Learning from Recounts

ELECTION LAW JOURNAL
Volume 17, Number 2, 2018

DOI: 10.1089/elj.2017.0440

Learning from Recounts

Stephen Ansolabehere, Barry C. Burden, Kenneth R. Mayer, and Charles Stewart III

ABSTRACT

We compare the results of two recent statewide recounts in Wisconsin—the 2011 Supreme Court election and the 2016 presidential election. Using the measure of absolute differences between the original tally and the recount, we find an error rate at the reporting unit level of 0.21% in 2011 and 0.59% in 2016. The 2016 error rate drops to 0.17% when write-in votes are removed from the analysis. We also find that paper ballots originally counted with optical scanners were counted more accurately than ballots originally counted by hand. To reach these conclusions, we address the methodology of measuring differences in election night and recounted vote tallies. The most commonly used measure to compare election-night and recounted tallies, the net difference, significantly understates the magnitude of errors in the original tally. We also develop a regression-based technique that estimates what the error rate should be if a ballot-by-ballot recount were possible. We conclude by discussing the implications for requiring post-election audits.

Keywords: voting machine, recount, electronic voting, optical scan, election recounts, Wisconsin

THE 2016 PRESIDENTIAL ELECTION was notable for many reasons, not the least of which was the fact that the integrity of vote counting was called into question. The overall accuracy of the count was challenged by both the winning candidate, Republican Donald Trump, and Green Party candidate Jill Stein. Questions raised by Stein in particular also generated a brief flurry of interest in the issue of post-election audits, a piece of election arcana that rarely sees the light of day in mainstream public discourse.

Stein's challenge to the accuracy of the vote count contained a direct attack on the use of computers to tally votes. Without a hand recount of the vote, she claimed, there would be no way of knowing whether computerized voting equipment had in fact accurately counted the vote, or worse, had been maliciously hacked.

Stein tried to force recounts in three battleground states—Michigan, Pennsylvania, and Wisconsin—that had given narrow and surprising majorities to Trump over the Democratic nominee, Hillary Clinton. In the end, she was successful only in Wisconsin. That recount produced revised vote totals for the two major party candidates that were virtually identical to the originally canvassed tally. Most likely because of this, the Wisconsin recount quickly vanished from the public mind, to become yet another footnote to an extraordinary election season.

However, there is more to be learned from a recount than whether the correct winner was declared or the vote margin was correct. A focus on changing vote margins masks richer information that emerges

Stephen Ansolabehere is the Frank G. Thomson Professor of Government at Harvard University in Cambridge, Massachusetts. Barry C. Burden is a professor in the Department of Political Science and Director of the Elections Research Center at the University of Wisconsin-Madison in Madison, Wisconsin. Kenneth R. Mayer is a professor in the Department of Political Science at the University of Wisconsin-Madison in Madison, Wisconsin. Charles Stewart III is the Kenan Sahin Distinguished Professor of Political Science at the Massachusetts Institute of Technology in Cambridge, Massachusetts. The authors are grateful for comments received from Martha Kropf, Craig Arceneaux, and the anonymous reviewers of this Journal, which significantly improved the analysis. The staff of the Wisconsin Elections Commission was also very helpful in providing the authors with data and answering questions about election practices in Wisconsin. Any remaining errors are entirely the fault of the authors.

during a recount that can help quantify the degree to which errors were made in the original count, including errors that had no bearing on the outcome.

In this article, we take advantage of the fact that Wisconsin has had two statewide recounts in recent years—after the 2016 presidential election and the 2011 state supreme court election—and that the state has kept detailed records about how the vote count changed from canvass to recount. In addition to detailed vote-count statistics, the state also employed a mix of voting technologies, which allows for a comparison of initial-count error rates, broken down by whether the votes were originally counted by hand, by scanner, or electronic voting machine.

We reach the following conclusions in this article.

1. The most common way of comparing the recounted victory margin with the election night victory margin significantly understates the number of errors made in the original count of ballots.
2. At least 0.59% of the ballots originally counted in the 2016 presidential election in Wisconsin were miscounted, compared to 0.21% in the state supreme court election that was recounted in 2011. The difference in these two error rates is due almost entirely to the miscounting of minor party and write-in ballots in 2016.
3. Scanning paper ballots produces a more accurate election night count than hand-counting ballots.
4. Differences in vote counts between election night and the recount are largely due to administrative factors, such as transcription errors, rather than the accuracy of the vote-tallying methods per se.

These conclusions are supported by the evidence generated by the 2011 and 2016 Wisconsin recounts. They are consistent with the prior empirical literature on recounts, and thus should provide guidance as academics and the media empirically analyze recounts in the future.

The remainder of this article proceeds as follows. First, we discuss the general issue of recounts as a method for measuring the accuracy of original vote counts. Second, we review the specific case of Wisconsin and the two recent recounted elections that provide the empirical fodder for this article. Third, we delve more deeply into these two recounts, exploring the relationship between voting technology and vote-count accuracy.

RECOUNTS AS A METHOD FOR MEASURING ACCURACY OF THE VOTE COUNT

Recounts provide an opportunity to gauge the accuracy of election night vote counting,[1] although few scholars have availed themselves of this opportunity.[2] The overall logic is simple. If there is a recount, the recounted results are taken to be the correct vote count, or the "ground truth" of the election. The difference between this more careful count and the tally conducted in the hectic hours immediately following the close of the polls is a measure of how accurately the ballots were counted in the first instance.

Although there might not be such as a thing as a perfectly correct vote count, recounts are designed to improve on the process used to count ballots on Election Day, and thus it is reasonable to consider the recounted tally as ground truth. The recount is done more slowly and deliberately; is conducted under the watchful eyes of election officials, journalists, and election observers from the campaigns; and is focused on getting a single contest right rather than processing the entire ballot. In this setting, previously undetected errors and miscounts often get revealed and remedied. For example, the 2016 recount in Wisconsin identified a few communities where tabulating machines failed to read some ballots because voters used the wrong kinds of pens to mark their scanned paper ballots.[3]

Taking the recounted vote as ground truth, then, the election night deviation from the recounted vote can be assumed to measure counting error. This error, however, can be measured in different ways and at different levels of aggregation.

---

[1]We use the term "election night vote count" as a synonym for the vote count that is produced for the original canvass of votes.
[2]The one scholarly paper we know of that has used recounts in such as way is Ansolabehere and Reeves (2012), which provides the methodological basis for this article. Also see Herron and Wand (2007), Atkeson, Alvarez, and Hall (2009), and Alvarez, Atkeson, and Hall (2013).
[3]The community where this affected the most votes was the City of Marinette, in Marinette County, which discovered that hundreds of absentee ballots had been marked in ways that "would cause problems during scan—i.e., red ink, ball point pen, incomplete connecting arrows, create in ballot, etc." See Marinette County Board of Canvass, Marinette County Unapproved Recount Minutes, pp. 38–53, available at <http://elections.wi.gov/sites/default/files/recount_2016/marinette_county_unapproved_recount_minutes_pdf_85823.pdf>.

Table 1a. Calculation of Vote-Count Error Using Hypothetical Vote Totals: Precinct Aggregation

		Election night				*Recount*				*Net difference*				*Absolute difference*			
Muni.	*Precinct*	*Brown*	*Garcia*	*Lee*	*Row total*	*Brown*	*Garcia*	*Lee*	*Row total*	*Brown*	*Garcia*	*Lee*	*Row total*	*Brown*	*Garcia*	*Lee*	*Row total*
A	1	377	207	4	588	379	202	5	586	2	−5	1	−2	2	5	1	8
A	2	300	169	4	473	303	166	3	472	3	−3	−1	−1	3	3	1	7
A	3	85	42	2	129	83	42	2	127	−2	0	0	−2	2	0	0	2
Municipality column total		762	418	10	1,190	765	410	10	1,185	3	−8	0	−5	7	8	2	17
B	1	478	303	10	791	481	300	11	792	3	−3	1	1	3	3	1	7
B	2	334	169	1	504	331	172	1	504	−3	3	0	0	3	3	0	6
B	3	312	224	3	539	310	223	5	538	−2	−1	2	−1	2	1	2	5
Municipality column total		1,124	696	14	1,834	1,122	695	17	1,834	−2	−1	3	0	8	7	3	18
State column total		1,886	1,114	24	3,024	1,887	1,105	27	3,019	1	−9	3	−5	15	15	5	35

Table 1b. Calculation of Vote-Count Error Using Hypothetical Vote Totals: Municipality Aggregation

	Election night				*Recount*				*Net difference*				*Absolute difference*			
Muni.	*Brown*	*Garcia*	*Lee*	*Row total*	*Brown*	*Garcia*	*Lee*	*Row total*	*Brown*	*Garcia*	*Lee*	*Row total*	*Brown*	*Garcia*	*Lee*	*Row total*
A	762	418	10	1,190	765	410	10	1,185	3	−8	0	−5	3	8	0	11
B	1,124	696	14	1,834	1,122	695	17	1,834	−2	−1	3	0	2	1	3	6
State column total	1,886	1,114	24	3,024	1,887	1,105	27	3,019	1	−9	3	−5	5	9	3	17

The most accurate assessment of counting errors would compare how each individual ballot was interpreted both on election night and in the recount. Indeed, the most rigorous post-election auditing techniques, such as risk-limiting audits, require scrutiny of individual ballots, even if they are selected based on sampling techniques (Lindeman and Stark 2012). The best measure of absolute vote-counting error would simply be the percentage of ballots that were interpreted differently on election night and in the recount.

