Risk-Limiting Post-Election Audits:
Why and How

Risk-Limiting Audits Working Group

October 2012 (version 1.1)

Contributing editors (affiliations for identification only):

Jennie Bretschneider, Office of the California Secretary of State
Sean Flaherty, Iowans for Voting Integrity
Susannah Goodman, Common Cause
Mark Halvorson, Citizens for Election Integrity Minnesota
Roger Johnston, Argonne National Laboratory
Mark Lindeman (executive editor), Columbia University
Ronald L. Rivest, Massachusetts Institute of Technology
Pam Smith, Verified Voting
Philip B. Stark, University of California, Berkeley

The **Risk-Limiting Audits Working Group** is a self-organized group of election officials, researchers, and advocates, created to explore risk-limiting audits as an election verification mechanism. The working group considers proposals for conducting risk-limiting audits and implementing appropriate laws and procedures.

Besides the contributing editors of this white paper, participants in working group discussions have included Mary Batcher, Justin Berardino, Robert Brownstone, Lynn Garland, Joseph Lorenzo Hall, Candice Hoke, Steve Huefner, Neal Kelley, Douglas A. Kellner, Jerome Lovato, Larry Norden, Steve Pierson, Barbara Simons, and Dan Wallach.

Table of Contents

Executive Summary

How can we determine whether vote-counting systems have performed well in specific elections? We need to check the evidence. Computerized systems may produce incorrect results due to programming errors or deliberate subversion. Even hand counts may be erroneous. Risk-limiting audits systematically check the election outcomes reported by vote-counting systems. Specifically, a risk-limiting audit checks some voted ballots or voter-verifiable records in search of strong evidence that the reported election outcome was correct – if it was. Specifically, if the reported outcome (usually the set of winner(s)) is *incorrect*, then a risk-limiting audit has a large, pre-specified minimum chance of leading to a full hand count that reveals the correct outcome. A risk-limiting audit can stop as soon as it finds strong evidence that the reported outcome was correct. (Closer elections generally entail checking more ballots.)

Risk-limiting audits are highly adaptable. If a voting system can be audited at all, a risk-limiting audit is feasible. Risk-limiting audits can check simple plurality contests, multi-winner contests, measures requiring a super-majority, and so on. Risk-limiting audits can be conducted for any number of contests in a single election, often using a single sample for greater efficiency. Risk-limiting audits can begin very soon after the election – perhaps even before all the votes have been counted – or later in the process, as long as time remains to correct any incorrect outcomes.

Any jurisdiction can conduct a risk-limiting audit, but the following conditions must be satisfied:

1. **Audit trail**: The audit requires a durable record that accurately records each voter's selections – a physical ballot or some alternative. If the voter has not verified the record, then we cannot know whether it accurately records the voter's selections.

2. **Preservation and security (chain of custody)**: To be assured of yielding correct results, the audit trail should be preserved inviolate. Procedures are needed to ensure, and to demonstrate, that the audit trail has been preserved, or at worst that any changes did not alter the outcome.

3. **Commitment to vote subtotals, or an alternative**: If the audit uses the voting system's reported subtotals (such as precinct vote totals for each candidate), then the subtotals used in the audit must verifiably be the same subtotals reported by the voting system, and must verifiably add up to the vote totals.

4. **Clear standards for determining whether a vote is valid and interpreting voter intent**: An audit can only yield reliable results if it applies explicit, previously established standards for what should count as a valid vote.

Generally, risk-limiting audit requirements can be enacted in relatively simple legislation (or the equivalent) that does the following:

- Defines a risk-limiting audit
- Says which election contests should be subject to risk-limiting audits, or how these contests should be chosen

- Determines the risk limit (the maximum chance that auditing a contest with an incorrect outcome will *not* lead to a full hand count)
- Sets the time frame for completing risk-limiting audits
- Establishes or requires procedures for checking the integrity of the audit trail, randomly selecting the audit sample, facilitating public observation, reporting audit results, and other aspects of the audit

(Post-election audit provisions should be harmonized with existing provisions for recounts and other ways of contesting or correcting election results. For instance, it may be appropriate to adjust the time frame for recounts, and/or to require full hand recounts in very close contests in lieu of auditing those contests.)

Most audit implementation details can be established in regulations and/or written procedures, subject to public comment and made available to the public. Omitting implementation details from the legislation can facilitate improvements in risk-limiting audit procedures based on experience, new equipment, and other changed circumstances.

Risk-limiting audits come in several varieties, with various benefits and requirements. Understanding the choices can help in deciding which contests to audit.

- The most efficient audits are *ballot-level comparison audits*, in which the voting system interpretation of individual ballots – typically recorded in Cast Vote Records – is compared with the audit interpretation of the same ballots. But many voting systems either do not record Cast Vote Records or provide no practical way to match them with the corresponding ballots. (Also, the use of Cast Vote Records can raise voter privacy concerns.) Sometimes it is feasible to perform a ballot-level comparison audit of a secondary system – a *transitive audit*.

- *Comparison audits at the batch level* – for instance, auditing the voting system's subtotals for certain precincts – sometimes are more feasible than ballot-level comparison audits, but they are less efficient. In smaller contests (for instance, those with under 100 batches), batch-level audits may require auditing large fractions of the ballots even when contests are not very close.

- *Ballot-polling audits* based on a random sample of ballots – without reference to the voting system interpretation of those ballots – may be more efficient than batch-level comparison audits, especially for contests that are not very close.

If ballot-level comparison audits are feasible, then it may be reasonable to conduct risk-limiting audits in all contests, or in many of them. If current voting systems do not support ballot-level comparison audits, a jurisdiction may choose to audit relatively few contests in the short term, until it can obtain more auditable systems.

Introducing Risk-Limiting Audits

Why audit?

State and federal voting system testing and certification help ensure that voting systems used in the U.S. can count ballots accurately and securely, while protecting voter privacy. But this kind of front-end regulation isn't enough. It is good to have evidence that voting systems can perform well in laboratory conditions, but how do we know how well they have performed in a particular election?

History tells us that election fraud can happen. Computer experts have shown that voting systems can be hacked.[1] But even if we set aside fraud and tampering, we know that no voting system can operate to perfection. Neither can humans. Machines misinterpret ballots; humans mismark ballots. Errors happen, as in a 2006 Iowa county primary election where many absentee votes were inadvertently awarded to the wrong candidates because of faulty ballot programming; a hand recount altered the outcome in two contests.[2] We cannot know how widespread those errors are – and whether they have actually altered election outcomes – unless we look. In March 2012, a routine audit in Palm Beach County, Florida revealed that two city council contests had been certified with the wrong outcomes![3]

Post-election audits[4] look for evidence that the original vote counts were substantially accurate. When (and as far as) an audit confirms the accuracy of the original count, it provides a basis for public confidence that the voting systems performed as they should. If the audit finds errors, then the errors can be corrected, and steps can be taken to prevent or reduce similar errors in the future.

How accurate is a "substantially" accurate vote count? The answer depends on the context, but we all want vote counts to be *at least* accurate enough to correctly determine the outcome: who won, or that a runoff is needed. Yet most post-election audits are not designed with that goal in mind! Nineteen states have some sort of post-election audit requirement, but most of those procedures are somewhat arbitrary. They sometimes require far more work than needed to

---

[1] On security issues pertaining to voting systems, two useful sources of information are California's Top-to-Bottom Review (TTBR) and Ohio's Project EVEREST. These studies, both completed in 2007, found major security flaws in every system submitted for inspection. Ohio's secretary of state concluded that "no system used in Ohio is without significant and serious risks to voting integrity" (Project EVEREST Report of Findings, available at http://www.sos.state.oh.us/sos/upload/everest/00-SecretarysEVERESTExecutiveReport.pdf, p. 76). Extensive documents from the TTBR are available at http://www.sos.ca.gov/voting-systems/oversight/top-to-bottom-review.htm.

[2] The incident is described in Sean Flaherty, "In an Age of Computerized Voting, Is It Possible to Maintain Voting Integrity?", available at www.votetrustusa.org/index.php?option=com_content&task=view&id=1460&Itemid=113.

[3] Jennifer Sorentrue, Dara Kam, and George Bennett, "Recount shows wrong winners declared in two Wellington election races," *Palm Beach Post News*, March 19, 2012, updated March 20, 2012, at http://www.palmbeachpost.com/news/recount-shows-wrong-winners-declared-in-two-wellington-2247602.html.

[4] Here we use "post-election audits" to refer to vote tabulation audits, which check whether votes were interpreted and tallied (totaled) correctly. Obviously other things should be checked after an election, as part of a broader post-election audit and/or other post-election procedures. We discuss some of those crucial checks in this paper.

confirm the election outcome; at other times, they provide far too little information. Risk-limiting audits are designed to solve the problem of verifying election outcomes.

What is a "risk-limiting audit"?

A risk-limiting audit checks some voted ballots[5] in search of strong evidence that the reported election outcome was correct – if it was. If the reported outcome is incorrect, then the audit usually will lead to a full hand count that reveals the correct outcome. By design, once the audit finds strong evidence that the reported outcome was correct, it can stop. Thus, the audit intelligently adapts to the facts of a particular election. (Closer elections generally entail checking more ballots.)

More formally, a risk-limiting audit is a procedure for manually checking a sample of ballots (or voter-verifiable records) that is guaranteed to have a large, pre-specified chance of correcting the reported outcome if the reported outcome is wrong. (An outcome is wrong if it disagrees with the outcome that a full hand count would show.)[6] The largest chance that an outcome will not be corrected by the audit – given that it is incorrect – is the risk limit in that audit. For instance, if the risk limit is 10%, then if the outcome is wrong, there is *at least* a 90% chance that the audit will lead to a full hand count that corrects it. The risk limit calculation is based on worst-case assumptions; the actual chance of correcting a wrong outcome may be much larger than 100% minus the risk limit, depending on how and why the outcome is wrong.

