A partial response to Merlin Klotz, Kenneth Horton, and other 7/5 comments
Mark Lindeman (writing for himself), 7/10/2017

Much of the content of the latest correspondence from Merlin Klotz and collaborators is already familiar, and I will not rehearse those arguments in detail. I think it is apparent that the Colorado legislature did not believe that the legacy audits of 1-7-514 C.R.S. met their new requirements for risk-limiting audits (RLAs) in 1-7-515 – and the legislature was right. It also is easy to confirm that many organizations and people, including Colorado election officials, have been involved in developing, implementing, and promoting RLAs. But some specifics here may bear further brief discussion.

- The fundamental problem remains that Klotz proposes to treat entire batches of ballots as if they constitute simple random samples of ballots. This is simply mistaken, much as ledger entries from, say, the first week of March and the second week of July cannot be treated as a simple random sample of all transactions in a year. Transactions from these two weeks may or may not be representative; much the same is true of batch counts.

- Kenneth Horton, as a justification for treating batches as simple random samples of ballots, avers that "Presently, there is no specific reason to believe that ballots are stacked in any way that would bias a sample towards a particular result." This is off point in at least three ways:
  - Perhaps most obvious, the lack of a "specific reason to believe" that a method is biased doesn't demonstrate that it isn't. Textbooks bristle with examples of sampling methods that were considered sufficiently random until they were shown not to be.
  - More subtly, "bias" is beside the point here. The problem with auditing just a few batches (or weeks) isn't that the sampling method is *biased*, but that we cannot effectively evaluate how representative the sample is. (Formally, these are *cluster samples* with just two clusters. A cluster sample of two batches tells us little about variability between batches, even if the batches contain hundreds of ballots.)
  - Even how the ballots are "stacked" is largely beside the point in evaluating scanner accuracy. Supposing for the sake of argument that the ballots were perfectly shuffled before being divided into batches, we still would not know whether the scanners interpreted and counted them consistently and accurately. This is what the tabulation audit is intended to determine, not to assume.

- Based on the faulty assumption just addressed, Klotz presents some dramatic but misleading workload comparisons. For instance, he cites Horton's calculation that a simple random sample of 298 ballots – if it finds no errors – suffices to confirm an election outcome with a reported, and true, winning margin of 2% (51%-49%) with 95% confidence. This figure in fact is similar to that for a sequential *ballot-level comparison*

RLA with a 5% risk limit.[1] The differences are that the comparison RLA actually does use a simple random sample of ballots, really can determine whether each ballot was interpreted correctly, and can readily expand as much or as little as necessary based on any miscounts found in the sample. A batch sample has none of these properties.

- Klotz's workload figures for RLAs appear to be based on ballot polling audits. This choice presumably reflects the fact that Klotz is from Douglas County, by far the largest of the few Colorado counties whose 2017 voting equipment cannot conduct ballot-level comparison audits. This reliance upon obsolescent technology is a matter of concern in its own right. It is not applicable to most counties – which already can conduct far more efficient and informative audits than Klotz here recommends – and should not be treated as a given.

- Klotz argues that the "Colorado Current Audit Protocol" comprises all election processes from ballot creation through election certification, whereas the RLAs in the draft rule are limited to one process among many. Of course, RLAs specifically supersede the legacy post-election audits, which are limited in the same way. To point to, say, "Ballot Proofing" as a rationale for scanting the vote tabulation audit makes little sense. Although some election processes presumably reduce the danger that scanners will perform inconsistently, they cannot be shown to *eliminate* that danger. The best way to assess tabulation accuracy is through direct evidence – and ballot-level comparison RLAs are an ideal way to gather that evidence.

---

[1] Using the method described in "A Gentle Introduction to Risk-Limiting Audits" (but adjusting the risk limit), a sequential RLA could terminate after 312 ballots if it found no discrepancies.