# Post-Election Audit Workgroup Report on Risk Limiting Audits

Date: September 30, 2022

## INTRODUCTION

Pennsylvania has made significant investments towards improving the security of its election system over the past few years. The state took a crucial step by upgrading its voting equipment to ensure the usage of paper ballots statewide going into the 2020 election cycle. Paper ballots allow voters across the state to confirm that their voted ballots reflect their intended choices. These voter-verifiable paper ballots make it possible to implement rigorous post-election audits that confirm that votes were counted accurately.

The state's election code requires each county to conduct a "statistical recount of a random sample of ballots" amounting to 2% of the votes cast, or 2,000 votes, whichever is less. While this audit is automatic and performed before the election results are certified, Pennsylvania can do more to confirm election outcomes and to bolster voter confidence without burdening election officials.

In January 2019, the Pennsylvania Department of State convened the post-election audit workgroup to study the implementation of robust post-election audits in the state. The workgroup is composed of county election directors, representatives of the Department of State, and other election experts. The primary objective of the workgroup is to identify and recommend robust post-election, pre-certification audit procedures that are suited to the election system and practices in place in the state. The post-election audit workgroup was tasked with identifying more rigorous alternatives to current post-election audit practices and with preparing for pilots to test these new risk limiting audit methodologies.

The aim of robust audit procedures is to strengthen election security and integrity, confirm the accuracy of election outcomes, and provide confidence to voters that their votes are being counted accurately. Accordingly, the workgroup decided to evaluate risk limiting audits (RLAs), an audit method recommended by the U.S. Department of Homeland Security, the U.S. Senate Select Intelligence Committee, the U.S. Election Assistance Commission, and many other experts as one element of a strong and resilient election infrastructure. An RLA is a best-practice post-election audit that manually checks a random sample of voters' paper ballots to provide statistical evidence that the reported election outcome is correct. If an error occurs that could impact the outcome of an election, the RLA is likely to detect that fact.

The workgroup began by piloting RLAs at a county level in 2019, with sandboxes (informal experiments) in Lancaster and Philadelphia counties and pilots in Mercer and Philadelphia counties. The Department of State then coordinated statewide pilots in 2020 and 2021 to provide all counties direct experience with RLA methods. The recommendations in this report are informed by these pilots, as well as extensive workgroup discussions. We believe that these recommendations provide a practical path to reinforcing security, ensuring accuracy, and bolstering public confidence in election results.

## BACKGROUND

### Best practices and key design principles for risk limiting tabulation audits

Following the success of its recent statewide upgrade of voting equipment, Pennsylvania is now in a position to implement more rigorous audit requirements which will enhance the overall security of the election system. Post-election RLAs serve as a check that the computers tabulating ballots functioned correctly, reinforcing public confidence in election outcomes. For

RLAs to be effective and informative, however, they must fulfill certain principles and best practices.[1]

**Manual examination.** First, ballots should be manually examined by auditors. As expressed above, the purpose of an RLA is to independently verify that the results derived from the existing tabulation procedures (mainly, the use of computers) are reliable. Thus, while it is appropriate in other contexts, such as a statutory recount, to rely on data produced from computers — for example, by using digital ballot images or data obtained by running the ballots through another set of scanners (perhaps of a different make or model) — an RLA should not rely on such computer-derived data. If any ballots were remade by election officials, auditors should be able to identify and audit the ballot originally voted by the voter.

**Comprehensive scope.** RLAs should also be comprehensive. All cast ballots should be eligible for review as part of the RLA, whether they were cast on election day, by mail, provisionally, or by UOCAVA voters. Ideally, no contest should be exempt from possible review, though some contests (such as top-of-ticket contests) may be prioritized.

**Pre-certification timing.** RLAs should be performed before election results are finalized and should be capable of detecting and correcting tabulation errors that alter outcomes. RLAs have a large chance of correcting a reported election outcome that is incorrect due to a tabulation error.

**Chain of custody and ballot-accounting procedures.** Strong chain of custody and ballot-accounting procedures are key prerequisites for best-practice RLAs. These procedures help ensure and demonstrate that the ballots provide a trustworthy record. Management and storage of paper ballots must also allow for easy ballot identification and retrieval since ballots are randomly selected for review.

**Standards for review and transparency.** RLAs should incorporate high-transparency standards. Observers should be able to observe the risk limiting audit in accordance with existing statutes for observing the canvass of votes.

All these principles are important, but risk limiting audit design must acknowledge tensions among them. For instance, as the workload of a risk limiting audit increases, it can become more difficult to provide opportunities for public observation. If a risk limiting audit requires reviewing a large number of ballots and/or many contests, election officials may divide the work between multiple teams. One or two observers cannot closely monitor all these teams. In the same vein, some types of risk limiting audits can be easier to observe than others because of differences in how many ballots are handled and examined. Observing full-hand counts poses particular challenges due to the sheer number of ballots involved.