Risk-limiting audits generally proceed by selecting an initial sample of ballots and interpreting them by hand, then determining whether the audit must expand. The number of ballots in the initial sample depends on various things, including the margin of victory in the contest: the narrower the margin, the larger the initial sample. In the most common kind of risk-limiting audit, the audit stops if auditors find very few differences between the voting system interpretation[7] and the audit interpretation of the votes on the sampled ballots.

Risk-limiting audits can lead to a full hand count. If the initial sample does not yield sufficiently strong evidence that the reported outcome is correct, then the audit escalates: additional ballots are selected and interpreted by hand. Escalation continues until there is sufficient evidence that the outcome is correct—or, if the evidence never becomes strong enough, until all votes have been hand counted. If there is a full hand count, it reveals whether the reported outcome was correct. The hand-count outcome replaces the original outcome if the original was incorrect.

Benefits of risk-limiting audits

Risk-limiting audits can confirm correct outcomes very efficiently, depending on the method used. The most efficient audits use individual ballots, comparing the audit interpretation of each ballot in the sample to the voting system interpretation of the ballot. This approach is much more

[5] Alternatively, but less desirably, the audit can use other voter-verifiable records, as discussed below.
[6] This definition assumes that the audit trail is intact – that no ballots have been added, removed, or altered – or at least sufficiently intact that the outcome according to the audit trail is still correct.
[7] The voting system is the entire process by which votes are counted, so the voting system interpretation may combine electronic and hand counts.

efficient than sampling precincts and counting all the ballots in each sampled precinct, a common approach. However, it is not always possible or practical to retrieve the voting system interpretation of each ballot. Fortunately, risk-limiting methods have been developed to accommodate a wide variety of voting systems. If a system can be audited at all, a risk-limiting audit is feasible.

Risk-limiting audits are highly adaptable – and not only in the variety of voting systems that can be audited. Risk-limiting audits can check simple plurality contests, multi-winner contests, measures requiring a super-majority, and so on. Risk-limiting audits can be conducted for any number of contests in a single election, often using a single sample for greater efficiency.[8] Risk-limiting audits can begin very soon after the election – in some cases even before all the votes have been counted – or later in the process, as long as time remains to correct any incorrect outcomes.

Beyond confirming outcomes, risk-limiting audits can yield extensive information about voting system accuracy. Because risk-limiting audits are designed to check election outcomes, some people mistakenly think that they can *only* check election outcomes. On the contrary, risk-limiting audits offer outstanding opportunities to rigorously and efficiently investigate voting system performance, fostering high quality and continual improvement.

Why not simply hand count all ballots in the first place?

Some jurisdictions do hand count all ballots, but that is no reason not to check their counts. Any method of counting votes – including hand counting – can produce wrong results. Routine risk-limiting audits provide incentives for accuracy in the original counts, and the audit results may be useful to improve counting procedures for hand counts as well as other methods.

In many jurisdictions, hand counting is time-consuming, labor-intensive, and prone to human error, especially as fatigue sets in. Machines are vulnerable to errors in their own right, but they never become distracted or bored. Combining machine counts with efficient audits provides two separate checks on the results, and may provide greater accuracy than either machine counts or hand counts alone.

What is the starting point for an audit?

Any jurisdiction can conduct a post-election risk-limiting audit, but the following conditions must be satisfied first:

1. **Audit trail:** The audit requires a durable record that accurately records each voter's selections. This record may be a physical ballot, or possibly a VVPR ("voter-verifiable paper record") from a direct recording electronic (DRE) voting machine.[9] If the voter has not verified the record, then we cannot know whether it accurately records the voter's

[8] As we discuss later, the most efficient approach to auditing multiple contests depends on circumstances.
[9] In principle, alternatives to a *paper* trail are possible, including voter-verified audio audit trails (VVAAT) and video audit trails (VVVAT). Whatever the merits of these alternatives, they are not widely available. See also the discussion of audit trails below.

selections. Paper ballots marked by voters are preferable to VVPRs, which voters may or may not review. (For simplicity, we will generally refer to the auditable records as "ballots.")

2. **Preservation and security (chain of custody):** For an audit to yield correct results, the ballots used in the audit should be the same ballots cast in the election – none added, none subtracted, none altered. Thus, procedures are needed to ensure, and to demonstrate, that the ballots have been preserved inviolate, or at worst that any changes did not alter the election outcome(s). These procedures may include ballot accounting, seals and locks to control access to the ballots, chain-of-custody records, video surveillance, and so on.

3. **Commitment to vote subtotals, or an alternative:** Most post-election audits rely on the voting system's reported vote subtotals for particular sets of ballots. (A set might consist, e.g., of all ballots cast in a precinct – or it might be an individual ballot.) Thus, these subtotals must exist, and they must be committed to. That is, like the audit trail itself, the subtotals used in the audit must verifiably be the same subtotals reported by the voting system, and must verifiably add up to the vote totals. Commitment can pose thorny issues, especially when sampling individual ballots. We will discuss these issues, and some risk-limiting audit methods that do not depend on a commitment to vote subtotals.

4. **Clear standards for determining whether a vote is valid and interpreting voter intent:** An audit can only yield "correct" results if it applies explicit, previously established standards for what counts as a valid vote.

What about other kinds of post-election audits?

Most current audit requirements[10] call for fixed-percentage, fixed-size, or tiered samples. For instance, New York has a fixed-percentage audit requirement: a random sample of 3% of voting machines or systems used in the election is drawn, and all votes on all ballots cast on those systems are audited. Wisconsin has a fixed-size requirement: In major general elections, 50 municipalities are randomly selected to audit some statewide contests. Oregon has a tiered requirement: The required sample depends on the reported margin of victory, ranging from 3% of precincts (or batches) in most cases to 10% of precincts if the reported margin is less than 1%. (Oregon mandates a full recount in certain cases if the reported margin is less than 0.2%.)

All these audits can provide useful information, but none is risk-limiting, and the difference matters. Some existing audit provisions offer no means whatsoever by which incorrect outcomes can be corrected. All these provisions mandate samples that are sometimes far too small to provide strong evidence about the outcome – or, conversely, needlessly large in many circumstances. Also, they have inadequate rules for deciding when an audit must continue – again, sometimes too permissive to provide strong evidence about the winner, and sometimes needlessly restrictive. Existing audit provisions should not be indiscriminately scrapped, but

[10] Comprehensive information about state audit requirements is available via Citizens for Election Integrity Minnesota's (CEIMN) State Audit Laws Searchable Database, http://www.ceimn.org/searchable_databases/state_audit_laws . CEIMN also maintains a State Recount Laws Searchable Database, at http://www.ceimn.org/ceimn-state-recount-laws-searchable-database/search.

replacing or integrating them with risk-limiting audits can provide better results, and actually reduce the burden in some cases.

Risk-limiting audits also can be combined with *discretionary partial recounts* or other partial hand counts. For instance, Minnesota (which does not conduct risk-limiting audits) allows candidates in certain contests to select up to three precincts to be hand-counted at their own expense. Other states give candidates wide discretion to request hand counts at their own expense. Many states formally or informally permit election officials to conduct partial hand counts to check voting system results. Such mechanisms provide ways of investigating anomalous results that might evade the audit, especially if the anomalies do not bring the outcome into question.

Implementing risk-limiting audits

Risk-limiting audits, like other policy requirements, can be implemented through any combination of legislation, formal rule-making, explicit operating procedures, and actual practice. For instance, pilot risk-limiting audits in California were framed by a state law (AB2023, 2010) that defined risk-limiting audits and set some broad parameters for conducting pilots; the details were left to the county election officials and consultants who actually conducted the pilots.[11]

In general, we recommend legislation that carefully defines the essential elements and prerequisites of a risk-limiting audit system while leaving most implementation details to rule-making and local jurisdictions. Given the diversity of election systems and political cultures across the United States, we cannot give comprehensive recommendations for what should happen at each level in every state. Accordingly, we address a range of issues here that should be of interest – to varying degrees – to state legislators and officials, local election officials, and other policy stakeholders who want to implement risk-limiting audits in their jurisdictions.

[11] For information on pilot audits conducted under AB 2023 through early 2012, see the March 1, 2012 Report to the Legislature at http://www.sos.ca.gov/admin/reports/2012/post-election-audit-report-20120301.pdf. Earlier pilots are discussed in Joseph Lorenzo Hall et al., "Implementing Risk-Limiting Post-Election Audits in California," paper presented at the 2009 EVT/WOTE annual meeting, available at http://static.usenix.org/events/evtwote09/tech/full_papers/hall.pdf; and Philip B. Stark, "Efficient Post-Election Audits of Multiple Contests: 2009 California tests," paper presented at the CELS 2009 4th Annual Conference on Empirical Legal Studies, available via http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1443314.

Risk-Limiting Audits: Scope and Methods

A basic question when designing and implementing risk-limiting audits is: Which election contests should be audited in this way? Ideally, every contest in every election should undergo a risk-limiting audit. But a jurisdiction may choose to audit some contests and not others.[12] Time and cost are important considerations – and the time and cost of auditing depends in part on the specific audit method used. We now turn to some discussion of methods, then return to the question of which contests to audit.

"Batch" audits and ballot-level audits

Most post-election audits – risk-limiting and other – have been "batch" audits: A jurisdiction selects certain batches of ballots to audit, such as the ballots cast in particular precincts, or on particular voting machines, and counts all the votes within each batch. Often these batches correspond to subtotals that are already routinely published. For instance, many jurisdictions publish canvass reports of vote counts in each precinct, and then audit some of the precincts. Batches can be substantially smaller than entire precincts.

