Voting Technologies and Ward Vote Data from Wisconsin

November 27, 2016

Did the outcome of voting for president in Wisconsin accurately reflect the intentions of the electors? Concerns have been raised about errors in vote counts produced using electronic technology—were machines hacked?—and a recount may occur.

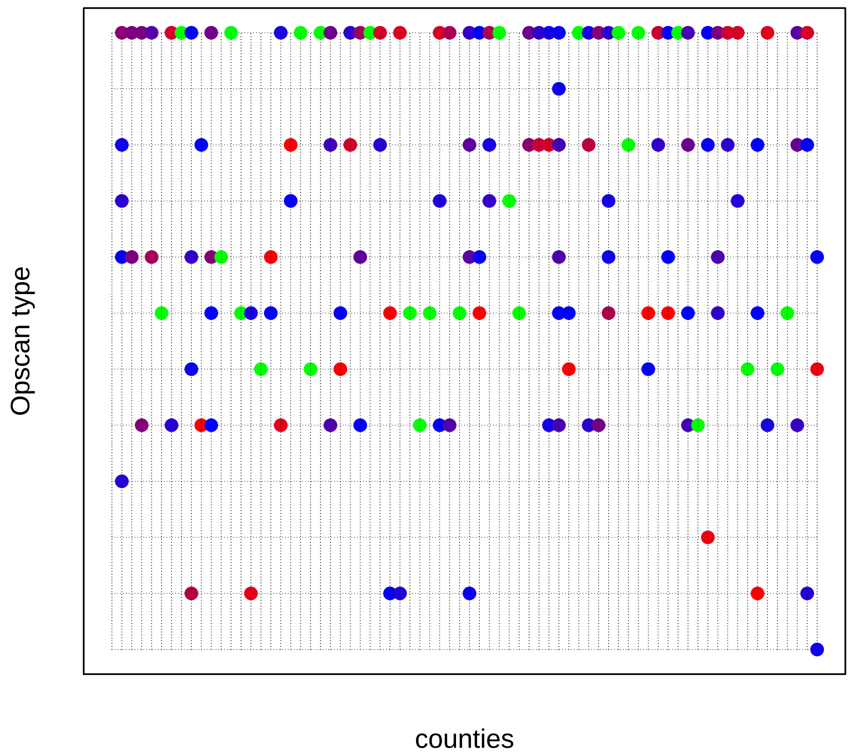
Some statistical analysis has been discussed using data for counties, but such analysis cannot resolve questions about whether voting technology—paper or electronics?—affected the votes for president. Technology, demographics and all other covariates vary within counties, and as argued here there is little reason to believe relationships among all those variables and votes are the same across the whole state or even across a single county.

Variation within counties and the likely diversity of relationships mean the county-level analysis does not give a reliable impression regarding whether voting technology relates to votes. With ward data we can do better, but an audit is needed to get a definitive answer.

That different voters in a county often used different voting technologies is clear from the figure. Many counties used multiple technologies, almost all used electronic vote tabulation technology, and some used both direct-record electronic (DRE) and optical scanner (Opscan) technologies (see the list of equipment used by each municipality). Opscan technologies mark votes on paper but tabulate the votes electronically.

Each horizontal line corresponds to a type of optical scanner technology, and each vertical line corresponds to a county. "None" for the Opscan type (the top row) reflects an unknown mix of DRE technologies and hand-tabulated paper ballots. Down the subsequent rows the other types are: (2) Dominion (Premier)-Accuvote-OS; (3) Dominion (Premier)/Command Central-Accuvote-OS; (4) Dominion (Sequoia)- Sequoia Insight; (5) Dominion (Sequoia)/Command Central- Sequoia Insight; (6) Dominion ImageCast Evolution; (7) ES&S DS200; (8) ES&S M100; (9) Optech- Eagle; (10) Optech/Command Central- Eagle; (11) Optech/Command Central- Eagle, Dominion (Sequoia)/Command Central- Sequoia Insight.

