The Performance of Optical Scanners and DREs in North Carolina's 2008 General Election

Mark Lindeman, Bard College

In the 2008 general election, North Carolina counties generally used either M100 optical scanners or iVotronic DREs (Direct Recording Electronic voting machines) – or, in some cases, both. Recently I was asked whether it was possible to compare the performance of the scanners and the DREs. In particular, did the results of the post-election audit provide any reason to think that the DREs had performed better than the scanners? The evidence available to me indicates that in fact, optically scanned paper ballots fared better than DREs in recording and tabulating voter intent. The biggest difference is in *residual votes* for president – the proportion of (paper or electronic) ballots for which no vote for president was counted.[1] The residual vote rate was higher on the iVotronics, and this difference does appear to have something to do with the equipment, not other differences between "op-scan counties" and "DRE counties." The audit revealed few problems on either kind of equipment.

Please note: What follows is not offered or intended as a definitive analysis of the election data. I note some particular loose ends, which are unlikely to alter the basic conclusions.

Voting equipment in North Carolina and scope of my analysis

The state Board of Elections has made available a list of equipment used in each of North Carolina's 100 counties.[2] Most counties used either the M100 optical scanner supplemented by the AutoMARK for ADA (Americans with Disabilities Act) compliance, or the iVotronic DREs. However, nine counties used the M100 on election day and DREs in "one-stop" early voting. Of these nine counties, six also used the DREs for ADA compliance on election day, while three used the AutoMARK.[3] To reckon exact statewide statistics, one would have to get complete results for these counties by mode of voting, as well as figures for other paper ballots (such as provisional ballots). For present purposes, I compare results for the 24 counties that used DREs both on election day and in early voting and the 67 counties that used optical scanners both times. Limiting the analysis to these counties also evades the complication that early voters may have been more or less motivated to vote in the presidential contest.[4] The vote and ballot counts used here are official results provided by the North Carolina State Board of Elections as of June 5, 2009.[5] My data file, including data from the 2000 Census, is available for download at http://www.ncvoter.net/downloads/Lindeman_Spreadsheet_08_GE_Undervote.xls.

---

[1] Residual votes can be "undervotes" (no vote recorded, or too few votes recorded) or "overvotes" (too many votes recorded). To add to the possible confusion, some election reporting software distinguishes between "blank votes" (no vote recorded) and undervotes (at least one vote recorded, but fewer than the maximum possible). Of course, presidential contests are always vote-for-one.

[2] *2008 General Election Voting Equipment – North Carolina*, published on the Web at http://www.sboe.state.nc.us/getdocument.aspx?id=295

[3] The nine counties were Bladen, Catawba (AutoMARK), Craven, Cumberland, Forsyth, Moore, New Hanover, Sampson (AutoMARK), and Union (AutoMARK).

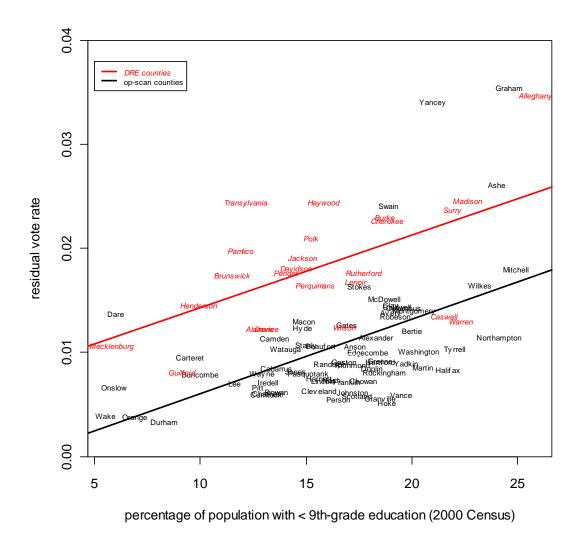[4] The nine excluded counties comprise about 16% of the North Carolina electorate, some 688,000 ballots out of 4.35 million votes cast. The residual vote rates in these counties range from 0.72% in New Hanover County to 1.55% in Catawba County, well within the range of counties included in the analysis. It is unlikely that including these counties would much affect the results reported here.

[5] The results are available in various formats at http://results.enr.clarityelections.com/NC/7937/14537/en/reports.html

Residual votes in op-scan and DRE counties

Overall, across the 91 counties in the analysis, the residual vote rate was 1.36% in "DRE counties" and 0.78% in "op-scan counties." (It was 1.03% in the nine mixed counties.) If we assume for a moment that the residual vote rate in DRE counties "should have" equaled the rate in op-scan counties, the difference of 0.58% equals roughly 7200 excess residual votes in these counties – that is, over 7000 North Carolinians who would have successfully voted for president if they had voted in an op-scan county instead of a DRE county.

Of course, op-scan counties and DRE counties vary in all sorts of ways, some of which might explain the difference in residual vote rate between them. However, while I have found several variables that help to "predict" (after the fact) the residual vote rate, those variables apparently have little if anything to do with the gap between op-scan and DRE counties. For instance, the proportion of each county's population with less than a ninth-grade education (as of the 2000 Census) is a moderately strong predictor of the residual vote rate. This relationship is depicted in the scatterplot below, with low-education levels on the horizontal axis and residual vote rate on the vertical axis. Separate best-fit lines are shown for the DRE counties (in red, names in italics) and the op-scan counties (in black, names in normal type).

