The State of Risk-Limiting Audit Implementation

Ginny Vander Roest, Implementation Manager, VotingWorks

In 2017, Colorado became the first state to implement risk-limiting audits (RLAs), a type of post-election tabulation audit. While RLAs are considered the "gold standard" of tabulation audits, other states and counties interested in implementing, knew their tabulation processes would not allow them to conduct RLAs in the same manner as Colorado. In the last four years, many states have successfully worked through those differences and are now conducting their own RLAs. Their efforts (highlighted below) are paving the way for easier adoption in every state (perhaps without requiring a legislation change in some).

Risk-limiting audits use a statistical sampling method to select ballots or batches of ballots for audit to ensure the tabulated outcome is correct to a desired level of statistical significance. The audit requires more ballots or batches to be sampled when contest margins are small and fewer when contest margins are larger. There are a few different types of RLAs and new types are being designed as implementation challenges are revealed.

The RLA type most piloted first is ballot polling. Ballot polling requires strong chain of custody and inventory of the audit trail (paper ballots), and allows any jurisdiction, regardless of voting method or voting equipment, to conduct an RLA.

A handful of states who started with ballot polling have since moved to piloting or implementing the batch-comparison RLA type. Like ballot polling, batch comparison requires strong chain of custody and inventory of the audit trail as well as a record of the candidate vote totals by batch. Batches can be as simple as all ballots for a specific precinct (all voting systems provide candidate totals for this type) or more complex like large early voting batches or random batches of ballots from any precinct run through a high speed tabulator (not all voting systems provide this easily).

The last RLA type in this discussion is ballot comparison. Ballot comparison also requires strong chain of custody and inventory of the audit trail as well as keeping each ballot in the order it was scanned, preferably with a unique identifier printed on the ballot at the time of scanning. Only high speed tabulators are capable of producing the cast vote record linked to a physical ballot.

As states have worked through implementation of these RLA types, they've been met with various challenges. Implementation in Colorado took eight years; many jurisdictions have been forced to move more quickly due to legislative directives. Most states have struggled to ensure all jurisdictions participate without a legal mandate and many are implementing while also maintaining a traditional audit program (usually a tabulation review of a fixed percentage of ballots) or implementing various new audit programs.

In 2018 & 2019, Michigan, Georgia, Pennsylvania, Rhode Island, and Virginia began piloting ballot polling RLAs by working with experts in the field and conducting small scale pilots. Different counties and jurisdictions verified the process worked on every paper ballot tabulation system and was truly implementable in states with tabulation methods different from Colorado's central count system. These pilots proved to be very successful and pushed each state to pilot statewide RLAs in 2020 elections.

Pilot programs are a valuable and necessary part of the process when implementing RLAs. We knew the success of local pilots in these states did not reveal enough about the process to truly show how it would scale when conducted statewide by all counties or jurisdictions simultaneously. More challenges would be discovered making it important to keep the pilot label (when not required by law) for the 2020 elections.

The most significant lesson learned in 2020 was that statewide conduct could be done. Not only with just one or two years of piloting but also during the height of the pandemic. Were they conducted to every high standard of an RLA? Not quite, but they were not far off and were better than any existing auditing methods. Presidential Primaries ended up being ideal first time statewide pilots with fairly wide contest margins providing a smaller number of ballots to review and a fairly light workload for any jurisdiction selected.

The November 2020 General election was quite different from the primary in most of the pilot states. Contest margins were much closer and in Georgia, the winning margin was only 11,779 votes or .23%, the closest margin in the country. Ballot polling RLAs quickly become a challenge when margins are under 2%, and in Georgia the sample size was over 1 million of the 5 million ballots tabulated. At that large of a number, a full hand tally was more efficient to ensure the outcome was verified (a full hand count is essentially a risk-limiting audit with a 0% risk-limit). While Michigan's and Pennsylvania's contest margins were not as close as Georgia's, they still required somewhat burdensome sample sizes.

Despite large sample sizes, the pilots were still successful, even Georgia's full hand tally of over 5 million ballots. However, RLAs were adopted for efficiency and to increase the public trust in the outcomes. These large sample sizes were not the most efficient and states questioned if they were sustainable. Ballot polling also proved to be difficult to describe to the public in a manner that was easily understood, and therefore it may not help to build the public confidence as intended.

In 2021 (and earlier in RI), GA, PA, and MI started piloting batch comparison audits. Batch comparison was less understood as there was no public calculator (there is now) to calculate sample sizes and provide states with an idea of what their typical sample sizes might be. Batch comparison also has the potential to be a very large workload in a small election because sample sizes do not change based on the number of ballots or batches in a jurisdiction. A contest with a 10% margin and a 5% risk-limit will require 32 batches of ballots to be reviewed whether there are 32 batches or 1,000 batches. RLAs are essentially more efficient the larger the election.

Batch comparison audits also scale up easier than ballot polling audits. For example, Georgia's 2020 Presidential race (.23% margin) would have required 1,274 batches to be audited, far fewer than the ~13,000 batches typical in their general elections. Batch comparison audits provide more intuitive results that are much closer to traditional audits. Tabulated batch totals are compared with the audit board's hand-count, allowing election officials and the public to see matches and discrepancies easily.

Given all of this information, those four states chose to move to batch comparison for their November 2022 General election. (Note: Virginia continued to use ballot polling as their contest margins generally allow for an efficient RLA). While one of the audits is still in process as of the submission of this paper (recounts and RLAs are the subject for a future paper), the other three were able to complete their RLAs within proposed timelines and successfully verify the outcomes. Of course, the margins in these states were a little larger than 2020 but they proved that a larger workload was still manageable.

In fact, Georgia chose to select more batches than those selected in the RLA. Once RLA batches were drawn, the selection tool drew additional batches so that at least two batches—one BMD-marked and one hand-marked—were selected in each county. Their original RLA sample size was 36 batches across 159 counties with 292 additional batches selected. Similarly, Michigan randomly selected additional batches for audit, based on their legacy performance audit selection requirements.

In just two general election cycles, these states have managed to implement the majority of RLA requirements. There are still tweaks to make and additional transparency measures to pursue, but they've built their own confidence through the process of piloting and will only improve over time.

There are likely many states who currently conduct traditional audits who could roll out batch-comparison RLAs (and maybe some extra batch selection) with very little change to current processes. This is especially true in states with wide margins sampling more than 2% of batches.

While all of this occurred, Nevada, Washington, and some California counties have been piloting ballot comparison successfully. These states are heavily vote-by-mail, with considerable practice in managing ballots including being able to find any exact ballot the voting system tabulated. This is no small task and should not be excluded from the conversation of RLA progress. Significant progress has been made in all RLA implementation in just four years.

This paper only skims the surface of RLA requirements, lessons learned, and pros and cons of implementing each type of RLA. While detailed information about RLAs can be found at risklimitingaudits.org, we'd recommend starting a conversation with election officials from one or a few of the states mentioned in this paper. We'd be remiss if we didn't mention that pilots have also been conducted in Boone County, MO, Connecticut, North Carolina, Ohio, South Carolina, and Texas, most of whom are just starting to explore RLA feasibility in their states.