Figure 1: Types of Voting Technology Used by Voters in Wisconsin Counties



Note: Each row represents a different type of Opscan technology (the top row is type "None," which means there is either DRE or hand-counted paper ballot technology). Counties in alphabetical order correspond to columns. Each green dot shows a county where all voters used the same kind of technology. Red-blue (purple) dots appear when different voters in a county used different technologies: the most frequently used technologies are more blue and the least frequently used are more red. The red-blue color proportion in each dot matches each county's proportion of voters using each type of technology.

A green dot appears when all the voters in a county used the same kind of technology. Purple dots appear when the technologies used in a county are diverse: the most frequently used technologies are more blue and the least frequently used are more red. In only 26 of the 72 counties were all votes recorded using the same kind of voting technology.

Even if all votes cast using technology type "None" were cast using a DRE system, most counties are internally heterogeneous with respect to whether votes were cast on paper. A county-level analysis is hard pressed to overcome that measurement error, which compromises the analysis. At best, measurement error reduces the apparent magnitude of effects technology type has on votes or on features of votes.

With ward-level data we can do a bit better to assess how technologies relate to features of votes. Wards are the smallest aggregation unit at which vote counts are reported in Wisconsin.

If we could obtain useful ward-level covariates like the demographic characteristics of each ward and the voting histories of the voters in each ward, we might attempt regression-style analysis using ward observations. Unfortunately ward covariates are scant, and we lack such data.

We can use the Election Forensics Toolkit (a website developed as part of a USAID-funded project) to look at features of the ward data. The features of the ward vote counts vary depending on the number of votes cast in the ward and on the type of voting technology used in the ward, so we consider subsets of the Wisconsin wards separately.

For the table a "Small" ward has less than 100 votes. All the statistics in the table that should cause concern occur for Small wards that use some kind of Opscan technology.

One statistic (`LastC`) is the mean of the last digits of the vote counts. At least for large vote counts, this article argues that each of the ten possible last digits of vote counts should occur equally often, in which case the mean is 4.5. Other patterns may suggest the counts were manipulated.

In the Small, Opscan wards the last digits of vote counts for Trump and for Clinton

Table 1: Distribution and Digit Tests, Wisconsin 2016, Wards

Ward Size	Opscan	Name	`2BL`	`LastC`	`P05s`	`C05s`	`DipT`	Obs
Small	Yes	Trump	3.809	3.458	0.448	0.177	0	96
			(3.008, 4.591)	(2.959, 3.969)	(0.354, 0.552)	(0.094, 0.25)	−−	
Small	Yes	Clinton	4.617	3.765	0.469	0.122	0	98
			(3.785, 5.52)	(3.215, 4.306)	(0.368, 0.571)	(0.061, 0.184)	−−	
Small	None	Trump	4.727	4.614	0.253	0.181	0.99	83
			(4.099, 5.328)	(4.024, 5.18)	(0.157, 0.349)	(0.096, 0.253)	−−	
Small	None	Clinton	4.014	4.049	0.148	0.185	0.987	81
			(3.32, 4.698)	(3.407, 4.691)	(0.074, 0.222)	(0.099, 0.259)	−−	
Big	Yes	Trump	4.257	4.474	0.194	0.195	0.994	2525
			(4.15, 4.369)	(4.357, 4.584)	(0.178, 0.21)	(0.18, 0.21)	−−	
Big	Yes	Clinton	4.191	4.44	0.2	0.204	0.862	2525
			(4.079, 4.305)	(4.334, 4.562)	(0.185, 0.217)	(0.189, 0.219)	−−	
Big	None	Trump	4.283	4.49	0.196	0.2	0.996	769
			(4.088, 4.485)	(4.3, 4.705)	(0.166, 0.224)	(0.169, 0.229)	−−	
Big	None	Clinton	3.973	4.493	0.209	0.208	0.992	769
			(3.766, 4.183)	(4.289, 4.706)	(0.179, 0.235)	(0.179, 0.237)	−−	

Note: ward vote counts of zero are omitted before computing statistics. "Obs," number of ward observations. Values in parentheses are 95% nonparametric bootstrap confidence intervals. Point estimates in red differ significantly from the values expected if there are no anomalies.

have means (`LastC`) that are much less than 4.5. Each "confidence interval" for a statistic give a range of estimates we could have observed given variations in the data that might have occurred by chance. The two `LastC` intervals do not include 4.5, which is why the point estimates are shown in red.