The best-fit line for DRE counties is about 0.8 percentage points above the op-scan best-fit line, suggesting an even larger effect of DREs on residual votes than the 0.58% difference cited above – in part because relatively

fewer DRE counties had high proportions of low-education inhabitants. The best-fit lines are nearly parallel, suggesting that education affected residual vote rates similarly in both DRE counties and op-scan counties. (Note that this analysis does not prove that education or DRE use affected residual vote rates in any way: some other factors(s) may account for the statistical relationships depicted here.)

Extending the analysis, I found that poverty rates (% below poverty level) and racial composition (% "black only") also were moderately strong predictors of residual vote rates: counties with more poor people tended to have higher residual vote rates, while counties with more African-Americans tended to have lower rates. These three education, poverty, and racial variables plus DRE use jointly "explain" over three-quarters of the variance in residual vote rates ($R^2 > 0.78$) in a Weighted Least Squares regression.[6] However, the estimated "effect" of DRE use (0.74%) remains greater than the 0.58% raw difference, hinting that even more than 7200 North Carolina voters may have "lost" their presidential votes because of the DREs.

Different modeling approaches yield somewhat different estimates, and no retrospective statistical analysis can reveal what would have happened if optical scanners had been used statewide. Nevertheless, all the evidence I have found so far indicates that the DREs somehow frustrated thousands of voters whose presidential preferences would have been counted had they voted with optical scanners. Probably both major candidates lost thousands of votes in this manner; there is no clear evidence that one candidate was disproportionately affected.

Post-election audit results

North Carolina audited the vote counts in 273 "precincts/places" under the methodological direction of Dr. William D. Kalsbeek, Professor of Biostatistics at the University of North Carolina. Kalsbeek has coauthored a report on the results[7]. The report concludes *inter alia* that despite the closeness of the election – Obama won by about 14,000 votes out of over 4.3 million ballots cast – "the statistical probability that Obama in fact defeated McCain in the North Carolina election is higher than 99.9 percent" (p. 11). The report compares M100 and iVotronics, concluding that "M100 machines were more likely to undercount the election votes than iVotronic machines" (p. 7) but were no more likely to overcount them. More specifically, the report estimates that M100s had an undercount rate of 0.07%, while iVotronics had an undercount rate of 0. (The overcount rates were estimated as 0.007% for M100s and 0.005% for iVotronics; these rates are tiny and statistically indistinguishable.)[8]

Any DRE devotee who hopes to use "undercounts" as a talking point against optical scanners faces two big objections. First, the M100's undercount rate (less than 0.1%) was tiny, both absolutely and in comparison with the DREs' disadvantage in residual vote rates. Second, the scanner undercounts, by definition, could have been corrected in a full manual recount – whereas any excess residual votes on DREs cannot be corrected because the associated paper records do not reveal voter preferences. The fact that hand counts can uncover

---

[6] Each county's weight in the regression is based on the number of ballots cast. (Ordinary Least Squares regression yields broadly similar results, as does logistic regression.)

[7] William D. Kalsbeek, Lei Zhang, and Hengrui Sun, "An Assessment of the Recount and the Certification of the Election Result for the November 2008 Election," March 2009, available at
http://www.sboe.state.nc.us/GetDocument.aspx?id=1321

[8] What I here call "undercounts" and "overcounts," Kalsbeek et al. call "undervotes" and "overvotes." The latter terms are already widely used with different meanings (see footnote 1).

and correct scanner errors should be construed as a benefit of op-scan systems, not a disadvantage. As it happens, the audit uncovered few problems with either system. The audit of M100s revealed just a 4-vote net shift across 221 precincts/places, while the 54 iVotronic precincts/places yielded just a 2-vote net shift (see pp. 24-27).[9] As Kalsbeek et al. conclude, these results constitute strong evidence that a full manual recount would not have changed the outcome of this very close presidential race. This finding illustrates an important benefit of post-election audits: they can provide a basis for public confidence in the accuracy of the vote count.

Post-election audits have important limits in assessing voting system performance. Kalsbeek et al. state that they used their measures of undercounts and overcounts [see footnote 8] to "compare the accuracy of voting machines" (p. 7). While these measures contain valuable information, they are very limited as accuracy measures, because we have no way to compare voters' intentions with how their votes were recorded. We can infer from the higher residual vote rate on DREs that apparently some intended votes never were recorded, but we cannot tell *from the audit* whether some intended votes were recorded for a different candidate. The iVotronic VVPATs (voter-verifiable paper audit trails) do allow voters to check whether their votes are being recorded correctly, although it is not known how widely they are used. (The existence of excess residual votes indicates that the VVPATs do not completely resolve the problem of voter intent.) Barack Obama's county vote shares in 2008 are strongly predicted by John Kerry's vote shares four years earlier, with no evident systematic difference between DRE counties and op-scan counties.[10] This result is consistent with the assumption that overall, neither voting method substantially advantaged one candidate, although no strong test of the assumption is possible.

Mark Lindeman is an assistant professor of political studies at Bard College. He is co-author of the textbook *Public Opinion* (Westview Press) and an executive editor of the Principles and Best Practices for Post-Election Audits (available at http://electionaudits.org).

---

[9] These figures are reckoned by summing Frequency * Value for frequency tables as follows:
    M100 diff @candidate2 (M100 change in McCain votes):+38
    M100 diff @candidate3 (M100 change in Obama votes): +34
    iVo diff @candidate2 (iVotronic change in McCain votes): -2
    iVo diff @candidate3 (iVotronic change in Obama votes): -4
Two audited "places" recorded both M100 and iVotronic votes.
[10] Specifically, if one regresses Obama vote share on Kerry vote share and DRE use (using either OLS or WLS), the slope of DRE use is small and not statistically significant. Other specifications yield similar results.