Later in this report, we discuss specific procedural elements that are integral to best-practice RLAs. Because some procedural details depend on basic design choices, we first turn to discussing risk limiting audit methods, how to choose among them, and how the choice may influence contest selection.

---

[1] Our discussion here draws on the widely endorsed Principles and Best Practices for Post-Election Tabulation Audits (December 2018), available at https://verifiedvoting.org/wp-content/uploads/2020/05/Principles-and-Best-Practices-For-Post-Election-Tabulation-Audits.pdf .

**Brief introduction to risk limiting audit methods**

A **risk limiting audit** of an election contest, by definition, has a predetermined minimum chance of leading to a full-hand count in that contest *if* a full-hand count would change the outcome. If the voted ballots contradict the reported outcome, a risk limiting audit is very likely to correct the outcome through a full-hand count. For instance, *if* an outcome is wrong, a risk limiting audit with a 10% **risk limit** has a guaranteed 90% *minimum* chance of correcting it. (The *actual* chance of correcting a wrong outcome may be much larger, depending on the specific differences between the reported and actual results.) When a risk limiting audit does not expand to a full-hand count, it provides strong evidence that the reported outcome was correct. A risk limiting audit can cover more than one **target contest**, providing strong evidence for each. Risk limiting audits also can be combined with less stringent audits of other contests.

Risk limiting audits, and related statistical audits, can use any of three basic methods, which vary in their requirements and other attributes: ballot comparison, ballot polling, and batch comparison. The first of these, **ballot comparison**, is not currently feasible in Pennsylvania because it requires the ability to match each individual ballot with its "cast vote record," the digital record of how the voting system interpreted that ballot. Typically, to implement ballot comparison audits, ballots are centrally scanned in an arbitrary order so that nothing links each ballot to the voter who cast it and imprinted with **sequential** serial numbers. In Pennsylvania, most ballots are cast in person; the hand-fed scanners in polling places cannot, and *must* not, imprint the ballots with sequential serial numbers that could be used to trace each ballot back to a voter. Short of centrally *re-tabulating* and imprinting these in-person ballots, they are not amenable to ballot comparison risk limiting audits.[2]

When ballot-comparison risk limiting audits are feasible, they generally require examining the fewest ballots. A random sample of ballots is selected; audit boards examine each ballot in the random sample and interpret the vote(s), if any, in the contest(s) being audited. Each audited ballot provides an independent check on voting system accuracy: The human interpretation of the ballot may or may not match how the voting system interpreted the ballot, as recorded in the cast vote record.

The other two methods, ballot polling and batch comparison, both have been successfully used in statewide RLA pilots in Pennsylvania. Each method has distinctive advantages and complications. Here we briefly summarize the methods and their general characteristics.

**Batch comparison** audits are the most common method for sampling ballots nationwide and are widely used in states that do not require risk limiting audits. In a batch-comparison risk limiting audit, the ballots have been counted and stored in separate batches, such as one batch per precinct. A random sample of *batches* is selected; audit boards hand-count the votes for each candidate/choice in each batch, and the risk limiting audit results for each batch are compared with the originally reported totals. Thus, batch comparison requires vote totals for each batch. Batch-comparison RLAs work best when ballots are divided into a large number of batches; for instance, the method generally is better suited for statewide elections than for local elections.

---

[2] Experts have proposed a variation of ballot comparison in which some, but not all, in-person ballots would have to be re-tabulated. This approach may be feasible in Pennsylvania in the future but has never been implemented at scale.

**Ballot-polling** risk limiting audits, like ballot-comparison risk limiting audits, involve selecting a random sample of ballots (not batches) and manually interpreting the vote(s) on each ballot. However, in a ballot-polling risk limiting audit, the manual interpretation of each ballot cannot be compared to the cast-vote record. Instead, ballot-polling RLAs use the RLA sample rather like responses to a public opinion survey: Loosely speaking, the RLA continues until the reported winner leads, in the RLA sample, by beyond the sample's "margin of error."[3] Compared to the other two risk limiting audit methods, ballot polling requires less information. The ballots still are stored in batches, and the risk limiting audit uses a **ballot manifest**: a record of how many ballots are stored in each batch. However, *vote* counts for each batch are not needed or used. Ballot polling often is more feasible than batch comparison in small contests with relatively few batches. As we discuss below, batch comparison has operational advantages in large contests.