Unfortunately, batch audits are not particularly efficient when the batches are large. Auditing individual ballots is much more efficient. For instance, sampling and checking 500 ballots from all over a jurisdiction reveals more than counting 500 ballots from just one precinct. The reason is that the errors that the audit is intended to detect are by no means guaranteed to be evenly spread across the jurisdiction. If you were testing a bowl of cake batter to see whether you accidentally added a tablespoon of salt, you would not want to rely on one large taste from one corner of the bowl, because the salt might not be evenly mixed. You would rather try several smaller tastes from different parts of the bowl. Similarly, if miscounts are concentrated in relatively few batches for some reason – such as isolated programming errors, bad pens in certain precincts, or fraud designed to evade detection in an audit – examining all the ballots in just one batch, or just a few batches, easily could miss them. It is far better to sample ballots from all over the jurisdiction. Thus, although batch audits are more familiar, ballot-level audits generally are more effective. Ballot-level audits do pose practical challenges.

Comparison audits, transitive audits, and ballot-polling audits

Most post-election audits – risk-limiting and other – have been comparison audits: The voting system tabulation of a batch (or its "count" of an individual ballot) is compared with the audit's tabulation of the same ballots. If the results match and the audit count is accurate, then either the voting system count was error-free, or any errors offset each other. A comparison audit offers the best means of assessing the voting system's accuracy. A ballot-level comparison audit is ideal for that purpose: Not only is it most efficient, but it reveals the most information about which ballots are being miscounted in what ways.

If ballot-level comparison audits are the most efficient, why discuss any other method? Because ballot-level comparison audits are not always feasible. Ballot-level comparisons entail matching

[12] Even if some contests are not audited, ensuring that all contests are potentially subject to audit may be beneficial. See the discussion on page 18.

a physical ballot with the voting system interpretation of that ballot. Most voting systems today provide no way to do that. Some voting machines do not even record the interpretations of individual ballots, but simply tally their interpretations. Others record the interpretation of each ballot as a Cast Vote Record (CVR), but offer no practical way to trace a CVR back to the corresponding ballot; some even take measures to prevent such a trace. Also, some methods of tying CVRs to ballots raise privacy concerns. We discuss possible solutions to these problems later.

While batch comparison audits almost always are feasible, two alternative methods reap some advantages by using ballot-level samples. We briefly introduce these methods here.

The first is called a transitive audit. In a transitive audit, the ballots in an election are re-interpreted by a secondary system, in a way that permits the secondary system's interpretation of each ballot to be matched with the corresponding ballot.[13] (This may entail adding identification numbers to the ballots, or rescanning the ballots in relatively small batches such that a ballot can be reliably identified as, say, "the 43rd ballot in batch #12.") If the tabulation by the secondary system yields the same outcome the voting system did – even if the vote totals differ – then we can conduct a ballot-level comparison audit on the secondary results. If that audit provides strong evidence that the secondary outcome is right, then it also provides strong evidence that the voting system outcome is right.[14]

An obvious disadvantage of a transitive audit is the need to rescan all the ballots – but the advantage of sampling individual ballots, instead of entire batches, may justify this expense if a ballot-level comparison audit using the system of record is impossible. A transitive audit independently retabulates the results from the audit trail; if that secondary tabulation agrees with the original, it provides some reassurance that the original is correct. A transitive audit can yield other benefits as well, because the secondary scans can be used for purposes beyond confirming the outcome. For instance, the scanned images can be analyzed and compared with the voting system counts to locate discrepancies and investigate their causes. The scans even can be published online, as the Election Transparency Project in Humboldt County, California, did in several elections – although this practice raises voter privacy concerns.

The second approach is a ballot-polling audit, so named because it resembles a public opinion poll. In a ballot-polling audit, a random sample of ballots is selected and the votes are tabulated. If the sample provides sufficiently strong evidence for the outcome reported by the voting system – much as a public opinion poll might provide strong evidence that one candidate is ahead – then

[13] In principle, voter-verifiable paper records could be used in a transitive audit, although we are not aware of any audit that has done so.

[14] The audit uses the transitive property: If the voting system outcome is the same as the secondary outcome, and the secondary outcome is correct, then the voting system outcome is correct.

Joseph A. Calandrino, J. Alex Halderman, and Edward W. Felten provide the first discussion known to us of a transitive audit (without using that name) in "Machine-Assisted Election Auditing" (paper presented at the 2007 USENIX/ACCURATE Electronic Voting Technology Workshop, available at http://static.usenix.org/event/evt07/tech/full_papers/calandrino/calandrino.pdf), with a useful discussion of prior art. In the specific method discussed in Calandrino et al., serial numbers are added to the ballots and they are rescanned.

the audit can stop. This method is not a comparison audit: It does not require (or use) any information about the voting system interpretation of the ballots in the audit sample.

A ballot-polling audit has the disadvantage that it provides less information about errors than a comparison audit does. If a ballot-polling audit of 100 ballots contains 70 votes for the reported winner versus 20 for the runner-up, then it provides strong evidence that the outcome was correct – but it does not reveal how many (if any) of those ballots were initially miscounted. Also, as we will see, ballot-polling audits are not efficient when the margin of victory is small.

Advantages and disadvantages: expected counting, workload, and cost

The counting to be done in a risk-limiting audit is not entirely predictable before the audit is performed. (If we could know in advance how many errors an audit would find, we wouldn't have to conduct the audit!) But we can offer some generalizations. Here we compare the expected audit burden for three audit methods – batch-level comparison audits, ballot-level comparison audits, and ballot-polling audits – in various election scenarios.

Audit burden is a multi-faceted concept: We distinguish among counting, workload, and cost. *Counting* refers to the number of ballots examined, counted, or interpreted. *Workload* refers to the person-hours invested in the audit. *Cost* includes all direct and indirect expenses. Counting can be estimated before the audit if we make some assumptions about the accuracy of the original count. Workload is harder to estimate, because jurisdictions handle their ballots and VVPRs in different ways, and because some of the auditing methods described in this paper are not yet widely used. Cost adds further dimensions of variability, including who conducts the audit at what rate of pay, and whether there are special costs such as additional warehouse space, transporting ballots, or re-scanning ballots. In past audits, the reported cost to audit voter-marked ballots has been variously estimated as ranging from 3 to 51 cents per vote (not ballot) audited. Ballot-level audits generally are more expensive *per vote* than batch-level audits, but usually entail auditing many fewer votes.[15] Costs for auditing VVPRs typically are higher because the records often are difficult to handle and to read. Even the largest auditing costs are small compared with other election costs, but we will not attempt to offer specific estimates. Instead, most of our analysis will focus on counting, but we will point out workload and cost impacts as we go.

For any particular audit method, if the original count is accurate, the amount of counting required to attain a particular risk limit depends heavily on the margin of victory – but usually not very much on the size of the contest. For instance, in a ballot-level comparison audit, if the margin is 5% of ballots cast in the contest, auditing 96 ballots may suffice to reach a 10% risk limit – whether the total number of ballots cast was one thousand or many millions. The good news here is that risk-limiting audits of large contests often require much less counting than many current audit laws mandate. On the other hand, in small contests, the percentage of ballots to be audited

[15] Some of these estimates are referenced in Pamela Smith's written testimony before the Committee on House Administration, Subcommittee on Elections, March 20, 2007 (http://www.verifiedvotingfoundation.org/downloads/PamelaSmithTestimonyFinal_2007mar20.pdf). Hall et al. (see footnote 11) give examples of ballot-level comparison audits with costs of $0.35, $0.44, $0.46, and $0.51 per ballot.

can be very large, especially in a batch-level audit. (Of course, if the original count is inaccurate, more counting may be required.)

To provide some basis for comparison, we will consider auditing two hypothetical contests based on actual ballot counts from California's 2008 presidential election. The first is a statewide election with about 13.7 million ballots cast in over 38,000 separate batches (averaging about 356 ballots per batch). Obviously, this election is very large! The second election is a countywide election in Amador County, with just over 19,000 ballots in 59 batches. In most California jurisdictions, each precinct is reported as two batches: one for votes cast by mail and one for votes cast in person. We exploit this fact to reduce our counting burden: Smaller batches are more efficient.

Here we report estimates of the average number of ballots to be audited under each method for various reported margins of victory, at a 10% risk limit, assuming that there are no errors in the original count. Most elections in most jurisdictions are not very close, but jurisdictions should be prepared to audit contests with a wide range of reported margins. (Small error rates do not make much difference in most cases, as we discuss later.) Reducing the risk limit to 1% would require roughly twice as much counting in most cases; the increases are smaller when the fractions of ballots to be sampled are large. The technical details are relegated to Appendix 2, but we emphasize that these estimates are illustrative. Also note that although we use California data, the batch-level audit modeled here does not use the simple random sample mandated in California's current 1% manual tally, but instead uses a more efficient method that substantially reduces the counting burden.[16]

[16] The batch-level audit modeled here has two advantages over current California practice. First, instead of giving each precinct an equal chance of being included in the sample, it uses a more efficient method in which batches that could contain more counting error are more likely to be sampled. Second, in the statewide examples the batches or ballots are drawn from the entire state as a whole, rather than separately from individual counties.