As this article points out, last-digit diagnostics have not been claimed to work when vote counts are small. So one view is that we have no reason to expect any particular result for those statistics, so there is nothing to worry about.

Even so, among Small wards the wards that use Opscan technologies exhibit anomalies while those using different technologies do not.

Another statistic (`C05s`) is the mean of a variable indicating whether the last digit of the vote count is zero or five. Based on the same rationale about digit frequencies as for `LastC`, `C05s` should be 0.2 if there are no problems. `C05s` being too large may mean that someone was sloppy and simply wrote down approximate numbers. `C05s` too small might mean that someone is faking the numbers (it has been found that 2 and 7 are favorite numbers for people trying to produce random numbers out of their heads).

In the Small, Opscan wards `C05s` for Clinton is too small, showing that vote counts for Clinton too rarely have a last digit of zero or five. Notably this statistic is significantly too large if ward vote counts of zero are included.

The `P05s` statistic, which is the mean of a variable indicating whether the last digit of the rounded percentage of a candidate's votes is zero or five, has a specific motivation from the idea that people who commit frauds want to allow their efforts to be detected in order to claim credit. Such "signaling" frequently occurs in Russian elections. Like `C05s`, `P05s` should be 0.2 if no signaling is occurring, but larger values of `P05s` are concerning.

Votes in the Small, Opscan wards exhibit a "signaling" pattern (`P05s`).

Having vote percentages concentrated around more than one distinct value, which would mean the distribution of percentages is multimodal, is also a potential problem. For instance, there might be a set of wards where a candidate received 30 percent of the votes and another cluster where the candidate received 60 percent.

In an elaborate model for election frauds, multimodality is an important indicator that one candidate is gaining fraudulent votes. We'd have to know how many voters registered in each ward to be able to estimate that model.

`DipT` is the p-value from a test that there is no multimodality that we can do without having the data needed for the fancier model.

Vote percentages in the Small, Opscan wards are significantly multimodal.

In contrast to the array of anomalies in the Small wards with Opscan technology, none of the statistics in Small wards without Opscan technology have values to worry about.

None of the statistics in "Big" wards have values to worry about, although additional analysis shows the Big wards set is diverse: some Opscan machines, particularly the Dominion (Sequoia)/Command Central- Sequoia Insight (in 209 Big wards) and the Dominion ImageCast Evolution (in 272 Big wards), exhibit anomalies.

The mean of the second digits in the vote counts (2BL) is a statistic some have argued can be used to detect frauds, but actually the statistic responds to many features of normal politics and is ambiguous. The second-digit mean (2BL) for Clinton in Big wards without Opscan systems deviates from the value Benford's Law would imply, but the observed value is not unusual (see Chapter 9 in this book).

Why do Small wards with Opscan technology (and several other kinds of wards) have anomalies, and do the anomalies mean the reported vote counts do not accurately reflect the intentions of the electors? Given all the information we have, it is hard to say.

A rigorous post-election audit, like some are trying to have happen in several states, is not subject to the limitations that prevent a full regression-style analysis nor to the interpretive uncertainty involved in using statistics like those from the Toolkit.

A crucial feature of an audit is that paper ballots are inspected directly by humans and not merely tabulated again by a machine, which can happen in a recount under some state recount procedures. An audit can tell us at least whether the votes marked on paper have been correctly tabulated by the machines.

A rigorous audit or a full recount that has humans manually checking the paper ballots can provide convincing evidence about who won the election. In the current environment, the reassurance such an audit may provide would contribute to the incoming government's legitimacy.

Walter R. Mebane, Jr. is a research associate at the Center for Political Studies, professor of political science and professor of statistics at the University of Michigan.