### Risk limiting audit pilots

By planning and conducting or coordinating multiple pilots at a county level and statewide, the workgroup has identified various factors that contribute to a successful risk limiting audit process, in order to inform recommendations such as those in this report. Concurrently, election officials around the state have gained practical experience with both batch-comparison and ballot-polling risk limiting audits. Here we briefly sketch the timeline of these pilots and key points about each one.

Page Break

| Pilot Chronology | | | | |
|---|---|---|---|---|
| **Date** | **County / Statewide** | **Contest(s) Audited** | **Method** | **Notes/Milestones** |
| June 27, 2019 | Lancaster (sandbox) | 3 contests from Nov. 2017 municipal election in 13 precincts | Ballot polling | 100 ballots audited |
| August 23, 2019 | Philadelphia (sandbox) | Mock election with 4,810 ballots in 40 divisions | Batch comparison | Focused on training, testing methods for hand counting batches |
| November 18, 2019 | Mercer | Constitutional amendment | Ballot polling | 79 ballots audited of 23,667 ballots cast |
| November 20, 2019 | Philadelphia | Mayor and constitutional amendment | Ballot polling | 60 ballots audited of over 300,000 ballots cast |
| August 2020 | Statewide | Presidential primary (Democratic and Republican) | Ballot polling | 400 ballots audited statewide, 53 participating counties |
| Following November 2020 general election | Statewide | Presidential election | Ballot polling | Over 47,000 ballots audited (of well over 6 million cast in the 63 participating counties) |

---

[3] Ballot-polling RLAs do not use the same "margin of error" formulas that surveys do, but the statistical theory is similar.

| May 2021 municipal primary | Statewide (Precertification) | Republican Justice of the Supreme Court | Batch comparison | 59 batches audited, 36 counties participating |
|---|---|---|---|---|
| May 2022 General Primary | Statewide | Democratic Lieutenant Governor | Batch comparison | 44 counties participating overall, 50 batches audited by 19 counties |

Additionally, the state planned a statewide batch-comparison RLA pilot following the November 2021 municipal election. Although this pilot, ultimately, was canceled due to an automatic recount in the Commonwealth Court contest, election officials throughout the state appeared well-prepared to conduct the pilot.

**Understanding risk limiting audit workload**

The workgroup was not tasked with choosing a single "best method" for audits in Pennsylvania, and we explicitly warn against codifying a specific method in statute. We instead recommend the General Assembly use general language to codify audits. Different methods have different strengths and requirements; audit methods should be able to adapt to unexpected changes in circumstances or election procedures without rewriting state law. That said, it is important to understand how the choice of audit methods affects counties.

We begin by discussing audit *workload*. Conceptually, workload in audits can be divided into five phases: preparing for the audit; retrieving the ballots to be audited; examining the votes on the ballots; submitting the risk limiting audit results; and wrapping up the risk limiting audit (for instance, returning the ballots to storage). In RLAs of close contests, retrieving and examining ballots often take the most work, and we focus on those phases here.

Some background on sample sizes is helpful to understand workload in RLAs. A crucial fact: the informativeness of a random sample generally depends primarily on the **sample size** (the *number* of units sampled), not the sampling *fraction* (what *proportion* of units is sampled), unless the fraction is very large. In public opinion surveys, interviewing 1,000 people has about a 3% margin of error whether the survey spans one city, a state, or the entire country. Similarly, in RLAs, the initial sample size depends heavily on the risk limiting audit method, the risk limit, and the reported margin (as a percentage of ballots cast), but much less on the total number of ballots cast.

The table below shows expected sample sizes to complete batch-comparison and ballot-polling RLAs at a 10% risk limit for various margins. These numbers cannot be compared directly. Batch-comparison risk limiting audits usually require *examining* many more ballots than ballot-polling risk limiting audits do, for the same margin, because a batch can contain hundreds or thousands of ballots. Retrieving ballots for batch-comparison risk limiting audits, by locating and opening the batches, usually is a small part of the work. Because ballot-polling samples are drawn at random from all ballots, they can require opening many different batches and retrieving ballots at random from each batch,[4] so *retrieval* usually takes substantially longer than examining the ballots.

---

[4] To obtain a trustworthy sample, these ballots must be retrieved truly at random; an ad hoc approach such as "grabbing some ballots from the middle of the batch" does not suffice. A detailed discussion of alternative methods appears in the Rhode Island RLA Working Group's *Pilot Implementation Study of*

Altering the risk limit changes the sample sizes, although not as much as one might expect. In general, compared to a 10% risk limit, a 20% risk limit requires about 30% less work; a 5% risk limit requires about 30% *more* work; a 1% risk limit requires double the work.