Figure 1: expected audit size in California election

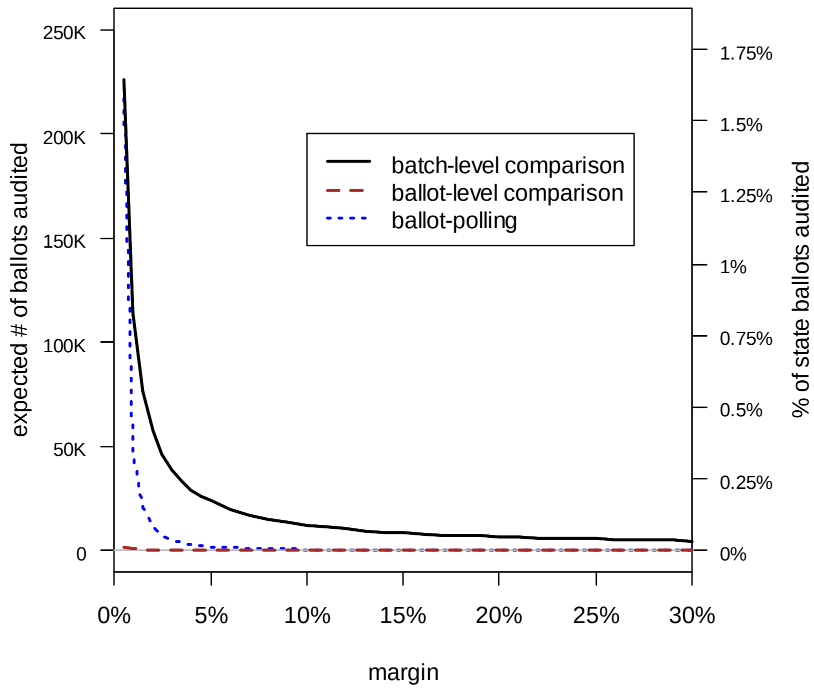


Figure 1 is a conventionally formatted graph of estimated average counting for the three methods in a statewide California election. (The smallest margin of victory shown is 0.5%.) The graph is somewhat lacking in detail because of the wide range of audit sizes (compare Figure 2 below). A few results stand out: (1) Ballot-level comparison audits entail hardly any counting compared with batch-level audits. (2) Ballot-polling audits do not require examining many ballots when the margin is greater than a few percent, but the workload expands rapidly for the smallest margins. (3) Although batch-level audits require the most counting, the *proportional* burden in a large election can be very small – here, less than 1.7% of ballots cast even for a margin of victory as small as 0.5 percent, if the sample of batches is drawn as described above.

Figure 2: expected audit size in CA election (logarithmic scale)

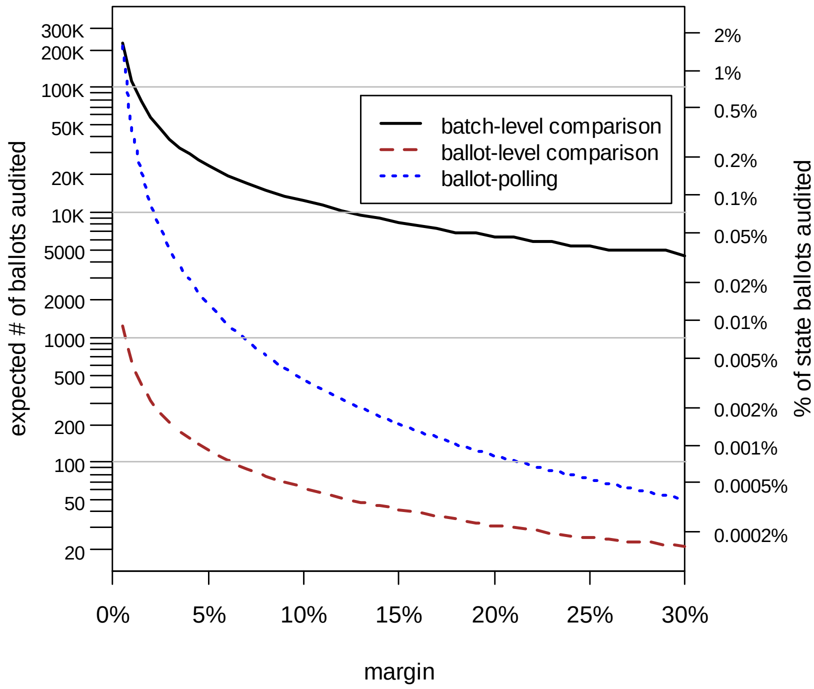


Figure 2 offers a different look at the same analysis of statewide audit burden. It uses a logarithmic vertical scale. For instance, the distance from 100 ballots to 1000 ballots is the same as the distance from 1000 ballots to 10,000 ballots. Although this format is less familiar than that of Figure 1, it provides more usable detail. We can see that a ballot-level comparison audit is expected to require fewer than 100 ballots for margins down to about 6%, and fewer than 1000 for margins down to about 0.6%. A ballot-polling audit is less efficient, but still requires fewer than 1000 ballots (in the entire state of California!) for margins down to about 7%. In this scenario, a ballot-polling audit is expected to require manually examining fewer ballots than a batch-level comparison audit for margins of 0.5% or more. However, because auditing (say) 300 ballots in 300 different batches requires more effort than counting 300 ballots in one batch, the breakeven point for workload surely is at a larger margin.

In Figure 2, notice that the curve for batch-level comparison audits has much the same shape as the curve for ballot-level comparison audits, although it is much higher. This is not coincidental. The expected number of *batches* to be audited in a batch-level comparison audit, in this example, is consistently a bit less than half the expected number of *ballots* to be audited in a ballot-level comparison audit. (The number of *ballots* to examine for a batch-level audit is much larger – how much larger depends on batch sizes and on how much miscount each batch can hold.) A similar relationship holds in other audits, constrained by the number of batches, as in the next example.

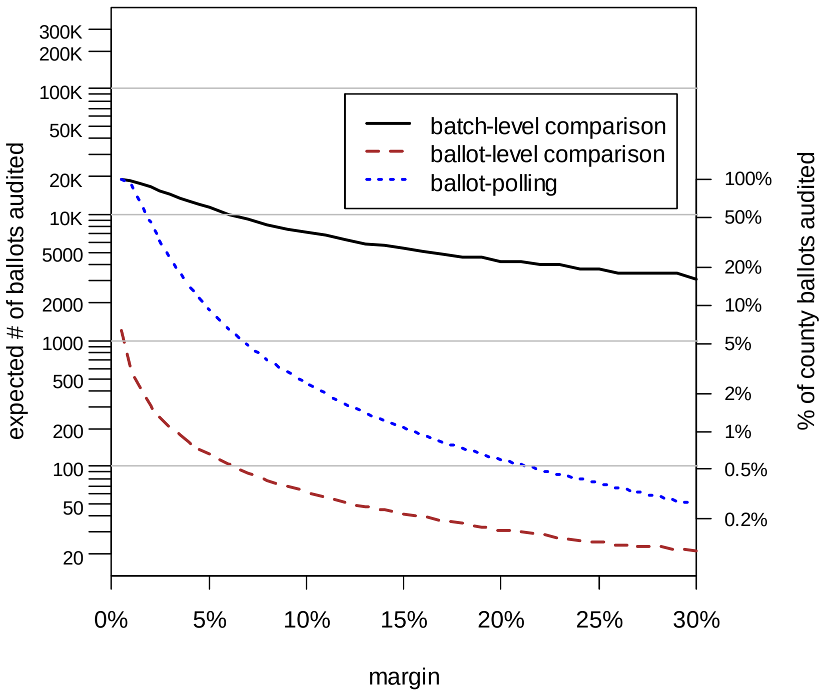
(Table 1 on page 45 reports the statewide analysis in tabular form. Again, we caution readers that the numbers should not be taken literally.)

Figure 3 depicts average audit counting estimates for countywide contests in Amador County, with just 19,006 total votes. Figure 3 is displayed on the same (logarithmic) scale as Figure 2 to facilitate comparison, but the right-hand scale shows percentages for the county rather than the state. In many cases, the difference in counting burden is not as large as one might expect. The expected counting for ballot-level comparison audits is practically indistinguishable from the statewide scenario.[17] The counting in ballot-polling audits is also similar for margins greater than about 5%. For smaller margins, the workload cannot increase as quickly as it does statewide, because it is impossible to audit more than 100% of the ballots. (At some point it would be easier to hand-count all the ballots than to do a ballot-polling audit of a great many of them.)

The counting required by batch-level audits in Amador County is somewhat lower for all margins, because the county's batches are relatively small. Even for a 30% margin the fraction of ballots to be audited in a batch-level audit is rather large. Although such a contest could be audited using just 10 batches, that number represents about 17% of the 59 batches in the county election. As the margin decreases, the counting for a batch-level audit flattens out, again because it is impossible to audit more than all 59 batches. Overall, batch-level audits are demanding for relatively small contests, because a large proportion of ballots must be counted even for moderately large margins. The greater efficiency of ballot-level comparison audits is accentuated in these smaller contests.

[17] The *workload* is probably less, because in small contests, ballots will be drawn from the same batch much more often than in a statewide audit. (It is easier, for instance, to inspect 20 ballots from one batch than to inspect 20 ballots from 20 different batches.) On the other hand, a statewide audit is spread over 58 counties, so the workload *in Amador County* to audit a local contest is generally greater than the county's share of the workload to audit a statewide contest with the same margin.

Figure 3: expected audit size in Amador election (logarithmic scale)



Election jurisdictions can be much smaller than Amador County. If a jurisdiction has only a few batches, it will often have to hand-count all of them to achieve a low risk limit using a batch-level audit. Ballot-level comparison audits generally require similar *numbers* of ballots (for a given margin) unless the fraction of total votes to be audited is quite substantial. That burden is often more manageable than the comparable burden for a batch audit, but full hand counts still are more likely in very small jurisdictions than in larger ones.[18]

Auditing multiple contests at once

So far we have only considered auditing one contest at a time – but many elections have multiple election contests, sometimes dozens. What if a jurisdiction conducts risk-limiting audits of several contests at once? Right now we are considering the *scope* of audits, so we need not discuss all the implementation details. The most important considerations are these:

It is possible to audit several or many contests using a comparison audit with a single sample. Auditing multiple contests with one sample is called a <u>simultaneous audit</u>. Simultaneous audits can sharply reduce the time spent locating and retrieving ballots. Simultaneous audits are

[18] For instance, a statewide ballot-level comparison audit for a margin of 2% is expected to take about 300 ballots. In a contest with just 600 total ballots, a ballot-level comparison audit for the same margin is expected to require about 236 ballots – about 40% instead of 50%. It would probably be easier just to hand-count all 600 ballots.

possible in both batch and ballot-level comparison audits. They even can span contests that appear in different jurisdictions or parts thereof, although this is not always efficient.

