on access, historical behavior and partisan demographics. For example, most jurisdictions in Washington and Colorado, where mail-in voting is the default, are likely to cast more than 90% of their ballots by mail. Conversely, we know that some jurisdictions in Texas, where access to mail-in voting is restricted, are unlikely to reach even 50% vote-by-mail. We document the rate of voting by mail by state in Chart 1 below.

## Model

We applied the following model:

*Jurisdiction turnout in 2016 \* our assumption about the share of requested absentee ballots \* the jurisdiction probability in 2016 of an individual sending back an absentee ballot \* the jurisdiction probability of a submitted absentee ballot getting rejected*

For example, assume a county in which 100 people voted in 2016, where the share of returned absentee ballots was 90% and where 10% of returned absentee ballots were rejected. We assume that 50% are going to vote absentee in 2020. Then, the expected number of rejected absentee ballots is:  $100 * 0.5 * 0.9 * 0.1 = 4.5 \sim 5$  mail ballots.

After we calculated the number of ballots rejected for every county across the United States, we merged the 2016 EAVS data in demographic data on race, poverty, unemployment, and household income from the [2018 American Community Survey Five-Year estimates](#).

**Chart 1: Rate of VBM by State in 2016**

State	VBM Participation in 2016
AL	0.00%
AR	2.22%
AZ	73.71%
CA	57.13%
CO	91.25%
CT	7.73%
DC	5.32%
DE	3.08%
FL	27.64%
GA	4.81%
HI	43.25%
IA	40.87%
ID	28.20%
IL	6.53%
IN	33.26%
KS	14.32%

KY	1.85%
LA	2.80%
MA	4.46%
MD	5.63%
ME	32.80%
MI	25.73%
MN	22.37%
MO	8.96%
MS	8.44%
MT	64.33%
NC	3.72%
ND	23.30%
NE	27.13%
NH	9.29%
NJ	8.73%
NM	5.75%
NV	6.40%
NY	4.87%
OH	21.33%
OK	6.75%
OR	99.14%
PA	4.24%
RI	8.23%
SC	23.27%
SD	28.43%
TN	2.09%
TX	4.95%
UT	68.72%
VA	13.40%
VT	28.35%
WA	98.20%

WI	4.77%
WV	1.67%
WY	30.98%
<b>Grand Total</b>	<b>23.99%</b>