Importantly, ballot-polling samples increase more rapidly than batch-comparison samples at smaller margins. For instance, with batch comparison, it takes about three times as many batches to audit at a 1% margin as it does at a 3% margin; with ballot polling, the smaller margin takes about nine times as many ballots.

### Initial sample sizes for RLAs with 10% risk limit at various contest margins

| contest margin | batch comparison | ballot polling[5] |
|---|---|---|
| 25% | 12 batches | 116 ballots |
| 20% | 14 batches | 179 ballots |
| 15% | 18 batches | 311 ballots |
| 10% | 26 batches | 701 ballots |
| 8% | 32 batches | 1,100 ballots |
| 5% | 50 batches | 2,780 ballots |
| 3% | 82 batches | 7,675 ballots |
| 2% | 122 batches | 17,272 ballots |
| 1% | 242 batches | 68,762 ballots |

A further complication of ballot-polling RLAs is that the results are less predictable than for batch-comparison risk limiting audits. A ballot-polling sample can underrepresent a winning candidate's vote share through random chance alone, even when the original count is highly accurate. The sample sizes in this table for ballot-polling RLAs provide a 90% chance that the RLA will reach the risk limit and end in one round. About one in ten RLAs would require continuing to a second round because the first-round results were inconclusive. One can reduce this chance by auditing even more ballots in the first round or one can audit fewer ballots with a greater chance of requiring a second round. Batch-comparison RLAs are far more predictable.

---

*Risk-Limiting Audit Methods in the State of Rhode Island*, available at https://verifiedvoting.org/wp-content/uploads/2020/07/RI-RLA-Report-2020.pdf.

[5] This is the ballot-polling sample size required to have a 90% chance of completion after one round, using the Minerva method.

Continuing to a second round is unlikely in a batch-comparison RLA unless the RLA finds very substantial discrepancies – not the minor discrepancies (often attributable to nonstandard voter marks or human error in counting) that risk limiting audits usually find.

Again, these sample sizes apply (approximately) unless the sampling fraction is very large. Thus, completing an RLA of a countywide contest may take as much work as completing an RLA of a statewide contest with the same margin – except that the work is done by just one county, not shared by 67. This fact is crucial when deciding how to audit contests that are not statewide.

**Choosing between risk limiting audit methods**

Most RLA pilots in Pennsylvania have used ballot polling. Ballot polling generally is best suited to RLAs of local contests with relatively few batches, where batch comparison might require auditing most or all of the voted ballots. (The August 2019 batch-comparison sandbox exercise in Philadelphia focused on hand-counting methods; it did not apply the statistical methods associated with RLAs). Also, ballot-polling RLAs do not require *vote* counts for each batch, only *ballot* counts. As we further describe below, timely obtaining batch-by-batch vote counts for both polling place and centrally counted mail ballots, using the various voting systems certified in Pennsylvania, poses logistical complications that election officials have needed time to work through. Ballot polling thus was far more feasible than batch comparison immediately after the November 2020 election.

However, workgroup members agree that batch comparison has important operational advantages in statewide contests. As we showed above, batch-comparison sample sizes increase more slowly than ballot-polling sample sizes for small margins. For the reported statewide margin of about 1.2% in the 2020 presidential election, an RLA of such a contest with a 10% risk limit would require sampling and hand-counting around 200 batches statewide, depending on operational details. Thus, the vast majority of batches need not be handled or audited. Even a considerably smaller margin would still yield a manageable sample size. The high likelihood of completing the RLA in one round also is advantageous.

Another advantage of batch comparison is the specificity of the results: the ability to determine that a risk limiting audit count matches a reported count exactly, or alternatively, to identify and isolate discrepancies. Batch-comparison RLAs usually find some combination of perfect matches and small discrepancies. For small discrepancies, it often is possible to identify one or two ballots that account for the differences and to explain why. If larger discrepancies are found, perhaps due to scanner misconfiguration or other problems that can affect multiple ballots, election officials can work to trace the problems to their source. Contrariwise, ballot-polling RLAs cannot distinguish small *counting* discrepancies from no discrepancies at all. For instance, in a ballot-polling RLA of a contest with a 5-point margin, the margin in a ballot-polling sample might be 4 points, or 6 points not because of any discrepancies between how the machines counted ballots and how the auditors did, but through random chance alone.