In a simultaneous audit, each contest must be audited on all the sampled ballots on which it appears. The counting to be done in a simultaneous audit heavily depends on the smallest margin as a fraction of all the ballots subject to the audit. (This fraction is called the diluted margin.) Therefore it can be inefficient to audit certain contests together in a simultaneous audit – for instance, a very close (or very small contest) together with several contests that are not very close, or contests that never appear on the same ballot. Such contests can still be audited at the same time using separate samples, which could include multiple simultaneous audits.

Simultaneous audits are not free lunches. They can substantially reduce overall effort, but auditing multiple contests on a ballot still takes more time than auditing just one.

Completing audits "in time"

The appropriate scope of an audit may depend on its timing – how long after the election it begins and ends, and how it interacts with various provisions of election law. By definition, a risk-limiting audit must be capable of correcting incorrect outcomes, either directly or by leading to a separate full hand count. In many states, election officials have a limited period of time to prepare and release certified results, followed by a period of time in which candidates can file challenges and request recounts. Ideally, election officials will complete all audits before certification, but this goal may be untenable with current election calendars. Risk-limiting audits rely upon complete or almost complete preliminary vote counts (or at least ballot counts),[19] which may not be available early in the certification period. Coordinating audits that cross county or other jurisdictional lines may be difficult, especially if the audit uncovers errors that lead to further auditing. Laws might allow an audit conducted entirely after certification to lead to a change in election outcomes, but this may raise other legal complications. Presidential elections pose a special challenge because the "safe harbor" date for choosing electors is just five weeks after Election Day. In the final analysis, election calendars should never pose an insuperable obstacle to risk-limiting audits, but they may impose short-run logistical constraints.

Choosing a risk limit

The previous analyses are for a 10% risk limit. As we mentioned above, a 1% risk limit often requires about twice as much counting as a 10% risk limit. (This is true both for comparison audits and for ballot-polling audits.) On the other hand, a 25% risk limit often requires about 40% less counting than a 10% risk limit. On what basis should a risk limit be chosen?

The choice of risk limit is a prudential judgment, but it is important to understand the meaning of "risk" in this context. An audit with a 10% risk limit is designed to have *at least* a 90% chance of leading to the correction of any incorrect election outcome, based on worst-case assumptions about how the votes and tabulation errors actually are distributed. For instance, the risk calculations essentially assume that the actual outcome is a tie vote, thus minimizing the number

[19] Ballot-polling audits do not rely upon vote counts, but do require knowledge of how many voted ballots there are, and where they can be found. See the discussions of ballot manifests below.

of votes that would have to be miscounted. For a batch audit, they assume that all the ballots in each batch may have been cast for any one of the candidates – no matter how implausible that might be – thus minimizing the number of batches that would have to contain miscounted ballots. A 10% risk limit does not imply a 10% chance that any outcome is wrong. On the contrary, it assumes that the outcome is wrong – in the way that is hardest to detect short of subverting the audit trail – and nevertheless guarantees at least a 90% chance of correcting it. Although audits with a 10% risk limit do not offer absolute assurance that every audited election outcome is right, they do provide strong evidence.

Because risk limits are conservative, a 10% or even higher risk limit is not unreasonable, especially if it allows risk-limiting audits to be extended to a greater range of contests. The choice of risk limits should consider the likely overall counting burden. As an extreme example, setting a 10% risk limit to audit just one statewide contest may be a false economy. Unless the contest is unusually close, the cost of the audit itself will be tiny (and may be dominated by reporting overhead), and the audit will not yield much information about how voting systems performed. If a state does choose to audit so narrowly, it would be reasonable to set a 1% risk limit and/or to establish some minimum audit size.

A state may choose to set a low risk limit for statewide contests and a higher risk limit for smaller contests. This choice may be construed as implying that local contests are less important than statewide contests; it can also be construed as a concession to practical challenges in auditing small contests. In considering this issue, one should closely consider the status quo. Currently, when states audit local contests at all, they generally audit the same percentage of batches as in statewide contests – which often means a much smaller sample size and, in many scenarios, little chance of even noticing miscounts that may change the outcomes. Thus, the current "equal treatment" in these states is far from equal in its implications for risk. Setting *any* risk limit for *any* contest would be a large step forward.

Implications for audit requirements

How widely to extend risk-limiting audit requirements ultimately is a political question, but an understanding of the available audit methods helps to inform the decision.

Ballot-level comparison audits are the most efficient approach, but often are not feasible with existing federally certified voting systems. Therefore, it may be desirable to introduce risk-limiting audits in phases; for instance, to require audits of a few important contests in the short run, to be extended once better voting systems are put in place.

Which contests should be audited? If and when ballot comparison audits are possible, it may be feasible to require risk-limiting audits of every contest. If a contest is decided by a double-digit percentage margin – as very many are – and if the original count was accurate, then its outcome generally can be confirmed at a 10% risk limit by checking fewer than 50 ballots, in some cases fewer than a dozen.[20] To be sure, small jurisdictions (including small towns and villages) may

[20] A pilot audit in Alameda County, California audited four contests simultaneously at a 10% risk limit by examining just 17 ballots. See the March 1, 2012 report cited in footnote 11, pages 12-13 (PDF pages 13-14).

be less sanguine about this requirement than larger jurisdictions. Moreover, the number of contests varies widely across the country. A state might exempt small jurisdictions from a risk-limiting audit requirement, or apply the requirement only to certain contests (perhaps selected at random). Or it might require risk-limiting audits only in federal and state contests, including district-based contests such as those for the state legislature.

Even the less efficient batch-level comparison audits offer a viable means for auditing statewide contests in most states – arguably in every state. Batch-level audits can work well in other large jurisdictions as well. Typically, they are affordable for congressional contests, which often span 400 to 600 precincts or more; they may be suitable for state legislative contests depending on the number of batches. Batch-level audits are least efficient when there are few batches, as in the Amador County example. Dividing large batches into smaller auditable batches can increase efficiency.[21]

To audit only one contest, a ballot-polling audit may be the most efficient approach, especially if the contest is not very close. In a large jurisdiction, auditing even a few thousand ballots selected randomly from throughout the jurisdiction may be manageable; most contests would require examining substantially fewer ballots than that. In closer contests (perhaps those with margins under 5 points), a batch-level comparison audit may be preferable for two reasons: It can reduce work – including the overhead of retrieving the ballots in the sample – and it provides more direct evidence about possible errors in the count. In very close contests, a full hand recount may be preferred. (We say more about full recounts later.)

Many jurisdictions seek a middle ground between auditing just one or two contests and auditing all possible contests. One useful approach is to choose some contests to audit at random, or by some discretionary means. For instance, Wisconsin's audit (which is not risk-limiting) is conducted as follows:

> A minimum of four (4) contests shall be audited, including the top candidate race on the ballot (either gubernatorial or presidential). The other audited contests shall be selected randomly by the State Elections Board from all other contests that appear on the ballot, but must appear on every ballot in the state. The State Elections Board may, with prior notification, direct that additional contests be audited.[22]

Similarly, local jurisdictions might be required to conduct risk-limiting audits not in all local contests, but in some contests selected at random, or perhaps chosen by election officials or political parties. In this way, it can never be known in advance that a particular contest *will not* be audited. This approach can deter fraud and enhance the credibility of the audit.

[21] For instance, some jurisdictions report absentee vote totals in large batches of several thousand votes. If the absentee ballots can be counted and reported in smaller batches for purposes of the audit, the required counting may be much reduced.

[22] Wisconsin State Elections Board, "Voting System Audit Requirements," November 2006, http://elections.state.wi.us/docview.asp?docid=9851&locid=47, page 4. (Similar language appears on page 6.)

Preparing and securing the audit trail

Defining the audit trail

A risk-limiting audit requires evidence of the voter's intent that is presented to the voter for verification before she casts her ballot, and is protected from loss, alteration, substitution, and dilution (through the addition of invalid records) thereafter. An audit of software artifacts (cast vote records, logs, etc.) alone may be useful for detecting certain malfunctions, but it *cannot* be risk-limiting, because it cannot measure or control the danger that the artifacts being audited misrepresent voter intent. The Resolution on Electronic Voting, endorsed by thousands of computer technology experts, attorneys, public policy analysts and others, elaborates:

> Computerized voting equipment is inherently subject to programming error, equipment malfunction, and malicious tampering. It is therefore crucial that voting equipment provide a voter-verifiable audit trail, by which we mean a permanent record of each vote that can be checked for accuracy by the voter before the vote is submitted, and is difficult or impossible to alter after it has been checked. Many of the electronic voting machines being purchased do not satisfy this requirement. Voting machines should not be purchased or used unless they provide a voter-verifiable audit trail; when such machines are already in use, they should be replaced or modified to provide a voter-verifiable audit trail. Providing a voter-verifiable audit trail should be one of the essential requirements for certification of new voting systems.[23]

The Association for Computing Machinery, the world's largest and oldest organization of computer professionals, has since 2004 taken a position in favor of voter-verified physical records:

> Voting systems should also enable each voter to inspect a physical (e.g., paper) record to verify that his or her vote has been accurately cast and to serve as an independent check on the result produced and stored by the system. Making those records permanent (i.e., not based solely in computer memory) provides a means by which an accurate recount may be conducted.[24]

By far the most efficient known means of providing a suitable audit trail is to base the election on paper ballots marked by the voters, either directly or through the use of a ballot marking device. Paper ballots are relatively durable, indelible, and easily handled. When voters mark their own ballots, they can verify their votes at the time they mark them, and again before casting their ballots. Some voters are unable to mark their ballots directly, or have difficulty doing so. Voting

[23] Resolution on Electronic Voting, available at http://verifiedvotingfoundation.org/article.php?id=5028
[24] "E-Voting," policy statement of the Association for Computing Machinery, available at http://usacm.acm.org/evoting/category.cfm?cat=14&E-Voting

systems should provide ballot marking devices, and means of verifying the ballots those devices produce, that are accessible to as many voters as possible.