Practically speaking, knowledge of how individual ballots were interpreted during the election night count is rarely preserved, and thus the most accurate measure of vote-count accuracy is generally unavailable to researchers and the public.[4] As a consequence, the comparison of the election night and the recount tallies must take place at some higher level of aggregation, usually the precinct (called the "ward" or "reporting unit" in the case of Wisconsin). As we discuss below, it may further be necessary to aggregate above the precinct level, such as at the local or state level, as the method used to calculate error can interact with the level of aggregate in ways that might be unanticipated.

There are two ways to measure the amount of vote-counting error revealed by the recount: (1) *net* error and (2) *absolute* error. The two methods are illustrated in Table 1 by means of hypothetical precinct-level vote totals for three candidates named Brown, Garcia, and Lee.

Table 1a displays hypothetical returns from a state with two municipalities, each of which has three precincts. The election night returns are reported first, and then the recounted returns, for the three candidates on the ballot. The final two sets of columns report, first, the net difference in returns for each candidate, and then the absolute value of those differences. For instance, in Municipality A, Precinct 1 (Precinct A-1), Garcia received five fewer votes in the recount than she received on election night, which results in a net difference of −5 votes, but an absolute difference of +5.

Because net errors may be positive or negative, the values will often cancel out. As a result, the

[4]This is changing, as vendors develop digital scanning technologies that preserve together both the image of each ballot and a record of how each ballot was interpreted—both on election night and in a recount, if it occurs.

magnitude of the sum of the net differences is significantly less than the summed absolute differences. This can happen both within precincts—note Precinct B-2, where three votes switched from Brown to Garcia, leaving no net change in total votes—and between precincts—note that Lee's gain of one vote in Precinct A-1 is offset by her loss of one vote in Precinct A-2. The net and absolute error calculations generally diverge as one includes more jurisdictions. With the results reported at the precinct level in Table 1a, the sum of all net differences is –5 and the sum of all absolute differences is 35. The error rates thus appear to be smaller using the net calculation (5 out of 3,019, or 0.17%) than the absolute calculation (35 out of 3,019, or 1.2%)

Table 1b shows what happens when we aggregate the vote totals up to the municipality level. Here, the statewide candidate vote totals are the same as before, as is the net difference. However, the sum of absolute differences is now significantly lower than before, 17 rather than 35 (a rate of 0.56%). Because of the associative property of addition, aggregating the net differences to the municipality level does not affect the final calculation of total net differences. The same associative property does not apply to summing absolute values.

This example illustrates two important properties of error rate measures using differences in election returns that we will use for the rest of this article. First, the absolute difference measure retains much more information about the actual amount of error in the system than the net difference measure.[5] Second, the absolute error will generally fall as vote totals are aggregated at increasingly higher levels.[6] Lower levels of aggregation are better at revealing errors that would otherwise be hidden.

These two properties have important implications for our analysis. We focus on the absolute error because it preserves more information. We will also report absolute errors at the smallest unit of aggregation possible, given the availability of data. When we report statewide totals, we will do so by summing across all individual reporting units, which are generally wards or aggregations of wards. We will only aggregate further, out of necessity. For instance, in comparing recount errors at the local level between 2011 and 2016, it will be necessary first to aggregate to the municipality level because ward boundaries and reporting units changed between these two years.

RECOUNTS IN WISCONSIN

Wisconsin provides a valuable context for studying the accuracy of elections. In recent years, there have been two statewide recounts that allow us to compare the original vote totals for each candidate with vote totals as corrected by the recount, numbers that we assume to be more accurate measures of the real intent of voters. The two elections took place under different circumstances with different numbers of candidates and much different levels of voter participation. This variety provides useful leverage and generalizability for our analysis.

The first recount happened after a nonpartisan state supreme court election on April 5, 2011, between sitting Justice David Prosser and challenger JoAnne Kloppenburg. Initial results following the statewide canvass showed Prosser as the winner 752,323 to 745,007, a margin of 7,316 votes or 0.49% of the total votes cast and counted.[7] Because the margin fell under the 0.5% threshold set by law, Kloppenburg was able to request a statewide recount without having to pay a fee. Her petition was motivated in part by concern about errors in Waukesha County, where her statewide election night lead vanished after the county clerk discovered that thousands of votes were unrecorded in the initial tally.[8] An agreement reached between

[5]In theory, it is possible to have a recount that reveals every ballot having been incorrectly counted (absolute error rate of 100%) and yet for the recounted vote totals to match the election night vote totals. This would be true, for instance, if the candidate names had been erroneously matched up with the locations of marks on a ballot, but that the total number of votes cast in the election night count equaled the total number of votes in the recount. With all the votes reallocated to a different candidate, the errors would balance out equally, even though all the ballots were counted incorrectly.

[6]To be more precise, the amount of absolute error must remain the same or decrease with each additional level of aggregation. If all the net errors at the lowest level of aggregation—the individual ballot, in this case—are non-negative, then aggregation will have no effect on the calculation of the total amount of absolute error. The divergence of net and absolute error at greater levels of aggregation depends on the mix of positive and negative errors at the lower levels.

[7]There were 1,550 "scattered" votes for other candidates.

[8]Returns from an entire municipality (the City of Brookfield) were uncounted because of a data entry error. Jason Stein, Laurel Walker, and Bill Glauber, "Corrected Brookfield Tally Puts Prosser Ahead After 7,500 Vote Gain," *Milwaukee Journal-Sentinel*, April 7, 2011.

the state and the candidates mandated hand recounts in parts of 31 counties that used Optech Eagle scanners.[9] After several weeks of recounting by county boards of canvassers, the state reported 752,694 for Prosser and 745,690 for Kloppenburg, a difference of 7,004 votes or 0.46%. The difference in the margin between the two candidates changed by only 0.03 percentage points between the election night count and the recount.

The second recount followed the November 2016 presidential election. The initial canvassed results produced a win for Donald Trump, awarding him 1,404,440 votes compared to 1,381,823 for Hillary Clinton, a difference of 22,617 or 0.76% of votes cast and counted. There were also significant votes for minor party candidates Gary Johnson, Jill Stein, and several write-in candidates. Stein requested a recount, motivated in part by Trump's assertions throughout the campaign about vote "rigging" of the elections, and concerns about voting machines and the effects of a new voter identification requirement. Because the margin between the top two candidates was above the threshold for a state-financed recount, the Stein campaign raised the roughly $2 million needed.[10]

The recount took approximately two weeks, with most counties recounting ballots by hand, a small number re-tabulating optical scan machines, and others using a mix of the two methods.[11] Following the recount, the state certified the results as 1,405,284 votes for Trump and 1,382,536 for Clinton, for a difference of 22,748 or 0.76% of the vote total. In other words, the recount revealed no change in the winner's vote margin, measured to two decimal places.