Additionally, a batch-comparison RLA can dovetail with the existing statutory requirement for a "statistical recount" of 2% of ballots (or 2,000 ballots, whichever is less) in each county. Although the workgroup strongly recommends replacing the current statutory provision with a new requirement built around robust methods such as RLAs, RLA implementation need not wait upon legislative action. In broad outline: counties can supply the necessary batch-level data to PA DOS; PA DOS then can select the random sample (as described below) and inform counties

which batches must be hand-counted in the contests specified for the RLA. If the target contests are not unusually close, some counties may then have to select *additional* batches to complete the statistical recount. (Of course, counties must comply with all requirements for the statistical recount, such as checking contests that are not target contests in the RLA). Contrariwise, it is *possible* that some counties will need to audit more batches to comply with the RLA requirement than to complete the statistical recount. If the RLA requires additional batches, only the specified contests must be hand-counted.

Based on these advantages and on experience from the pilots, the workgroup strongly recommends replacing the existing 2% statistical sample with RLAs of randomly selected statewide and county-level contests that confirm the overall accuracy of election results (see further discussion below). As an alternative to replacing the 2% statistical sample, the General Assembly could adopt a hybrid approach that would utilize an RLA of selected statewide contests and provide for batch-comparison audits of a maximum number of ballots for county and local contests. Without a statutory change, the statewide RLA will supplement the 2% statistical sample.

**Contests to be audited**

When considering which contests to audit, it matters that "auditing a contest" has more than one meaning. RLAs (and similar statistical audits) specify one or more **target contests** and seek strong evidence for the outcomes of those contests. (As explained earlier, the risk limit determines how strong the evidence must be). It is common to audit additional contests with respect to the same batches or ballots that were randomly selected for the RLA. This approach is called **opportunistic auditing** because it takes advantage of the fact that audit boards already are examining these ballots; the additional contests audited can be called **opportunistic contests.** No risk limit is set for opportunistic contests, but the audit still can provide important evidence about the accuracy of the vote count.

Target and opportunistic contests should be selected with all goals of the RLA in mind. RLAs must be broad enough to support the overall accuracy of vote counting, while being small enough for election officials to conduct them carefully. Enormous, complicated hand counts are demanding and error prone. Tally sheets become increasingly hard to use when more than three or four contests are included.

> The workgroup recommends selecting some statewide contests (which could be retention votes or referenda) as *target* contests for the RLA, and conducting *opportunistic* audits of smaller contests as they occur in the audit sample. Specifically, we recommend:
>
> - Offices to be voted on by the electors of the state at large that are contested (have more candidates or choices than vote opportunities) and are not subject to an automatic recount under 25 P.S. § 3154(g)(1) should be eligible to be target contests.
> - The top-of-ballot eligible contest to be voted on by the electors of the state at large should always be a target contest.
> - An additional target contest should be selected at random from among all other eligible contests to be voted on by the electors of the state at large.
> - In no case would more than two contests to be voted on by the electors of the state at large be selected.

- In no case would more than one local contest be selected.
- At least one local contest in each county should be audited opportunistically, to be selected according to guidelines directed by PA DOS.
- To ensure that all counties are included, the audit sample should include at least one randomly selected batch from each county.
- In the event of a statewide recount, a precertification RLA may not always be feasible given statutory obligations.

Contests that qualify for an automatic recount should not be eligible to be target contests because these contests will be the subject of a full recount. However, we recommend that they be audited in counties that do not conduct manual (hand-count) recounts.

In the future it may be feasible to audit additional target or opportunistic contests.

## Operational considerations

RLAs require that election officials follow a specific set of procedures to ensure that risk limiting audits are conducted efficiently, accurately, and transparently. Here we sketch the major steps involved in preparing for and conducting an RLA. For simplicity, we primarily consider batch-comparison risk limiting audits. This discussion informs some of our recommendations below.

*Ballot accounting, storage, and pre-risk limiting audit reporting*

RLAs require specific ballot accounting information to be available at the time the random sample is selected. All RLA methods require election officials to prepare a **ballot manifest**, which lists the various batches in which ballots are stored and how many ballots are stored in each batch. The ballot manifest cannot rely solely on voting system data; the ballot counts for each batch must be attested by independent records, such as pollbook records for regular election day ballots and manual scanner logs for central-count (absentee, mail, and provisional) ballots. (Workgroup participants have commented that preparing a "skeleton" or outline of the ballot manifest before the election saves time later on). For batch-comparison risk limiting audits, election officials must also export **batch-level vote counts**, by candidate or choice, from the voting system. County officials upload their ballot manifests and batch-level vote counts to the RLA software. All counties must submit their data files before the random sample can be selected and the risk limiting audit itself can begin.

These requirements entail some careful planning. County election officials must decide in advance their strategy for dividing ballots into batches (**"batching"** the ballots), obtaining ballot counts and vote counts for each batch, and storing the batches for easy retrieval. If some polling places have more than one scanner, will election officials use one batch per *scanner* or one batch per *precinct*? Will election officials sort central-count ballots by precinct or will some central-count batches contain ballots from several or many different precincts? If officials sort central-count ballots, will they combine these ballots with election-day ballots to create one large batch per precinct, or will they report election-day and central-count batches separately? The answers will determine how they account for and store their ballots, as well as how they obtain vote counts for each batch.