As an alternative, electronic voting machines equipped with voter-verifiable paper records (VVPRs)[25] may provide adequate records for a risk-limiting audit if voters in fact check the VVPRs, and if the records are complete and intact at the time of the audit. Both these conditions are problematic. Research studies[26] have found that many voters fail to check VVPRs, and fail to notice deliberately introduced errors even when they do check. (The term voter-*verifiable* paper audit trail underscores that the records may not have been verified.) Unfortunately, this danger is not merely abstract: Security reviews of currently deployed voting systems have demonstrated the feasibility of attacks that produce false VVPRs.

Moreover, even when the VVPRs are correct in principle, they may not be usable due to printer failures or subsequent damage. A collaborative unofficial audit of a 2006 election in Cuyahoga County, Ohio, found that almost 10% of VVPRs were “either destroyed, blank, illegible, missing, taped together or otherwise compromised."[27] Often VVPRs are produced on long spools (sometimes derisively called “toilet paper rolls”) that make it hard to read and handle individual records, impairing their usability even under the best circumstances.

If it is found that VVPRs or paper ballots have been damaged or compromised, a risk-limiting audit should treat the evidence in the most pessimistic possible way – for instance, by assuming that a true and correct paper record would reveal that all those ballots showed votes for the runner-up. If so many VVPRs or paper ballots are missing, damaged, or compromised that these pessimistic assumptions would alter the outcome, then the outcome should be deemed inexorably uncertain. Whether to order a revote in such circumstances, or to make do with some alternative resolution, is a policy question beyond the scope of this paper.

In principle, it is possible to deploy voter-verifiable audit trails that are not paper. These records could be usable in risk-limiting audits provided that voters generally verify them and that they are amenable to auditing. It has been argued, for instance, that Voter-Verifiable Audio Audit Trails – audible to voters and simultaneously recorded onto write-once DVDs – could help voters detect errors. The prospect of auditing possibly millions of hours of audio records without heavy reliance on software is daunting; at any rate, no such systems have been deployed. It is unknown when and whether viable alternatives to paper-based audit trails will be available.

Securing the audit trail

No matter how attractive the inherent properties of an audit trail, it is only as reliable as it is secure. Past elections have been tainted by allegations – and even strong evidence – of ballot box “stuffing” after the election; anecdotes abound of ballots gone lost. Auditing or recounting an

[25] VVPRs sometimes, confusingly, are called VVPATs (“voter-verifiable paper audit trails”). “VVPAT” properly refers to the complete set of VVPRs that document votes in a particular election.

[26] For instance, Sarah P. Everett, “The Usability of Electronic Voting Machines and How Votes Can Be Changed Without Detection,” Ph.D. dissertation, available at http://chil.rice.edu/research/pdf/EverettDissertation.pdf

[27] Marc L. Songini, “Paper Trail Flawed in Ohio Election, Study Finds.” *Computerworld*, August 21, 2006, http://www.computerworld.com/s/article/9002610/Paper_Trail_Flawed_in_Ohio_Election_Study_Finds.

untrustworthy audit trail yields untrustworthy results. Moreover, it is highly desirable not only to assert that the audit trail has been secured, but to be able to demonstrate that it has. Some analysts speak of a compliance audit to verify that the preconditions for a risk-limiting audit have been satisfied.

A comprehensive discussion of audit trail security is beyond the scope of this paper, but the following points should be considered. These suggested requirements are based on the work of Roger Johnston of the Vulnerability Assessment Team at Argonne National Laboratory.

The first group of points addresses issues to be scrutinized in a compliance audit or concurrent with the risk-limiting audit.

- A random sample of machines or containers shall[28] be closely scrutinized for evidence of tampering. (This sample could be coterminous with the sample in a batch-level comparison audit, or it could be separate.)

- Ballot containers shall be sealed with tamper-indicating seals. When ballot containers (and similar packages) are opened, the seal serial numbers shall be carefully compared with the numbers on the data logs.

- In a ballot-level audit, the ballots in the containers that are opened for the audit (or some minimum number thereof) shall be counted for comparison with the reported counts. Any discrepancies in these counts can be incorporated in the risk calculations.

- A chain of custody log shall be kept for each ballot container; each recipient of the asset shall sign the log and legibly print his or her name, along with the date. The chain of custody log shall be reviewed for each ballot container opened during the risk-limiting audit (and/or for a sample of containers selected in the compliance audit).

The next set of points addresses policies that help to secure the audit trail.

- Poll workers shall take an oath to defend the integrity and security of the election process.

- Poll workers and election officials who open locks or tamper-indicating seals on voting machines or ballot containers shall have training on how to detect tampering with the locks or seals. This training shall involve at least 10-20 minutes per type of seal used, and shall include both photographic and hands-on examples of tampering.

- Locks on voting machines shall not all open with the same key.

- Data logs of seal serial numbers shall be secured.

- Used seals shall be thoroughly destroyed on a prescribed date after the election.

[28] Our use of “shall” reflects the recommendation that these points be required, although we do not provide detailed implementation language here.

Additional security procedures are not integral to risk-limiting audits per se, but can be beneficial in forestalling attacks.

- Each jurisdiction shall convene an Election Security Board, independent of election officials, to review and analyze election security. Each jurisdiction shall appoint a Chief Election Security Officer (who may also have other duties).

- Each jurisdiction shall randomly select 1% of voting machines and intensively examine them for evidence of tampering – ideally *before* the election, otherwise within six weeks after the election.

- Each jurisdiction shall conduct background checks, repeated every 3 to 5 years, on all election officials, technicians, contractors, or volunteers who prepare, maintain, repair, test, inspect, or transport voting machines or compile substantial amounts of election results.

Trustworthy audits: the virtue lies in the details

Risk-limiting audits are easy to define, and in broad outline they are fairly easy to implement: Draw a sample, look at the ballots in the sample, do some math to see if more counting is required. However, some implementation details need careful attention.

Public observation and transparency

Risk-limiting audits provide one means for citizens to monitor how well election systems are functioning. Audits provide valuable information to election officials, but crucially, they inform the public and provide evidence as to whether reported election outcomes are correct. The *Principles and Best Practices for Post-Election Audits*[29] state the case as follows:

> Elections belong to the public. The public must be allowed to observe, verify, and point out procedural mistakes in all phases of the audit without interfering with the process. The following conditions must be met:
>
> a. Detailed auditing procedures are developed well in advance of elections, with reasonable opportunities for public comment. These include procedures for selecting audit units [i.e., ballots or batches to be audited], sorting the paper records and counting the votes, and determining when more units need to be audited and when the audit can end. There is adequate notice to allow the public to witness and verify each phase of the audit.
>
> b. The public is given sufficient access to witness and verify the random selection of the audit units as well as the manual count with reasonable opportunities for public comment. Election officials have the authority to prevent the public from hampering the proceedings.

These principles impose important responsibilities both on election officials and on public observers. When all parties take these responsibilities seriously – but not grimly – audit observation builds positive relationships between election officials and the citizens they serve.

Good audits are confidence-building exercises; not-so-good audits are more like sullen skirmishes. In the past, some audit observers and would-be observers have reported events like these: never receiving advance notice of audits despite statutory or regulatory requirements; being confined in one corner of a room with no meaningful opportunity to observe; receiving no information about the procedures to be used; having no opportunity to ask basic questions; witnessing unambiguous violations of written procedures but being unable to persuade officials to refer to, or conform with, those procedures. Contrariwise, many other observers have reported interacting cordially with election officials and workers, in some cases politely making suggestions that were immediately adopted, and generally forming a favorable opinion of the

---

[29] *Principles and Best Practices for Post-Election Audits*, September 2008, available at http://electionaudits.org/files/bestpracticesfinal_0.pdf. (Here we quote from page 6.)

audit and other election processes.[30] Clear written procedures, made available in advance of the audit, help observers and other interested citizens understand how the audit evinces the integrity of the results.

Some states provide for partisan observers in certain election audit processes. We recommend that audits be explicitly open to non-partisan observers as well. All interested individuals and groups should be permitted to observe the audit process to the greatest possible extent. Effective audit observation can increase public confidence in the audit and in the integrity of elections, by making the process more transparent and providing an independent verification of the results.

Observability includes not only direct public observation of audits, but clear reporting of the audit findings. Election officials should systematically report audit results, identifying any differences between the audit and voting system counts, and explaining them if possible. These reports need not be long in order to be informative and reassuring.[31] The audit results should be forwarded to state election officials, who in turn should compile a summary statewide report – perhaps within 30 days of completion of the audit. Among other things, this report should integrate results from local officials in a consistent, comprehensible, searchable format.[32] Such a report may enable state officials to detect patterns of error they otherwise may have missed, or simply to document how well voting systems performed. A national database of audit results would give additional information.

Basic elements of public observation of audits, including notice, access, and reporting, probably belong in statute. Procedural details can be handled in various ways. Questionnaires for observers designed by the Verified Voting Foundation[33] and Citizens for Election Integrity Minnesota[34] provide useful criteria for evaluating observation procedures.