Our review of press accounts and minutes of the county election boards makes it clear that the recount was more complicated than a simple rerunning of the tabulation process. First, the state-mandated recount procedures required counties to review other administrative practices in the wards and reporting units, and to balance the number of ballots in hand with turnout data from poll books. Discrepancies discovered here could lead to a change in the vote count even before the votes were re-tallied. For instance, if it was decided that disputed absentee ballots had been improperly included in the vote count, then a number of ballots equal to the number of improperly included ballots must be randomly removed from the counting. This is called a "drawdown" of absentee ballots. A similar procedure is followed if the number of ballots found in the ballot box exceeds the number of voters accounted for on the poll list.[12]

From the perspective of judging the accuracy of the original count, one of the most consequential set of decisions that local officials reviewed during the recount was how carefully write-in ballots had been counted and recorded on election night. It is easy to imagine how the interaction of Wisconsin law and the functioning of voting technologies, especially optical scanners, would make it likely that many write-in votes are not counted on election night. Wisconsin election law states that there is "no requirement for a voter to make an X or other mark, fill in an oval, or connect an arrow in order to cast a write-in vote.[13] What this means practically is that poll workers must visually inspect each ballot to ensure that all write-in votes have been accounted for. In the case of scanned ballots specifically, inspectors cannot rely on ballots with write-in votes being diverted to the auxiliary ballot box because the voter had filled in the oval next to the write-in line; even ballots that had not been diverted must be examined to see if they contain a write-in vote without the corresponding oval or arrow being marked.[14]

Even aside from the issue of accurately accounting for all write-in votes on election night, municipal and county officials make clerical errors on election night that are then corrected in the recount.

---

[9]Wisconsin Elections Commission, "2011 Supreme Court Statewide Recount Information," <http://elections.wi.gov/node/1719>.

[10]A state law adopted in 2015 lowered the threshold for a "free" recount to 0.25% of votes cast and counted. The 2016 margin did not fall below that threshold (or even the previous threshold of 0.5%), so Stein was required to reimburse state and local election officials for expenses related to the recount.

[11]Wisconsin Board of Elections, "Presidential Recount County Cost Estimate and Recount Method," <http://elections.wi.gov/sites/default/files/story/presidential_recount_county_cost_estimate_and_reco_16238.pdf>.

[12]Wisconsin Elections Commission, *Election Day Manual for Wisconsin Election Officials*, July 2016, p. 101, <http://elections.wi.gov/sites/default/files/publication/65/election_day_manual_july_2016_pdf_12281.pdf>.

[13]Ibid., p. 107.

[14]Write-in votes cast on DREs also require close attention in order to be reported accurately. The DRE used in Wisconsin is the AVC Edge. A voter wishing to vote a write-in candidate touches a "write-in" button, which brings up a keyboard for the voter to indicate his or her choice. At the end of the voting day, the results tape indicates the number of write-in votes for each office. A separate write-in report lists all the write-in candidates for each race.

Unfortunately, the only systematic information we have to compare the election night and recounted tallies comes from the election returns published by the state election commission.[15] When there are differences between the two tallies, we cannot reliably distinguish between discrepancies caused by machine errors and clerical errors.[16]

Our analysis of the 2011 and 2016 recounts begins with data provided by the Wisconsin Elections Commission that reported the original vote totals for each candidate and the final vote totals after recounts.[17] The data are provided at the level of a "reporting unit." In many municipalities, the reporting unit is the same as a ward (or what would be called a precinct in most states), but municipalities can combine multiple wards into a single reporting unit.[18] The state had 3,636 reporting units in 2016 with an average of 818 votes cast in each. This is the lowest level of aggregation available to us. Aggregation to the municipality level is necessary to merge the 2011 and 2016 data, because of the many changes to reporting units between the two elections. Even the number of municipalities changed slightly over time; there were 1,879 in 2011 and 1,886 in 2016.[19]

Table 2 presents *absolute* error rates from both elections computed at four increasingly large levels of analysis: reporting unit, municipality, county, and state. It also reports *net* error rates, which are unchanged by the level of aggregation. The table confirms our previous observations that absolute error rates are generally greater in magnitude than the value of corresponding net error rates, and that absolute error rates generally decline as the level of aggregation increases. As we aggregate upward from reporting units to municipalities, to counties, and then to the state, the estimated absolute error rates in both elections decrease.

At the reporting unit, the lowest level of aggregation available, the absolute error was 0.21% in 2011 and 0.59% in 2016. That is, a conservative estimate is that about one out of every 475 ballots in 2011 and one out of every 170 ballots in 2016 was miscounted in the election night tabulation. Because the statistics were not calculated at the individual ballot level, this number represents a lower bound on the true error rate.

Although individual ballot data are unavailable, we can use regression to extrapolate from these four levels of aggregation (state, county, municipality, and reporting unit) to estimate the true individual ballot-level absolute error rates. The technique we use is simply to regress the absolute error rates for each level of aggregation on the logged average number of voters at each level of aggregation. The dependent variables are taken from Table 2, and are simply the absolute error rates for the four levels of analysis. The independent variables are the corresponding average number of

TABLE 2. ERRORS IN TWO STATEWIDE RECOUNTS IN WISCONSIN

Level of aggregation	*2011 Supreme Court*	*2016 president*
	Absolute error	
Reporting Unit	3,181 (0.21%)	17,681 (0.59%)
Municipality	2,309 (0.15%)	15,343 (0.52%)
County	1,617 (0.11%)	12,871 (0.43%)
State	1,223 (0.082%)	6,901 (0.23%)
	Net error	
	1,233 (0.082%)	397 (0.013%)

[15]We explored using the minutes of county election boards as a data source for distinguishing specific reasons why the recounted tally did not match the election night tally. Although the information contained in these minutes is invaluable for developing a general understanding of the practical details of the recount process, it is not systematic enough to be used for the purposes discussed here.

[16]So far as we know, the only state that allows the public to distinguish different reasons why the election night tally might differ from the recount, or even the official canvass, is Virginia, which posts a change log on its website that documents the source of every deviation between the election night tally and the official election returns. See Virginia Department of Elections, "Changes to Unofficial Results Activity," <https://www.elections.virginia.gov/resultsreports/dataproject/ChangesUnofficialResults.html>.

[17]The 2011 data comparing the original and recount was downloaded from <http://elections.wi.gov/sites/default/files/COUNTY_BY_COUNTY_FOR_SPRING_2011_ELECTION_AND_RECOUNT.xls>. The 2016 data comparing the original and recount was downloaded from <http://elections.wi.gov/sites/default/files/Ward%20by%20Ward%20Original%20and%20Recount%20President%20of%20the%20United%20States.xlsx>.

[18]Wisconsin municipalities with populations under 35,000 can combine individual wards into aggregate reporting units. See Wisconsin Statutes § 5.15(6)(b).

[19]We lose only seven observations in the merging process, all from municipalities that appeared in the 2016 vote data but were not in the 2011 data.

votes at each level of aggregation, transformed to natural logarithms.[20]

The results of the regressions are reported in Table 3. The associated graphs are shown in Figure 1. The correlations between the aggregate error rates and the logarithm of the average number of voters represented at each succeeding level of aggregation is very high. The R^2 for the 2011 regression is .81, and .99 for 2016. To estimate the true individual ballot-level error rate for each year, we extrapolate the regression line to the point where the number of voters equals 1. Because the logarithm of 1 is zero, the estimated value we are looking for is the intercept of the regression. In 2011, the estimated individual ballot-level error rate is 0.262%, compared to 0.857% in 2016.[21] These rough extrapolations translate to errors in one out of every 382 ballots in 2011 and one of every 117 ballots in 2016. As both Figure 1a and 1b illustrate, the 95% confidence intervals of the prediction are quite large. Furthermore, because this is an out-of-sample prediction, the true prediction error is likely greater than shown and calculated here.[22]

The higher absolute error rate reported in Table 2 for 2016 compared to 2011 might suggest that votes were less accurately counted than five years before, an alarming conclusion if we believe the elections systems have improved over time. However, the structures of the two elections were so different that we are reluctant to draw this conclusion.[23] Most notably, the 2016 election featured many more candidates—those officially listed on the ballot, candidates qualified as registered write-ins, and the scattering vote[24]—which provided more opportunities for error to be introduced into the counting than in 2011.

Further examination of the recount patterns makes it clear that the absolute error rate in 2016 was driven largely by write-in candidates. This is evident in Table 4, which shows the absolute error (calculated at the reporting unit level before summing) for the seven candidates on the ballot, plus quantities associated with the nine registered write-in candidates and the scattering vote.[25]

This breakdown shows that the absolute error rates associated with Trump and Clinton were 0.159% and 0.161%, respectively, both of which are slightly lower than the absolute error rates in 2011. The absolute error rates for the minor party candidates on the ballot were higher, ranging from 0.273% for Johnson to 3.567% for De La Fuente.[26] Finally, the absolute error rates for the write-in candidates were in a league of their own: nearly 27% for the registered write-in candidates and over 37% for the scattering vote.[27]

Table 3. Regression Results Used to Predict Absolute Error Rate of Individual Ballot Recounts

	2011	*2016*
Natural log of the average number of voters	–0.0134	–0.0417
	(0.0046)	(0.0036)
Intercept	0.262	0.857
	(0.045)	(0.037)
N	4	4
R^2	.81	.99
Adj. R^2	.72	.98

[20]These average values in 2011 were 416 for the reporting unit, 930 for the municipality, 20,818 for the county, and 1,498,880 for the state. The corresponding numbers for 2016 were 818, 1,840, 41,330, and 2,975,753.