Counties should not all be required to use the same batching method. For instance, sorting central-count ballots by precinct may be easy in some counties and very difficult in others. As

we further discuss below, voting systems may vary in their support for various approaches. The Department of State can provide valuable guidance to counties on best practices for obtaining the batch data and checking it against other election data.

With regard to batching strategy, it is important to know that, in general, large batches are more likely to be selected than small batches. It is desirable to avoid very large batches with many thousands of ballots, which are challenging to hand-count in risk limiting audits.[6] (The sheer *number* of batches is not very important. Dividing a 5,000-ballot batch into two smaller batches may be worthwhile; splitting a 20-ballot batch is not).

Ballot-storage practices are key to smooth risk limiting audits. County election directors have commented on how they have modified and improved their practices based on their pilot experiences. For instance, some jurisdictions have invested in storage racks to avoid tall stacks of ballot containers where the bottommost container is hard to access. Some have improved their container labeling and organization to make it easy to find specific batches. To manage batch size, some counties have adopted new procedures to ensure that ballots counted on different scanners are kept separate and can be treated as separate batches. Some have set a uniform, modest size for central-count batches, and they preprint labels with unique batch names for each batch.

*Voting system support for data*

Pennsylvania presently employs voting equipment from five different manufacturers.[7] Batch-comparison RLAs require batch-level vote counts to be exported from the voting system. The May 2021 batch pilot identified challenges in obtaining the desired data from some voting systems. Some voting systems require combining multiple reports to obtain batch-level vote counts. Others provide the data directly, but in a pdf format, which requires conversion to be compatible with the risk limiting audit software. County election officials must be aware of their voting system's limitations and procedures. For instance, counties realized during a pilot that one voting system "locks down" certain reporting functions while write-in ballots are being adjudicated, potentially delaying access to batch-level vote totals. It is important that the Department of State work with voting equipment vendors to ensure that it is possible to obtain the required data as easily and efficiently as possible and to provide county election directors with clear instructions on how to do so. Any method for exporting data from the voting system and importing it into the risk limiting audit software (currently Arlo) must maintain the security and integrity of both the data and the voting system.

*Generating the random sample*

Random selection is critical to the validity and credibility of post-election RLAs. When election officials use their own discretion to select batches, or when they apply a *non-random* rule (such as "select the smallest batch or combination of batches that contains at least 2% of all ballots cast"), there can be no assurance that the resulting sample is representative.

The risk limiting audit software used for RLAs in Pennsylvania uses the most common and recommended approach for generating random risk limiting audit samples. Rather than

---

[6] Even in ballot-polling audits, very large batches can be hard to work with.

[7] Election Systems & Software, Dominion Voting Systems, Clear Ballot, Hart InterCivic, and Unisyn Voting Solutions.

separately select each ballot or batch through some random procedure – akin to generating hundreds or thousands of winning lottery numbers – officials generate just one random "seed" in a public event, often called the seed ceremony. Typically, participants take turns selecting one of 20 ten-sided dice and rolling each one to generate a 20-digit random number, which is the seed. Officials then enter the seed into the risk limiting audit software; the batch data files have already been uploaded into the risk limiting audit software. Arlo uses an open-source *pseudo-random number generator* (PRNG) to convert the seed into a series of random numbers, each of which corresponds to a batch (or ballot, depending on the method).

Once the random seed is entered into the risk limiting audit software and the RLA is "launched," each county can immediately see what batches or ballots it has to review as part of the risk limiting audit.

*Conducting the risk limiting audit*

Once the random sample for the RLA is drawn, election officials retrieve and audit the selected batches.[8] Often the risk limiting audit begins the day after the random sample selection, accommodating possible last-minute delays, and in some cases allowing time for the randomly selected batches to be transferred from secure storage before the risk limiting audit begins.

County officials often put together multiple audit teams to help them with retrieving and examining audited ballots since this is the most time-intensive part of the risk limiting audit process. Officials can enter their RLA results into the audit software as they finish each batch or when all batches have been counted.

To ensure accurate hand counts in batch-risk limiting audits – especially of large batches – auditors should follow well-defined procedures. Some points to consider:

*Manual tally sheets:* Each vote should be recorded individually on a paper tally sheet. Often, two talliers keep independent tally sheets, which they compare and reconcile periodically to guard against errors.

