

Comments on RLA methods, workload and goals (3/13/2017)

From: Mark Lindeman (writing for myself)

To: Colorado Risk-Limiting Audit Representative Group

I have seen comments by Merlin Klotz that appear to reflect some confusion about methods and goals of risk-limiting audits. Let me first make a few points about methods.

Imagine a two-candidate contest in which Alice reportedly gets 53% of the vote and Bob gets 47% – close, but not very close. Consider three ways of auditing 100 ballots in that contest:

1. Randomly¹ sample 100 ballots and count votes for Alice and for Bob (“ballot-polling”).
2. Randomly sample 100 ballots, compare each one to the machine interpretation (Cast Vote Record), and see how many were miscounted (“ballot-level comparison”).
3. Randomly sample one *batch* of 100 ballots, count votes for Alice and Bob, and compare that count to the machine count of that batch (“batch comparison”).

#1 is not very efficient. You might get more votes for Alice, or for Bob (or a tie!). Even if you get a few more votes for Alice than for Bob, that won’t provide much evidence about who won.

#2 works a lot better. If you count 100 ballots and find no errors in the original count, you have fairly strong evidence that the error rate is low. Indeed, there’s less than a 5% chance of this audit result if the error rate is actually so high that Bob won.

#3 might sound like #2, but it has a big problem: the error rate might vary enormously across batches. Finding that one batch was counted correctly² tells you almost nothing about any other batch. Suppose that for 10% of batches, due to a programming error, every vote for Alice was counted as a vote for Bob and vice versa. If you audit one batch of 100 ballots, you only have a 10% chance of finding that problem. If you audit 100 separate ballots, you are almost certain to.

This efficiency advantage is why we advocate ballot-level comparison audits wherever possible.

Merlin Klotz suggests that “For 95% confidence that the outcome is within +/- 3% on 300,000 ballots cast the sample size needs to be 1,063.” That calculation assumes that you are only counting votes on the sampled ballots – in effect, that you are doing a ballot-polling audit.³ If you can check whether each ballot was counted correctly, 100 randomly sampled ballots can get you close to the equivalent of 95% confidence (a 5% risk limit). (Counting eleven batches of 100 is not nearly as useful for that purpose, even though it comprises many more ballots.)

¹ I’m using “random[ly]” as shorthand for a *Simple Random Sample*, in which each unit (ballot or batch) is sampled independently of the others and with equal probability. (For batch audits, equal probability is not always best – but for my purposes here, that doesn’t matter.)

² Strictly, you can’t even know that: maybe some ballots were miscounted in opposite ways.

³ We would use different math for ballot-polling, but again that doesn’t matter here.

Auditing multiple statewide contests

Two quick points here. First, it is valid to audit multiple contests on the same ballot; for statewide contests, it is simple. So auditing (say) eleven statewide contests may not require sampling eleven times as many ballots. (Of course, in RLAs, contests with smaller margins typically require more sampling – but the samples can overlap.)

Second, *auditing has multiple goals*. Obtaining strong evidence that statewide election outcomes are correct is one possible goal – one that I enthusiastically favor. To decide how to meet other goals, first we have to state those goals. In the previous example, suppose that Alice and Bob were candidates for governor. It can be said (not quite correctly) that we can confirm the result of their contest at a 1% risk limit by comparing just 160 ballots sampled randomly from around the state to the corresponding CVRs.⁴ Obviously, that does not inherently mean that we should audit only 160 ballots in the entire state. We will probably want to know more about election processes in each county. At the same time, it does not inherently mean that we should audit the governor’s contest on 160 ballots in each county. That work is not needed in order to confirm the statewide outcome, and we may have better ideas about what counties should do with their auditing efforts.

It simply isn’t the case that anyone’s “RLA theory” would require some counties to pull batches containing “more than 300% of the ballots cast.” That conjecture misses three separate points: the distinction between ballot-polling and ballot-level comparisons; the possibility of auditing multiple contests on the same ballot; and the need to specify particular auditing goals before determining how to satisfy them. As the previous example shows, statewide ballot-level comparison audits can be very efficient indeed.

“Risk-limiting” is a feature, not a blueprint or straitjacket

Generalizing about risk-limiting audits is a bit like generalizing about cars with adaptive cruise control: they don’t necessarily have much in common. The math of RLA is designed to support various strategies for providing strong evidence that the outcomes in particular contests weren’t altered by tabulator error. Obviously, the math doesn’t stand alone, and it doesn’t answer every important question about audit design; that is not its purpose. It works with other design aspects, not against them.

⁴ This figure assumes that every ballot in the state has a corresponding CVR, and that no errors are found. Other circumstances might require more work (or possibly less!).