[21]The 95% confidence intervals of these predictions are 0.216% and 0.172% for 2011 and 2016, respectively.

[22]The extrapolation technique employed here is valid only if the relationship between the absolute error rate and the average number of voters continues to be linear beyond the bounds of the values of the independent variables. See Tufte (1974, pp. 32–33). Whether this assumption actually holds in practice awaits the availability of data from ballot-level post-election audits, which should become more common with spread of risk-limiting audits and the adoption of digital ballot scanners that retain information about how each ballot was interpreted by the scanner.

[23]In addition, the comparison of the two years in Table 2 demonstrates that the net error rate can even go down between two elections while the absolute error rate goes up.

[24]The scattering vote consists of write-in votes cast for candidates, but not reported on an individual candidate basis. In Wisconsin, the only write-in votes that are reported on an individual candidate basis are those for registered write-in candidates. See Wisconsin Elections Commission, "Reporting 'Scattering' Votes," <http://elections.wi.gov/node/3283>. Also see Wisconsin Statutes §§ 7.50(2)(d) and 7.50(2)(em).

[25]Nine write-in candidates were eligible to receive votes and have their tallies individually reported. Among the nine write-in candidates who were eligible to receive votes, conservative Evan McMullin drew the largest number. The 2016 election saw an exceptionally high number of write-in votes in Wisconsin. See Matt DeFour, "More Write-Ins This Year in Wisconsin Than All Previous Presidential Elections Combined," *Wisconsin State Journal*, December 2, 2016.

[26]The other five tickets were led by Darrell Castle (Constitution), Gary Johnson (Libertarian), Jill Stein (Wisconsin Green), Monica Moorehead (Workers World), and Rocky Roque De La Fuente (American Delta).

[27]These estimates count all the qualified write-in candidates and all the scattering vote as two candidates. Thus, the absolute error rate for these two sets of candidates is likely an underestimate, owing to the fact that we have calculated these rate after aggregating across numerous candidates.

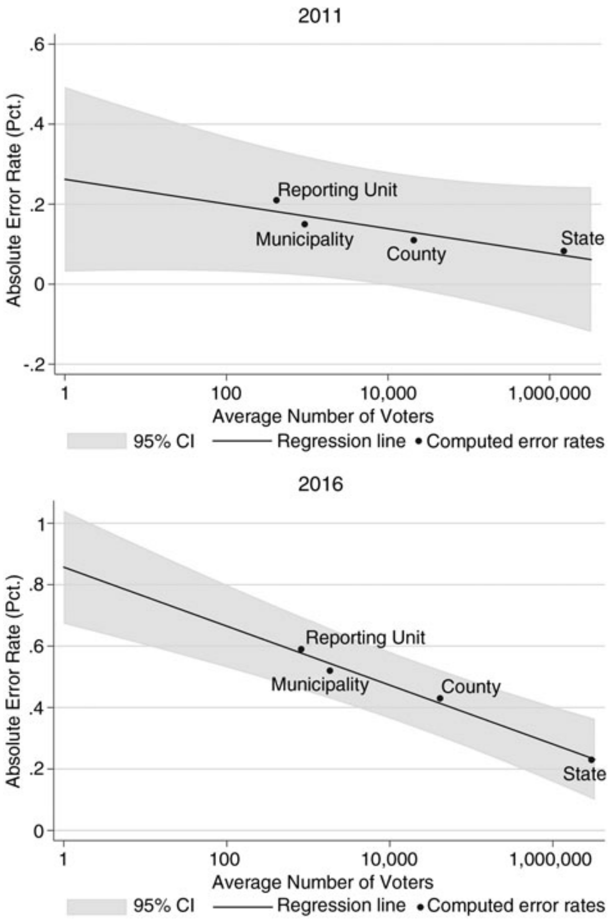

FIG. 1. Relations of absolute error rate to the number of voters at different levels of aggregation in Wisconsin.

This large contribution of write-ins to the absolute error rate led to our further investigation of the recount statistics, which revealed that many counties simply did not recount write-in ballots in 2016, either on election night, in the recount, or both.

This is illustrated in Figure 2, which displays four scatterplots that report the county-level percentage of the vote attributed to four types of candidates on the Wisconsin presidential ballot: major party candidates (Trump and Clinton), minor party candidates (five other candidates listed on the ballot), registered write-in candidates (nine candidates certified by the state to receive write-ins), and the scattering votes (all other write-in candidates). Circles in the scatterplots are sized proportional to the number of votes counted in each county on election night.

TABLE 4. ABSOLUTE ERROR RATES BY CANDIDATE IN 2016

Candidate	*Original votes*	*Absolute difference*	*Absolute rate*
Trump	1,404,440	2,236	0.159%
Clinton	1,381,823	2,227	0.161%
Johnson	106,585	291	0.273%
Stein	31,006	160	0.516%
Castle	12,156	88	0.724%
Moorehead	1,769	15	0.848%
De La Fuente	1,514	54	3.567%
Registered write-in	10,458	2,818	26.946%
Scattering	26,002	9,724	37.397%
Total	2,975,753	17,613	0.592%

The percentage of votes attributed to the two major candidates and the five minor candidates for both the election night count and the recount are quite similar, but with some instructive differences. With one exception, the minor party vote share remained virtually unchanged in the recount. In almost every county, the vote-share of the registered write-in candidates increased in the recount.[28] This is entirely consistent with the discussion above related to the counting of write-in votes on election night. What most likely explains the nearly uniform increase in registered write-in votes statewide is that in the recount, all ballots were scrutinized. This uncovered a number of ballots that contained write-in votes but lacked marks in the oval next to the write-in line.

Finally, the vote share of the scattering vote varied considerably among about half the counties. There were substantial inconsistencies in counting of the scattering vote. Fifteen of Wisconsin's 72 counties reported zero scattering votes in both counts, three counties reported a positive scattering vote on election night but zero in the recount, and six reported zero scattering votes on election night and a positive number of scattering votes in the recount. These patterns strike us as unlikely reflections of the actual distribution of scattering votes across the state. Most likely, the scattering vote that did exist was not counted at all in some counties (or at least not reported on the tally sheets). In counties where there was at least some counting of the

[28]The one notable exception here was Menominee County, Wisconsin's least populous county, which is coextensive with the Menominee Indian Reservation. It stands apart on the right side of the first scatterplot and left side of the second scatterplot. In the election night vote count, the county's sole reporting unit recorded zero votes for any candidate other than Trump and Clinton. In the recount, two votes were removed from Trump's count and one from Clinton. In addition, Castle (3 votes), Johnson (11), and Stein (17) were credited with votes. No write-in votes were recorded in the recount.