Committing to auditable results

To conduct effective post-election audits, the necessary election result data must be exported from voting systems in a format that is easy for both people and software to "read." These data should be committed to, that is, declared in a manner that prevents them from being undetectably

[30] David Wagner's "Thoughts on the Nov 16, 2006 1% Manual Tally in Yolo County" (available on Yolo County's election website at http://www.yoloelections.org/news/snews/reactions.pdf) exemplify constructive interactions between election officials and observers.

[31] As an example, consider Cuyahoga County's (Ohio) post-election audit summary for the 2010 general election. The one-page summary (available at http://boe.cuyahogacounty.us/pdf_boe/en-US/11022010PostElectionAuditSummary.pdf) reports 7 discrepancies out of 34,996 ballots audited, and offers the best available explanation for each discrepancy. The accompanying detailed spreadsheet report, and other audit reports and summaries, are available at http://boe.cuyahogacounty.us/en-US/audit-materials.aspx.

[32] In North Carolina, William Kalsbeek, the independent statistician charged with supervising the audit, has routinely produced both a written report summarizing and characterizing the results, and a spreadsheet with detailed results in a common format, available on request.

[33] "Observing the Manual Audit, Precinct Version," Verified Voting Foundation, available at http://www.verifiedvotingfoundation.org/downloads/AuditQ-Precinct.pdf.

[34] "Appendix C: Audit Observation Report Form," pages 42-45 in Mark Halvorson and Sarah Martyn Crowell, *Eyes on the Vote Count: Non-partisan Observer Reports of Minnesota's 2008 Post-Election Audit and Recount*, Citizens for Election Integrity Minnesota, available at http://www.ceimn.org/files/ceimn.report_color.pdf.

changed later. (This property is sometimes described as *indelible* commitment.) Data reporting is a challenge in itself, but some audit methods impose additional challenges.

Election data problems are complicated because elections are complicated. Currently, many states have heterogeneous voting systems with a variety of vote capture devices. Even within a given precinct, there are often multiple types of vote capture devices; there might be ballot scanners to tabulate voter-marked paper ballots and direct recording electronic (DRE) voting machines at the same polling place, and central-count ballot scanners for absentee and provisional ballots. Current election data systems sometimes cannot report detailed results intelligibly. Some export the details only in formats suitable for printing tables for human readers, but inconvenient for further quantitative analysis.

Unreadable or inscrutable election data are intolerable, and not only for audits. Detailed election results must be provided in a single standard, structured, machine-readable format that is easy to export and then analyze. The data should be clearly documented to facilitate analysis. The format should be the same for all makes of equipment; ideally, all jurisdictions will adopt a common standard such as the Election Markup Language (EML).[35] Providing data in a common format not only facilitates auditing but also enables interested citizens to analyze and understand the results.

If all audits were batch-level comparison audits based on the commonly reported results – most often precinct totals or precinct subtotals by vote type – then we would have little more to say about the reporting issue: The detailed election results would be directly usable for the audit.[36] Unfortunately, several complications can arise.

Some jurisdictions' heavy use of vote-by-mail ballots can complicate batch-level audits. Typically, these ballots are counted by central scanners in large batches as they are received, but the canvass results are reported by precinct. To audit these ballots at all entails some way of harmonizing what is reported with what can be audited. Either the ballots must all be sorted by precinct, some other means must be found of retrieving all mail ballots for selected precincts, or the election data system must be able to report results by physical batch *as well as* by precinct. Then the mail ballots can be scanned and reported in batches of any desired size. Sorting mail ballots by precinct is feasible in many jurisdictions (and already required in some), but better election data systems provide valuable flexibility.

[35] The IEEE Standards Association's Project 1622, Standard for Voting Equipment Electronic Data Interchange, has the long-term purpose of attaining a common definition of election data. (See the group page at http://grouper.ieee.org/groups/1622/.) Initial efforts have focused on extending the OASIS Election Markup Language (OASIS EML) standard to support electronic distribution of blank ballots.

[36] However, we must warn that some jurisdictions currently report precinct subtotals that can undermine voter privacy, as further discussed below. If a canvass report indicates that five voters in a precinct cast provisional ballots, and all five voted for a particular candidate, then anyone who can determine the identity of those voters also knows how they voted. Election officials should take care, as the U.S. Census Bureau does, to protect people's privacy when reporting geographic details.

It is important to commit to the number of ballots in each physical batch – which, again, may not correspond to the (typically precinct-level) results in a canvass report. These counts comprise the ballot manifest, which is further described in the section on random selection.

Committing to ballot-level results raises considerations that merit their own section.

Voter privacy and other concerns in ballot-level comparison audits

Conceptually, a ballot-level comparison audit is simply a special kind of batch-level comparison audit. Suppose that every ballot has a preprinted unique identification number, and that scanners read these numbers and store them as part of the Cast Vote Records (CVRs) that record ballot interpretations. Then it should be relatively simple to select a random sample of CVRs, retrieve the corresponding ballots, and check whether the audit interpretation of the ballot matches the CVRs. All the CVRs could be published online so that anyone who wants can confirm that they add up to the announced vote totals.[37] Conceivably, actual digital images of the ballots – or at least the audited ballots – could be published so that people can draw their own conclusions. Is anything wrong with this picture? Many people think so.

Publishing CVRs – especially with preprinted ballot identification numbers – and/or publishing digital images of the ballots, would greatly increase the available information about the election, but at the cost of undermining voter privacy. Opinions differ about whether this cost, or part of it, is worth paying. Voter privacy has two aspects. One aspect is whether and to what extent voters *involuntarily* reveal their votes to others. For instance, pre-numbered ballots linked to CVRs presumably allow pollworkers to note which ballots they have given to whom, and thereby to determine how individual voters voted. The second aspect is whether and to what extent voters can *voluntarily* reveal their votes. For instance, if images of the ballots are published, then voters could put some distinctive marks on the ballots – or even write their names—unless the images are somehow "cleaned" before publication. Involuntary violations of privacy may be construed as worse than voluntary violations, but both are concerns. If I can voluntarily reveal how I voted, then not only can I sell my vote, but conceivably I can be coerced to cast it in a particular way. Free elections may benefit from denying voters the "freedom" to *prove* how they have voted.

CVRs pose a tradeoff between observability and privacy: to publish or not to publish? If the CVRs are not published, then only a limited number of people can confirm that the CVRs used in the audit add up to the reported vote totals. (Even doing that is not a trivial problem; we discuss it below.) If the CVRs are published, then voters might be able to reveal their identities through pattern voting: casting votes for a combination of candidates that is very likely to be unique within a precinct or batch. If a ballot has ten contests with three candidates per contest, a voter can vote as directed in one contest and vote the other nine contests in almost 20,000 distinct ways – even without allowing for write-ins. There is a path between these alternatives, called SOBA (briefly discussed below), which has its own drawbacks but is worth considering.

[37] The remaining challenge, discussed below, is to demonstrate that the published CVRs are the CVRs used in the audit.

Some people argue that the danger of pattern voting is worth accepting on one or both of two grounds: that large-scale vote-selling and/or coercion would be too complex to evade detection, and that the wide use of mail ballots already facilitates vote-selling and coercion. Reasonable people can disagree about these issues. Here we will consider both CVR publication and alternatives.

A ballot-level comparison audit entails some means to compare a CVR with the corresponding physical ballot. One approach, described earlier, is a transitive audit. A transitive audit reduces the privacy compromise because it does not require printing identifiers on the ballots before the ballots are cast. This solution does not resolve the dilemma of whether to publish CVRs, but at least it protects privacy better than preprinting identification numbers on ballots. An alternative is to add embossers to the voting system scanners, so that the ballots are imprinted with identification numbers at the time they are cast, numbers that cannot be linked to the voter identity. Note, however, that using serial numbers (in numerical order) – or even publishing the CVRs in the order in which votes were cast – may compromise voter privacy, especially if voters are required to sign a list at the polling place.

How can an election official commit to a set of CVRs without publishing them? One approach is to copy the CVRs to a write-once DVD and to calculate a cryptographic hash for the set of CVRs. A cryptographic hash (or hash value) is a distinctive digital "signature" based on a set of data, using a particular hash function. Changing any part of the data – even swapping two non-identical pieces – radically alters the hash. It would be difficult to prove physically that a DVD has not been altered, but it is relatively simple to prove that a set of CVRs has not been altered: If it produces the same hash, it is essentially certain to be the identical data. So, in principle, observers can borrow the DVD (or a copy thereof, to be destroyed at the end of the audit) and verify that it contains the same CVRs being used in the audit and that those CVRs match the reported results. Of course, to verify those facts may require some technical savvy – and if observers have the technical means to independently verify a hash and a vote count, won't they also have the means to copy the CVRs? It seems hard to guarantee that voter privacy will not be compromised, but this approach certainly is more cautious than publishing the CVRs.

A more elaborate effort to protect voter privacy without compromising transparency is called SOBA, for Secrecy-preserving Observable Ballot-level Audit.[38] The basic idea behind SOBA is that the CVRs are published in pieces – one vote per record, instead of an entire ballot per record – along with a cryptographically shrouded link that associates each piece with the corresponding ballot. Because all the votes are published, anyone can verify the contest results. The shrouded links, by themselves, do not allow people to reconstruct which votes came from the same ballot. However, when a ballot is sampled, election officials reveal additional information that, when combined with the shrouded links, proves that particular CVR pieces indeed came from the same ballot. Perhaps the biggest problem with SOBA is the difficulty of explaining in a few sentences how it works. People who did not actually participate in the audit – and perhaps some people who did participate in the audit – might well feel that a published set of individual *votes* with

[38] Josh Benaloh et al., "SOBA: Secrecy-preserving Observable Ballot-level Audit." Paper presented at the 2011 EVT/WOTE annual conference. Available at http://static.usenix.org/event/evtwote11/tech/final_files/Benaloh.pdf.

shrouded links is less convincing than a published set of complete ballot interpretations. Also note that SOBA (per se) does not solve the problem of identifying physical ballots, through numbers or otherwise. Nevertheless, SOBA does offer a way to publish very detailed data without compromising voter privacy.