*Staffing:* Auditors should work in teams of *at least* two people, preferably three or four, to ensure that at least two people examine each ballot, and that at least two people tally or observe the tallying. Some common models:

- Two auditors, each with a tally sheet, separately review and tally each ballot, calling out their interpretations as they go.
- Three auditors: a caller, a tallier, and an observer who sits between them and confirms the interpretations and the tallies.
- Four auditors: two callers and two talliers.

In addition, staffers should be designated to handle ballot containers, track chain of custody, and organize the ballots into smaller sub-batches (see below); to carry ballots to and from the

---

[8] As mentioned above, in ballot-polling audits, retrieving the individual ballots selected for audit from across the many containers of ballots generally takes most of the time of the audit. In batch-comparison audits, generally most time is spent hand-counting the ballots in the selected batches.

risk limiting audit teams; and to enter tally totals into the risk limiting audit software. Data entry should be done by at least a team of two to help catch any errors.

The total staffing required for the RLA may depend on how many ballots must be counted in the time available. Although the exact ballot counts cannot be known until after the public seed ceremony, PA DOS can and should provide estimates to counties based on the risk limit, the target contest margins, and other data. The *speed* of hand counts also depends on the number and complexity of contests to be tallied and on how efficiently ballots can move through the process (which can be affected by available space). Commonly, an audit team can tally three or four contests for about 200-250 ballots per hour, allowing time for checks and short breaks. It is prudent to allow for delays.

*Voter intent standards and dispute resolution:* Auditors should be given directions on determining voter intent and valid votes. A clear mechanism should be in place for resolving any disagreements between auditors over the interpretation of a particular ballot. Minor discrepancies between machine counts and auditor interpretation of voter intent are normal and should not prolong risk limiting audits.

*Sub-batching:* Auditors should subdivide batches of ballots into smaller sub-batches of a known size for counting, rather than working with an entire large batch at a time. Tallying 50 or 100 ballots at a time reduces errors by breaking up the work and providing natural opportunities for cross-checks.

*Risk limiting audit calendar*

To ensure that precertification RLAs can be conducted successfully, election directors must first have sufficient time to complete other election processes. A statewide RLA cannot launch until all counties have completed tabulating ballots and have submitted certain data. The state also plays a role in ensuring that the risk limiting audit data submitted by counties are correct. Any delays in processing and tabulating ballots will put pressure on the RLA timeline. (For instance, as noted before, in one pilot, an unexpected number of write-ins delayed some counties' submission of risk limiting audit data while they completed adjudicating the write-in votes). As the number of ballots submitted by mail in the state has increased, providing election officials time to pre-canvass ballots prior to election day would allow unofficial vote totals to be available sooner, giving election officials more time to audit prior to certification. Because of possible delays, the workgroup does not recommend establishing a full RLA calendar in a directive.

The workload of the RLA itself should be very manageable even for unusually close statewide target contests. At the same time, because of the need for coordination across all 67 counties, it is important to allow for exceptions to deadlines. DOS should be given explicit statutory authority to extend the county certification deadline if more time is needed to complete the RLA.

*Division of responsibilities*

Successful risk limiting audits depend on close collaboration between the Department of State and county officials, but also on clarity about basic roles and responsibilities. We see the basic division as follows.

- The Department of State should organize and direct RLAs of statewide contests, in partnership with counties, who conduct the risk limiting audit. This coordinating role provides coherence and consistency. The Department of State should:
  - o Prescribe the method and best practices to be followed by counties conducting the risk limiting audit.
  - o Contract for the statewide risk limiting audit software and provide administration and support.
  - o Conduct the random seed ceremony.
  - o Develop plain language guidance and procedures needed to prepare for and conduct post-election risk limiting audits.
  - o Prescribe any necessary forms, including a standard tally sheet template
- Each county should *conduct* its portion of the statewide RLA and should *organize and direct* its opportunistic audits of county contests. These responsibilities include:
  - o Creating, checking, and timely uploading batch data files.
  - o Randomly selecting an eligible contest for opportunistic auditing at a public event.
  - o Recruiting, training, and supervising people to conduct the RLA in accordance with directives and guidance.
  - o Timely providing all risk limiting audit results, and any additional documentation needed for opportunistic audits.

These responsibilities entail some material costs on both sides. *The Department of State* will incur initial costs to formulate necessary directives, guidance, procedures, and forms and possibly to develop tools for data reporting. It will also incur ongoing costs to coordinate statewide RLAs, assist counties, and manage data and result reporting. *Counties* will incur initial planning and training costs and may also invest in equipment to enhance ballot management and storage (such as ballot boxes and shelving). Counties also will incur ongoing staff and supply costs to prepare for and conduct RLAs. In many cases the ongoing costs may be similar to or smaller than (or largely incorporated within) the existing statistical recount. To ease possible burdens on counties in the event that a risk limiting audit requires more work than expected, it may be desirable to follow Wisconsin's example and establish a statutory formula and appropriate funding for partial reimbursement audit costs, based on the number of ballots to be examined in each county.