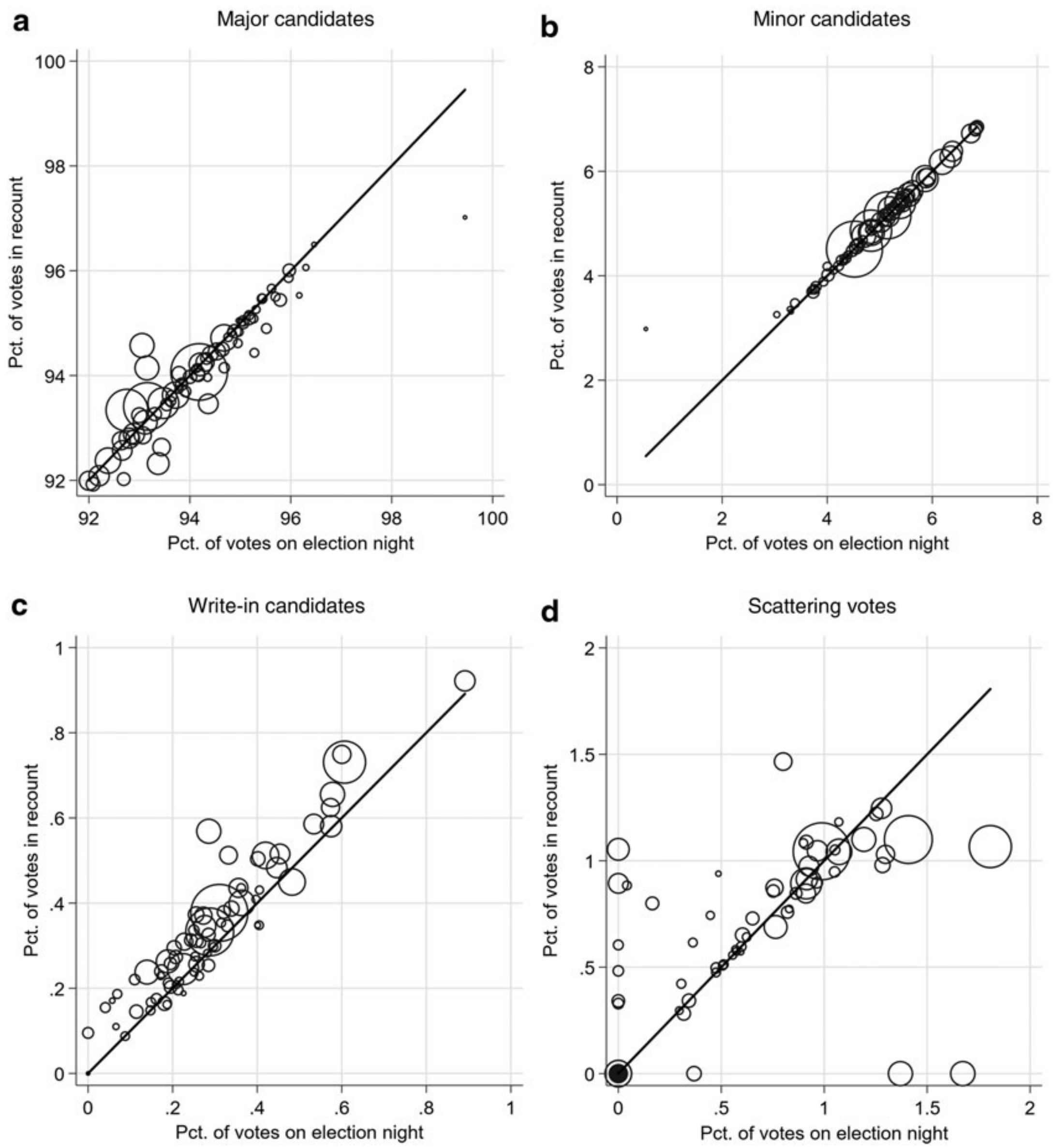

FIG. 2. Comparison of votes received by categories of candidates in Wisconsin on election night and in the recount 2016. *Note:* Major candidates = Trump and Clinton; minor candidates = other candidates printed on the statewide ballot (Castle, Johnson, Stein, Moorehead, and De La Fuente); write-in candidates = candidates officially registered to receive write-in votes (Fox, McMullin, Maturen, Schoenke, Keniston, Kotlikoff, Hoefling, Maldonado, and Soltysik); scattering votes = all other write-in candidates. Size of circles is proportional to the number of votes counted in each county on election night.

scattering vote, there was considerable variability in how thoroughly the municipalities and reporting units accounted for the scattering vote during the two tallies.

The ways municipalities and counties handled write-ins, both in the original count and the recount, had a significant impact on the overall error rate, as calculated by comparing the original canvass with the recount. It is not difficult to see why write-ins have such outsized influence on counting errors. Write-ins present specific challenges for both voters and election officials.[29] For scanned paper ballots, accurately counting write-in votes depends on poll workers carefully examining every ballot by hand, either to catch write-ins that did not have the

[29]The document "Reporting 'Scattering' Votes," cited above, notes uncertainty about how recent changes to Wisconsin election law affect the counting of write-in ballots, and states its purpose as ensuring "that all counties are reporting scattering votes uniformly. . . ." Clearly, the state elections board has achieved limited success in this regard.

corresponding oval marked, or to record write-ins correctly when ballots are counted by hand. Compared to a vote for a candidate printed on the ballot, write-in ballots provide more ways for a ballot to be inaccurately counted.

Write-in candidates contributed only about one percent of total votes cast but accounted for roughly half the absolute errors in the original vote tally. If only Trump and Clinton had been on the ballot in 2016, with no write-ins allowed, the absolute error rate would have been only 0.16%, slightly less than the 2011 recount, *even if we include the write-in vote from 2011*. If no write-ins had been allowed at all in 2016, and we only consider the seven candidates printed on the statewide ballot, the absolute error rate would have been 0.17%—which is also below the 2011 error rate. The addition of the registered write-in candidates raised the error rate to 0.27%, which is slightly higher than 2011. Finally, when we add the scattering vote, the error rate more than doubles, to 0.57%.

All of this suggests that simply using the pure absolute difference between the election night count and the recount complicates the idea of using the recount as ground truth. At the very least, the 15 counties that failed to count unregistered write-ins in either round of counting, plus the three counties that had counted the scattering vote on election night but not in the recount, should be excluded from any analysis that includes write-in votes as the basis for calculating vote count errors.

On top of that, the scattering-vote graph in Figure 2 also suggests that some municipalities and reporting units did not recount the scattering vote at all, even though other municipalities and reporting units in the same county did. In other cases, the reduction in the scattering vote may be the result of the correction of other errors. A good example is Waukesha County, which saw the total number of scattering votes reduced from 4,319 to 2,534, a drop that is due to double-reporting of write-in votes during the original count.[30]

For all these reasons, it appears that the best apples-to-apples comparison of error rates in Wisconsin focuses on the candidates printed on the ballot, excluding write-in candidates, both registered and unregistered.

With this in mind, Table 5 recalculates error rates for 2011 and 2016, this time using only the votes cast for candidates printed on the ballot. Focusing only on the candidates printed on the ballot in each election, the absolute error rates revealed by each recount are comparable for each degree of aggregation across the two years. Putting the two elections on a more common footing by comparing only the votes for listed candidates shows that the error rate did not increase over time. In fact, if we compared only the error rates for the top two candidates in each election, we would find a drop in the error rate between 2011 and 2016.

Table 5. Errors in Wisconsin Recounts, Using Only Candidates Printed on Ballot

Level of aggregation	*2011 Supreme Court*	*2016 president*
	Absolute error	
Reporting unit	2,762 (0.18%)	5,071 (0.17%)
Municipality	1,978 (0.13%)	4,093 (0.14%)
County	1,354 (0.090%)	2,581 (0.087%)
State	1,054 (0.070%)	1,731 (0.058%)
	Net error	
	1,054 (0.070%)	1,707 (0.057%)

We conclude this section by comparing error rates in 2011 and 2016 at the municipality level. The comparison within municipalities over time is useful because it reveals whether errors are largely idiosyncratic and thus display little continuity from one election to the next or whether they are endemic to particular jurisdictions and thus display significant continuity over time. Figure 3 graphs the absolute error rates in 2016 against 2011 where the circle sizes are again weighted by the number of votes cast in 2016. Figure 3a shows the error rate including write-in votes; Figure 3b shows the error rates calculated using only the candidates on the ballot. To assist in legibility that might be impeded by a small number of extreme outliers, the error rates have been transformed by taking cube roots.

The overall correlation between the absolute 2011 and 2016 error rates is a mere .058 when we

[30]See Waukesha County Board of Canvass, "Waukesha County Recount Summary and General BOC Meeting Minutes," p. 23, available at <http://elections.wi.gov/sites/default/files/recount_2016/waukesha_county_recount_summary_and_general_boc_me_13166.pdf>.

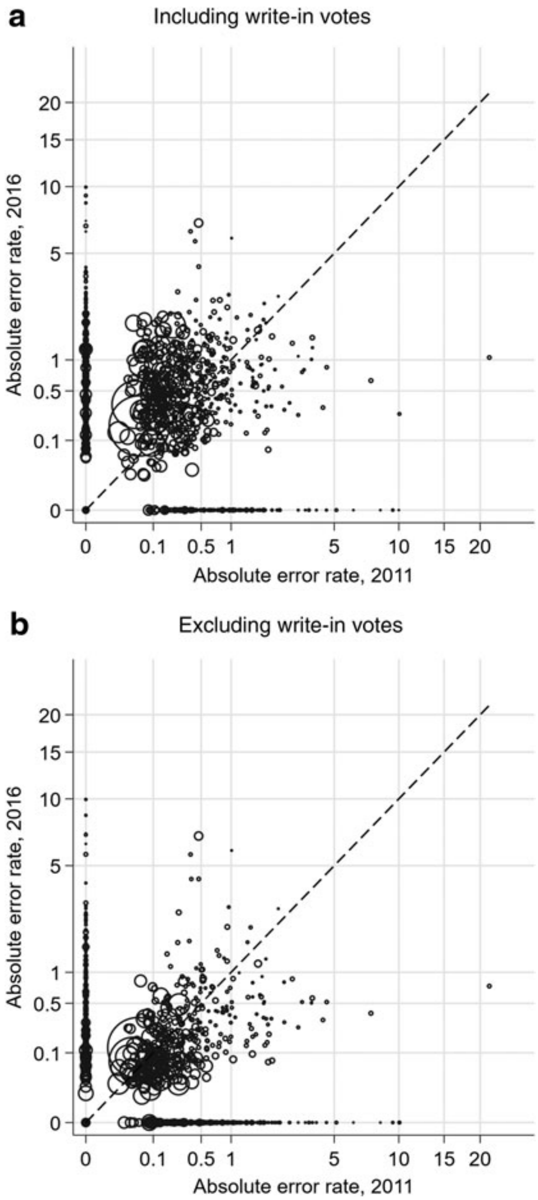

FIG. 3. Scatterplot of absolute error rates in Wisconsin in 2011 and 2016. *Note:* The error rate along the vertical and horizontal scales have been transformed by a cube root.

include write-in votes and .059 when we exclude them.[31] Figure 3 suggests one reason why the error rates are so weakly connected across the two elections: the modal error is 0%. Including write-ins, municipalities reporting no errors accounted for 63% of observations in 2011 and 38% of observations in 2016; excluding write-ins, these percentages rise to 66% and 56%, respectively. However, the large number of zeros is not responsible for the low correlation between 2011 and 2016; eliminating the municipalities with 0% error rates in either year produces a similarly low over-time correlation of .097 including write-ins and .14 excluding them.

It therefore seems that errors in vote tabulation are not endemic to particular communities over time but rather vary in somewhat unpredictable ways from one election to the next. This lack of relationship over time is in contrast to polling place "incidents," which show a substantial amount of continuity in their prevalence in particular communities over time in Wisconsin (Burden et al. 2017). These two facts are not as incongruous as it initially seems. This is in part because many incidents are in fact "benign" or even successful resolutions of potential concerns such as spoiled ballots. In these instances, the remedy by a poll worker on the "front end" helps to avoid a tabulation problem on the "back end."

VOTING TECHNOLOGY AND WISCONSIN RECOUNTS

A significant controversy surrounding the 2016 recount in Wisconsin was a claim that vote counts produced using computerized equipment—both ballot scanners and direct recording electronic (DRE) devices—are inherently suspicious and prone to error.[32] If this is true, then it was especially important to recount Wisconsin's votes, because the margin of victory was tight, and so many of Wisconsin's ballots had been counted on equipment that relied on computers to do the tabulation.

The criticism of computerized vote-tallying equipment as being unreliable, or at least less reliable than hand-counting paper ballots, is open to empirical test in states such as Wisconsin that rely on a mix of voting technologies to count the ballots. The most obvious test to conduct is whether paper ballots originally cast on paper and counted by scanner showed more discrepancies between the election night tally and the recount, compared to ballots that were originally

[31]The observations were weighted by total ballots cast in 2016. The correlations after transforming the rates by taking cube roots are higher, but still a meagre .073 and .10, respectively.

[32]J. Alex Halderman, "Want to Know if the Election Was Hacked? Look at the Ballots," *medium.com*, November 23, 2016, <https://medium.com/@jhalderm/want-to-know-if-the-election-was-hacked-look-at-the-ballots-c61a6113b0ba#.cph8nrhce>.

cast on paper and counted by hand. This is the cleanest test because paper ballots are verifiable by the voter, so the only material difference is the method of tabulation.

Including DREs in the comparison creates an ambiguous test because it is impossible for the voter to independently verify whether the votes he or she cast on the touchscreen were in fact recorded faithfully by the DRE's internal memory. A "hand" recount of DRE votes in Wisconsin means reviewing the paper record that is produced for each voter. Because these records are not subject to interpretation about voter intent, if there is a difference between election night and recount tallies in reporting units that use DREs, it is likely due to procedural issues related to the handling of absentee ballots in the reporting unit or transcription errors, not differences in how the DREs reported the outcomes from one time to the next.

The 2011 recount is about as clean a test as possible, since the recount was conducted entirely by hand, regardless of how the ballots were originally cast. The one significant departure from a clean test in 2011 is that a small fraction of ballots in 2011 were cast on DREs. We nonetheless report discrepancy statistics separately for ballots originally cast on DREs, because they help quantify vote-counting errors due to purely clerical errors.

The 2016 recount does not provide as clean a test as 2011. Although most jurisdictions recounted all their ballots by hand, even those that had been originally counted with scanners, some recounted optically scanned ballots by running them through the scanners again. Unfortunately, state records from the 2016 recount do not always clearly delineate which reporting units were recounted by hand and which were recounted by scanner. For the most part, counties reported that all ballots in their jurisdiction were recounted either by hand (51 of Wisconsin's 72 counties) or by scanner (nine counties). However, twelve counties reported that they employed a mix of optical scan and hand recounts, without specifying which municipalities used which recount methods.

We have scrutinized the minutes of the county election boards, with an eye toward discerning whether it was possible to determine the recount methods used across specific reporting units or municipalities in these twelve counties. On the whole, we were unsuccessful in producing a clean coding of the precise use of recount methods within these

Table 6. Net Counting Errors by Dominant Voting Technology in Wisconsin in 2011 (Absolute Errors Measured Based Only on Candidates Printed on Ballot)

Technology	*Mean absolute error*	*Number of reporting units*	*Number of ballots*
DRE	0.128%	270	49,283
Hand-counted paper	0.276%	179	66,705
Scanned paper	0.152%	1,911	1,050,670
Other	0.278%	1,084	332,222
Total	0.184%	3,444	1,498,880

DRE, direct recording electronic device.

counties. Therefore, we treat these twelve "mixed" counties separately from the counties that were either 100% hand or scanner recounts.

Furthermore, state records are not always clear about which method was used to count ballots on election night. A pre-election Wisconsin Elections Commission report on the voting technologies used by each municipality in 2016 is sometimes at odds with a post-election report that identifies the equipment used by each reporting unit.[33] In light of this disagreement of sources, we choose the post-election report, because it provides fine-grained information about how many ballots were counted by each type of voting technology at the level of reporting unit, whereas the pre-election report only provides information about voting technologies at the municipality level.

Turning to 2011 first, state records indicate that 81.3% of ballots were originally counted by scanners, 10.9% were counted by DREs, and 7.8% were counted by hand. On the whole, one technology type dominated each reporting unit, but even within reporting units there was some heterogeneity of usage. This is illustrated in Figure 4a, which shows the distribution of ballots counted by the three main voting technologies in each reporting unit in each year.

The spikes at 0% and 100% (indicating that all ballots in a reporting unit were cast via a single method) make it possible to classify most reporting

[33]Compare, for instance, the report of municipality voting equipment for 2016 (<https://web.archive.org/web/20170113061148/http://elections.wi.gov/sites/default/files/page/179/voting_eq_list_12_2016_xlsx_16214.xlsx) with the 2016 post-election report>, <https://web.archive.org/web/20170315013617/http://elections.wi.gov/sites/default/files/publication/2016_presidential_and_general_election_el_190_2017_83144.xlsx>).