## Recommendations

After careful consideration of the different audit methods and the mechanics of audit implementation, the workgroup puts forward a series of recommendations directed to the Legislature and the Department of State. The fundamental requirements for robust post-election precertification audits should be defined in statute. However, many of the procedural details are best left to directives and guidance issued by the Department of State to allow the flexibility to adapt to new advances in the field or to adjust procedures as needed. The workgroup has also identified where it should continue its work.

*Legislative Recommendations*

The workgroup recommends that the Department of State and the County Commissioners Association of Pennsylvania (CCAP) engage the Legislature to pass legislation to accomplish the following goals:

● Repeal Section 1117-A of the Election Code, 25 P.S. § 3031.17, requiring a statistical recount of a random sample of ballots in each county, including at least two percent of the votes cast or 2,000 votes, whichever is the lesser.
● Replace this Section with a new law requiring all counties and the state to implement enhanced post-election audits, conducted after every primary and November election and before the results are certified, that utilize a Department of State-approved audit method that conforms to a pre-specified standard of statistical confidence that the election outcome is correct.
  o All Department-approved audit methods should be consistent with best practices in the field and may include but are not limited to risk limiting audits.

We further recommend that the Legislature allow election officials to pre-canvass mail and absentee ballots prior to election day. This is the norm in the majority of states, and it is necessary to ensure timely reporting of both unofficial election returns and certified election returns. Ideally, election officials would be able to not only verify the information on the outside ballot envelopes prior to election day, but also to extract the ballots, prepare the ballots for scanning, and scan the ballots without releasing any vote counts.

The Legislature may decide to create additional provisions, such as specifying certain races to be audited. However, we recommend leaving the Department of State with broad discretion to implement, in consultation with local election officials, appropriate rules and procedures for robust, efficient and effective precertification audits. In some other states, excessively specific statutory requirements have hobbled audit implementation.


*Recommendations to the Department of State*

The workgroup recommends that the Department of State analyze and work toward implementation of the following goals, to the extent consistent with its statutory authority:
● The Department of State should formulate a directive requiring that beginning in November 2022, in every primary and November election where there is at least one statewide contest, the state should conduct the following post-election precertification audit(s) to supplement the 2% statistical sample:
  o The race that is determined to be the top of the ticket; and
  o If there is more than one statewide race, one additional randomly selected race.
● For audits of statewide elections, the directive should require that counties utilize the batch-level comparison audit method, which audits all the ballots from randomly selected batches of ballots. If the current two percent statutory audit requirement is not repealed, the batch-level comparison audit can be used concurrently with the current two percent requirement.
● The directive should further require that beginning in November 2022, in every primary and November election, the counties should conduct the following post-election precertification audits(s):
  o At least one county-level race randomly selected by the county board of elections to be audited in parallel with the statewide audited race(s).
  o In the event there are no countywide races on the ballot, the Department of State should formulate standards for random selection of municipal-level races that may be audited, if applicable.

- The directive further should require basic procedural elements that assure the validity and transparency of the audit, as described in the report, including:
  - o Ballot manifest data required to be submitted by counties and published by the Department of State before the random selection of races.
  - o In addition to the uniform Standards for What Constitutes a Vote, any additional procedural instructions for conducting the audit hand counts and reporting the results.
- The Department should prepare additional guidance and forms as appropriate to support counties in conducting the audit.


*Workgroup continuity*
Since it first convened in 2019, the workgroup has studied and tested the implementation of robust audits in Pennsylvania, amassing considerable experience with audit practices in the process. The workgroup recommends that it continue to meet since it still has a critical role in ensuring the successful implementation of audits in the state. Election directors and representatives of the Department of State provide the workgroup with invaluable expertise in election practices in the state. Outside experts, with knowledge of elections and election audits, contribute technical advice on audits as well as a broader understanding of election best practices.

The workgroup could assess audit procedures on an ongoing basis, making recommendations for improvement. As RLAs are conducted across multiple election cycles, the workgroup could evaluate whether the risk limiting audit timeline requires any adjustments with respect to other election deadlines. To keep Pennsylvania at the forefront of developments in the field, the workgroup could also study whether new audit methods would be practical or useful in Pennsylvania, organizing pilots as part of the process. Following the implementation of robust risk limiting audits in the state, the workgroup proposes to issue a report in late 2022 or early 2023 with further recommendations.