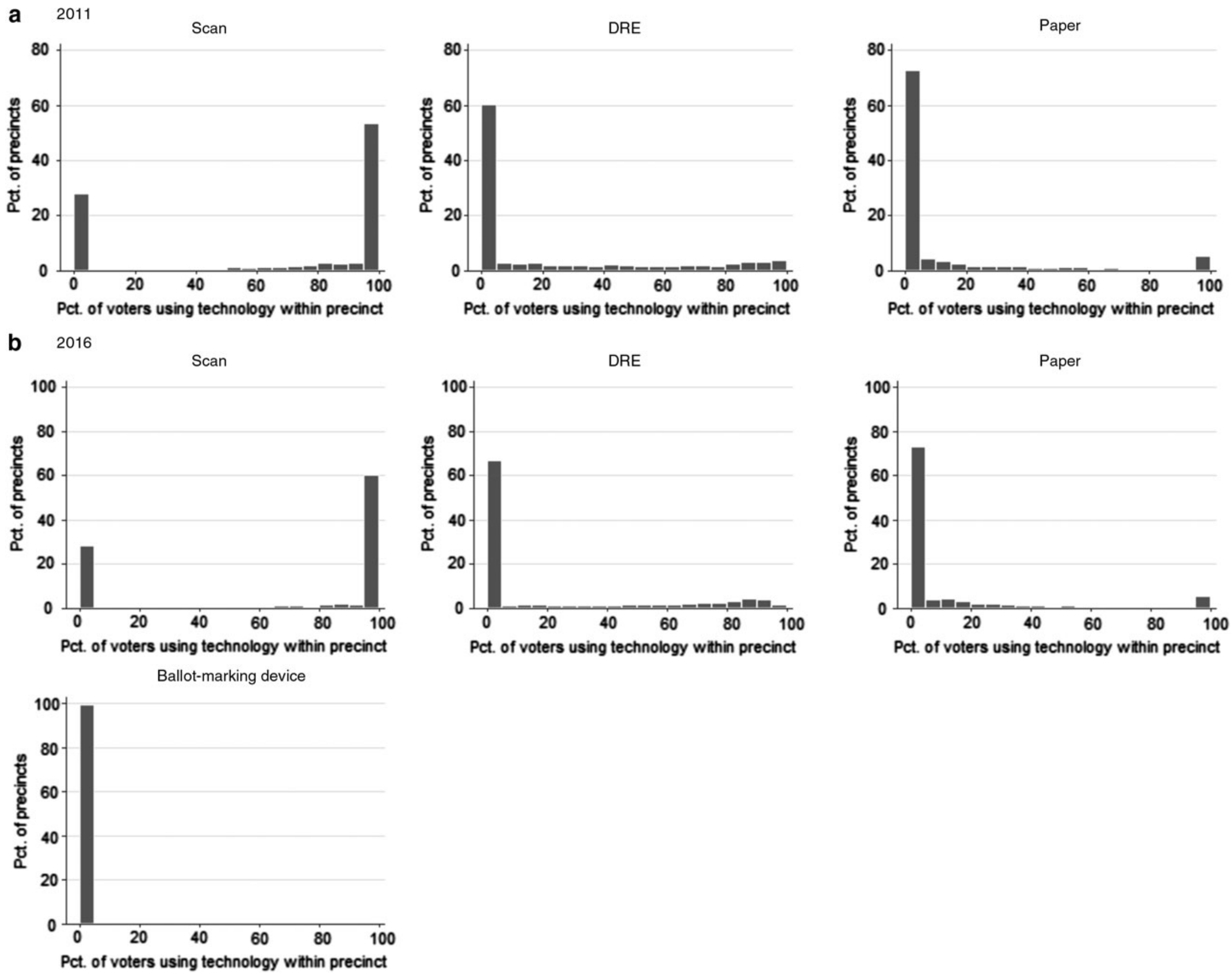

FIG. 4. Usage of voting technology in Wisconsin in 2011 and 2016. DRE, direct recording electronic device.

Table 7a. Net Counting Errors by Dominant Voting Technology in Wisconsin in 2016 (Absolute Errors Measured Based Only on Candidates Printed on Ballot): All Counties

Technology	*Mean absolute error*	*Number of reporting units*	*Number of ballots*
Ballot-marking device	0.000%	3	790
DRE	0.160%	185	54,859
Hand-counted paper	0.183%	194	169,635
Scanned paper	0.132%	2,145	2,205,278
Other	0.344%	1,109	508,731
Total	0.173%	3,636	2,939,293

Table 7b. Net Counting Errors by Dominant Voting Technology in Wisconsin in 2016 (Absolute Errors Measured Based Only on Candidates Printed on Ballot): Counties with Hand Recounts

Technology	*Mean absolute error*	*Number of reporting units*	*Number of ballots*
Ballot-marking device	0.000%	3	790
DRE	0.113%	139	44,080
Hand-counted paper	0.243%	153	114,348
Scanned paper	0.122%	1,056	1,065,495
Other	0.423%	757	328,359
Total	0.191%	2,108	1,553,072

Table 7c. Net Counting Errors by Dominant Voting Technology in Wisconsin in 2016 (Absolute Errors Measured Based Only on Candidates Printed on Ballot): Counties with Machine Recounts (Including Mixed Counties)

Technology	*Mean absolute error*	*Number of reporting units*	*Number of ballots*
Ballot-marking device	NA	0	0
DRE	0.350%	46	10,779
Hand-counted paper	0.060%	41	55,287
Scanned paper	0.142%	1,089	1,139,873
Other	0.201%	352	180,372
Total	0.148%	1,528	1,386,311

units into a category—predominantly scanner, DRE, paper, or ballot-marking device (for 2016). We classify a unit into one of these categories if at least 90% of ballots were counted using the associated technology. If no technology was used in a reporting unit to count more than 90% of a reporting unit's votes, it was assigned to an "other" category.

Table 6 reports the results of this analysis for 2011, the election in which the recount was done entirely by hand. In 2011, ballots originally counted by hand, and in reporting units with a mix of technology use (the "other" category) had the largest mean absolute error, at 0.276% and 0.278% respectively, whereas ballots originally counted on DREs[34] and on scanned paper had the lowest error (0.128% and 0.152%).[35]

We now turn our attention to 2016. Here we report the mean absolute error for all reporting units (Table 7a), and then separate the results for counties with full hand recounts (Table 7b) and for counties with at least partial machine recounts (Table 7c). Doing this allows for a more apples-to-apples comparison with the 2011 election among those counties that used hand recounts in 2016. Beginning with the counties with pure hand recounts, we see that reporting units that used DREs and scanned paper for the election night tally had the smallest mean absolute error (0.113% and 0.122%), much lower rates than in reporting units that used hand-counted paper (0.243%), and a mix of technologies (0.423%).[36] Although the presence of the "other" reporting units complicates things in Table 7c, the results for the hand recount in Table 7b reinforce our findings from the 2011 election.

As we demonstrated earlier, the major complicating factor in considering the election night-recount comparison is the matter of registered write-in votes and scattering votes.[37] As Figure 1 showed, registered write-in votes and scattering votes fared differently in the 2016 recount. On the one hand, virtually every county had more registered write-in votes in the recount than on election night.

[34]As noted above, the difference between election night and the recount for DREs is due either to clerical errors, such as transcription mistakes, or to differences in how absentee ballots were counted.

[35]An analysis of variance rejects the null hypothesis that the four counting methods had equal error rates at very high levels of certainty ($F=11.40$, $p<.00005$).

[36]A ballot-marking device (BMD) is a hybrid voting technology, which uses a touchscreen to receive a voter's choices, but then produces a paper ballot to be scanned. BMDs were grouped with DREs in the 2011 recount, but were reported as a separate category in 2016. We leave aside the two reporting units that used ballot-marking devices, because of the small numbers.

[37]Recall that the scattering vote consists of write-in votes cast for unregistered candidates and is not reported separately by candidate.

Table 8a. Counting Errors among Write-In Votes in Wisconsin in 2016: Registered Write-in Candidates

Technology	Mean absolute error	Mean net error	Number of reporting units	Number of ballots
Ballot-marking device	0.0%	0.0%	3	2
DRE	72.1%	0.0%	185	122
Hand-counted paper	45.5%	7.5%	194	683
Scanned paper	35.2%	11.8%	2,145	8,287
Other	128.4%	49.7%	1,109	1,364
Total	26.9%	16.3%	3,636	10,458

Note: Calculated using only counties that hand-counted the recount.

Table 8b. Counting Errors among Write-In Votes in Wisconsin in 2016: Scattering Vote

Technology	Mean absolute error	Mean net error	Number of reporting units	Number of ballots
Ballot-marking device	75.0%	75.0%	3	4
DRE	55.7%	−27.7%	185	307
Hand-counted paper	29.5%	−10.6%	194	1,675
Scanned paper	37.1%	−15.2%	2,145	21,977
Other	44.4%	17.8%	1,109	2,039
Total	37.4%	−12.5%	3,636	26,002

Note: Calculated using only counties that hand-counted the recount.

Indeed, the recount reported 1,928 more votes (12,386 versus 10,458) for the registered write-in candidates than were counted on election night. On the other hand, some counties had more scattering votes in the recount than on election night, some had fewer in the recount, and some reported precisely no scattering votes in either tally.

In the end, the recount reported 22,764 scattering write-in votes, compared to 26,002 counted on election night, for a reduction of 3,238. These differences amount to net error rates of 18.4% and −12.5% for the registered and scattering write-in votes, respectively, compared to the net error rate of 0.057% for candidates who were printed on the ballot (as reported in Table 5). While a negative error rate might seem nonsensical at first, these numbers show it is results from the "uncounting" of ballots in the recount that had been included in the election night count. When we calculate the *absolute* error rates for the registered and scattering write-in candidates alone, the rates are 27.6% and 37.4%, respectively, compared to the absolute error rate of 0.17% for the candidates printed on the ballot. These error rates for all of the write-in candidates, whether registered or not, are between two and three orders of magnitude greater than the error rates for candidates printed on the ballot.

An important question is whether some voting technologies were more prone to write-in counting errors than others. To answer this question, we calculated the counting error for write-in candidates, breaking down the percentages according to the technology used to count ballots on election night. Here, we look only at ballots that were recounted by hand, although the conclusions remain the same if we examine all ballots, regardless of how they were recounted.

Table 8 reports the results of this examination. Although there is some variation in the error rates by voting technology, all technologies showed substantial counting errors, both among the registered write-in candidates and the scattering vote. Because all write-in votes had to ultimately be tallied by hand, both on election night and in the recount, it seems most likely that these large counting errors are due to choices made by local election officials about how diligently to pursue these hand counts.

CONCLUSION

Recounts are significant in elections for many reasons. Most obviously, they provide an opportunity to double-check the counting that was done in the days immediately following the election, to either confirm or overturn the initial verdict of the election officials who administered the election. Using the proper measurement strategy, recounts also provide a glimpse into the accuracy of initial vote counts. When recounts are held in jurisdictions that use different methods to count ballots on election night, they can also provide additional insight into the relative accuracy of the different tabulation methods that are used to adjudicate winners and losers in most elections.

Using the mean absolute difference metric, we found that at least 0.21% of ballots were counted differently when they were recounted in 2011 and 0.59% in 2016. Using linear regression to extrapolate these error rates to the level of the individual ballot, we found that these error rates could have been as large as 0.26% in 2011 and 0.86% in 2016. Stated another way, these latter statistics are equivalent to one ballot in 385 in 2011 and one ballot in 116 in 2016.

The errors rates in 2016 were so much greater than in 2011 because the structure of the ballots was so different. The 2011 supreme court election was a competitive, simple two-candidate affair with little interest in minor party or write-in candidates. The resulting ballot was simple and easy to count. The 2016 presidential election was a complicated, closely contested affair with two major-party candidates, five minor-party candidates, nine registered write-in candidates, and countless idiosyncratic candidates who received a scattering of votes. The resulting ballot was long. A significant number of voters who chose to write in a candidate could easily make a mistake marking their ballots. Counting the ballots, especially the write-in votes, was tedious and prone to error, especially in the context of a busy presidential election night. Confining our attention to the candidates actually printed on the 2016 presidential ballot, the absolute error rate was nearly identical to 2011.

We suspect that many informed observers of election administration would be surprised to learn that ballots were so frequently miscounted, especially in a state that has a reputation for well-run elections. However, in comparison to the only other study of this type, which was conducted by Ansolabehere and Reeves (2012), who examined New Hampshire recount data, the error rates we observed in Wisconsin were small. Examining elections from 2000 to 2004, Ansolabehere and Reeves found an average absolute error rate among ballots initially counted by hand 1.98% and a rate among optically scanned ballots of 0.95%.

In reaching our conclusion about the magnitude of the counting error revealed by the Wisconsin recount, we have had to be attentive to the issue of measurement. We have shown that the most common statistic reported by the press to describe the difference in the vote count, which simply compares the original canvass to the recount, can dramatically underestimate the magnitude of the errors made in counting votes on election night. This measure, the net error rate, tends to cancel out errors, so that even if the number of miscounted ballots is relatively large, the net error rate can look small. Based on a comparison of the absolute and net errors reported in Table 5, which focuses on errors made in counting votes for candidates printed on the ballot, the magnitude of the absolute error was approximately three times that of the net error.[38] If individual ballot information was available to compute the true absolute error rate (or we rely on extrapolation from a regression analysis to approximate it), the net error rate would be even further from the actual accuracy rate.

It has been suggested to us that the net error rate is all that should matter to the public and students of election administration, because the purpose of elections is to choose leaders based on their popular support. If vote-counting errors tend to balance out across candidates, then the problem of vote-counting inaccuracies is minor. We disagree. Like all areas of democratic accountability, the legitimacy of elections rests on a public demonstration that the electoral process was managed competently. We do not believe that the goal of election administration should be to make sure that mistakes balance out, but rather, that mistakes be minimized, and that the few mistakes that remain not systematically advantage one candidate above the other. To that end, election administration should strive *both* to minimize absolute error *and* to have net error equal zero.

We also propose in this article a simple regression method that allows us to simulate the size of the absolute error we would observe if we were able to calculate the absolute error on a ballot-by-ballot basis, rather than having to rely on aggregate election returns. As ballot-based auditing techniques, such as risk-limiting audits, become more common, it will be possible to test the accuracy of this method directly.

We end with the ostensible topic that prompted the recount in the first place—skepticism about the accuracy of vote counts conducted with the assistance of computers. We find, as did Ansolabehere and Reeves, that vote counts originally conducted by computerized scanners were, on average, more accurate than votes that were originally tallied by hand. This finding should not be surprising, either to people who have administered elections or to those who have a grasp of the extension of automation into the workplace. Computers tend to be more accurate than humans in performing long, tedious, repetitive tasks. The demanding election night environment only drives a bigger wedge between human and machine performance.

[38]For instance, in 2011, the absolute error was 0.18%, compared to a net error of 0.070%. In this case, the absolute error was 2.6 times greater than the net error.

The fact that the average scanner is more accurate than the average human in counting ballots on election night is not an argument against checking the work of computers. Quite the opposite. The statistics presented in Table 2, for instance, show that, *at a minimum*, 0.59% of all ballots counted for president on election night in 2016 contained a counting error, which works out to one ballot out of every 169 cast. The regression technique described in this article suggests that if information had been retained about how each ballot was originally interpreted, 0.85% of ballots, or one out of every 117 cast, would have been shown to be in error. That ballot-counting errors can be so high in a state such as Wisconsin, which has a reputation for running clean elections, calls for greater attention to be paid to the initial vote count, and to the criteria used by vote counters in interpreting ballots. However, the analysis that compares error rates of optically scanned ballots with that of hand-counted ballots reveals the need for ballot-level audits of hand-counted ballots, too.

REFERENCES

Alvarez, R. Michael, Lonna Rae Atkeson, and Thad E. Hall. 2013. *Evaluating Elections: A Handbook of Methods and Standards*. New York, NY: Cambridge University Press.

Ansolabehere, Stephen and Andrew Reeves. 2012. "Using Recounts to Measure the Accuracy of Vote Tabulations: Evidence from New Hampshire Elections 1946–2002." In *Confirming Elections: Creating Confidence and Integrity through Election Auditing,* eds. R. Michael Alvarez, Lonna Rae Atkeson, and Thad E. Hall. New York: Palgrave.

Atkeson, Lonna Rae, R. Michael Alvarez, and Thad E. Hall. 2009. "The New Mexico 2006 Post Election Audit Report," University of New Mexico. <https://polisci.unm.edu/common/c-sved/papers/the-2006-post-election-audit-report.pdf>.

Burden, Barry C., David T. Canon, Kenneth R. Mayer, Donald P. Moynihan, and Jacob R. Neiheisel. 2017. "What Happens at the Polling Place: Using Administrative Data to Look Inside Elections." *Public Administration Review* 77:354–364.

Herron, Michael C. and Jonathan Wand. 2007. "Assessing Partisan Bias in Voting Technology: The Case of the 2004 New Hampshire Recount." *Electoral Studies* 26:247–61.

Lindeman, Mark and Philip B. Stark. 2012. "A Gentle Introduction to Risk-Limiting Audits." *IEEE Security and Privacy* 10: 42–49.

Tufte, Edward R. 1974. *Data Analysis for Politics and Policy.* Englewood Cliffs, N.J.: Prentice-Hall.

Address correspondence to:
Charles Stewart III
Department of Political Science
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

E-mail: cstewart@mit.edu