The various complications in implementing ballot comparison audits must be weighed against the large efficiency advantages.

Randomness: "too important to leave to chance"

Risk-limiting audits crucially depend on random samples. Formally, the statistical calculations assume randomness, rather than "cherry-picking" particular ballots or batches. (It is perfectly fine for a risk-limiting audit to include some consciously chosen batches, but these must not be treated as if they were selected at random.) Informally, people – quite rightly – find audits most credible when no one can control which ballots are audited, and when the ballots are inspected as soon as possible after they are selected.

Unfortunately, some methods used to draw "random" audit samples are flawed – some fatally so. Here is how an observer of the 3% partial recount in one Ohio county after the 2004 general election describes the county's selection procedure that year (other counties have sometimes used similar procedures):

> The total number of votes cast in Morrow County was 16,694. Three percent of this would be 501. They [election officials] picked Harmony Township [North] to do the hand count on because it has 517 ballots cast.
>
> At this point Observer S. spoke up and said that the precinct had to be selected at random.
>
> He explained that he felt that random was something like putting 36 precincts' numbers in a bowl and pulling one, if it was short [of 3%], we would have to pull another one.
>
> This was not well taken. They said that they felt how they selected was random and that the board had the right to choose. They showed us in writing where they were getting this, and of course it is all in the interpretation of "random."[39]

In reality, "the right to choose" is roughly the antithesis of random selection, not an interpretation of it. Perhaps the officials meant that they had not selected the precinct before the election, so the selection was effectively random. But regardless of when the officials made their choice – a datum that observers could in no way verify – it was anything but effectively random. This preference for selecting a precinct of a certain size arbitrarily exempted most precincts from

[39] Morrow County, Ohio 2004 Ballot Recount Observer Report, December 17, 2004: Report by Green Party Observer, available at http://www.iwantmyvote.com/recount/ohio_reports/counties/morrow.php.

any chance of selection. In the worst case, an attacker could anticipate and exploit this preference to subvert the election while evading detection. At best, this approach undermines the representativeness of the sample and the credibility of the partial recount. Since 2004, Ohio state election officials have worked to prevent similar misinterpretations.

While some failings of random selection methods leap off the page, others are more subtle. Some methods are demonstrably biased toward auditing some votes rather than others.[40] Some random selection procedures may work reasonably well, but it is impossible to know or to prove that they did. For instance, some jurisdictions have printed the precinct names on slips of paper, mixed the slips in a box, then drawn slips. But how can observers be sure that every precinct name appeared on exactly one slip? And how can anyone be sure that the slips are fully mixed?

A good random selection method has a few essential properties. Nobody should be able to *influence* the selection to favor certain ballots or sets of ballots over others. Nobody should be able to *predict* that certain ballots or sets of ballots will be favored over others, except to the extent that the audit method deliberately does so.[41] And the preceding factors should be *observable*: As many people as possible should be able to confirm that the selection method was fair.

One approach to creating samples with these properties is to use a single source of randomness that is considered to be unpredictable and immune from influence. Certain websites provide sequences of random or quasi-random numbers that, in principle, should be beyond anyone's capacity to influence or to predict. These should not be used in audits because they fail the observability criterion: How can observers be certain that the numbers they obtain using a website really are random? What is to prevent someone from hacking the sequence? Similarly, software programs such as Microsoft Excel should not be relied upon. A reasonable alternative may be a ping-pong ball machine such as many state-run lotteries use to choose their winning numbers. These machines can be used for audit samples if stakeholders trust them. However, it may not be obvious how to demonstrate that a particular machine is fair, and it may not be practical for all jurisdictions to obtain such machines.

An alternative approach is to use multiple sources of physical randomness under the limited influence of multiple observers. For instance, suppose that several observers at an audit – some who represent the candidates, others who are independent – each roll two translucent ten-sided dice (of different colors), and the digits are combined to form a random number. If the dice are not perfectly fair, the number may not be perfectly random. However, if the dice are "fair enough" that the observers cannot detect any bias, then any departure from randomness – even if

[40] For instance, one audit procedure mandated listing all the municipalities in a county in alphabetical order, randomly choosing a municipality to start with, and then selecting one precinct per municipality in sequence until the audit quota was filled. Of course, this meant that large cities and small villages received roughly the same attention in the audit although most voters were in the cities. ("Biased" is a statistical term of art; we do not think that anyone consciously intended to favor village votes over city votes.)

[41] The most efficient batch comparison methods have a greater chance of selecting batches that could contain more error than others. Colloquially, this fact is a feature of these methods, not a bug; it cannot be exploited to undermine the effectiveness of those methods.

it is known – may well be difficult to exploit. Because multiple observers participate, none of them can control the outcome without the collusion of all the rest. (Combining this approach with the PRNG method described next can provide additional protection against influence.)

For audits with relatively small samples, ping-pong balls or die rolls may suffice. For larger samples, such methods can be used to provide a seed – loosely speaking, a starting point – to a pseudo-random number generator (PRNG) that uses a publicly known, reliable computational algorithm. Given a particular seed, a PRNG always produces the same numbers in the same sequence. In this respect, it is not "random" (hence the "pseudo" in the name) – and this feature is in fact useful for observability. If an audit team announces that it used, say, the well-known SHA-256 hash algorithm in a particular way[42] with a seed of 592907385202 to produce a sequence of 78 numbers, everyone who can use (or write) a program that implements SHA-256 can reproduce the sequence. A good PRNG has the quality that a small change in the seed produces a large change in the sequence, greatly complicating any attempt to influence the sequence.[43] The PRNG based on SHA-256 is a very good way to produce 256-bit pseudo-random numbers (that is, numbers with up to 78 decimal digits). Longer seeds are preferable to shorter ones.

Ping-pong balls, dice, and PRNGs produce random (or pseudo-random) numbers; these numbers need to be "translated" to particular batches or ballots. The computational details are a bit too dense for this white paper, but have been described elsewhere.[44] Auditors and observers need a ballot manifest that explains where each ballot that bears the contests being audited can be found. The ability to produce and distribute this ballot manifest is an important prerequisite of the audit. (As we sketch below, a similar requirement applies to batch comparison audits.)

For instance, suppose that all the ballots for an election are contained in three ballot boxes: A, B, and C. The ballot manifest might report that box A contains 400 ballots, box B contains 250 ballots, and box C contains 500 ballots – making 1150 ballots in all. We can say, then, that A contains ballots 1 through 400; B contains ballots 401 through 650; and C contains ballots 651 through 1150. (These numbers are not serial numbers printed on the ballots.) Now, for each ballot to be audited, a random number is converted to a ballot number between 1 and 1150; each

[42] Ron Rivest has proposed a simple and robust method. His method is implemented in an interactive Javascript web page at http://statistics.berkeley.edu/~stark/Java/Html/sha256Rand.htm, and in the "audit tools" page http://statistics.berkeley.edu/~stark/Vote/auditTools.htm. Rivest's own Python implementation, with extensive explanatory comments, is at http://people.csail.mit.edu/rivest/sampler.py.

[43] For instance – to illustrate the idea – suppose that someone attempts to subvert an audit by tampering with the results in low-numbered precincts and loading the dice so that they disproportionately produce high digits (7, 8, and 9). If a good PRNG is used, this attack fails because a high-digit seed is no more likely than a low-digit seed to produce a sequence with high numbers. The number of "likely" sequences is reduced, but that does the attacker little good because the likely sequences have very little in common.

[44] Arel Cordero, David Wagner, and David Dill, "The Role of Dice in Election Audits – Extended Abstract," June 26, 2006, available at http://www.cs.berkeley.edu/~daw/papers/dice-wote06.pdf, explains a general approach of mapping from random numbers to ballot batches. Mark Lindeman and Philip Stark, "A Gentle Introduction to Risk-Limiting Audits" (*IEEE Security and Privacy*, forthcoming; preprint of March 16, 2012, available at http://statistics.berkeley.edu/~stark/Preprints/gentle12.pdf) explains the use of ballot manifests.

number is equally likely to be selected.[45] Suppose one such number is 443: This number corresponds to the 43rd ballot in box B. Provided that the auditors can identify and retrieve the 43rd ballot in box B – and so on for the other ballots in the sample – they should obtain a fair simple random sample of all the ballots. (This is a sample with replacement: Each ballot can be selected more than once. If so, the ballot is only inspected once, but it counts as many times as it is selected.)

For efficient batch comparison audits, the procedure is broadly similar but more intricate. In addition to enumerating ballots, the manifest enumerates the error bounds – a measure of how much miscount possibly could be found in each batch, which depends on the number of ballots and the vote counts.[46] Before the sample is drawn, each batch is assigned part of the range from 0 to 1 based on its error bound: The more error it could contain, the larger its share of the range. To identify the sample, each random number is converted to a fractional number between 0 and 1, and the batch whose range contains that fraction is selected.[47] (Again, a batch can be selected more than once, but is only audited once. However, any errors found in a batch "count" as many times as the batch is selected.) The details can be many depending on the numbers of batches, contests being audited, and candidates in those contests – but they all boil down to simple arithmetic operations, and can be computed in software *and checked by observers*.

The preceding discussion assumes that all the random sampling for an audit is conducted at a single central site. If it is preferred for multiple jurisdictions to draw random samples independently, the details become more complicated, but not intractably so. (See the discussion of multiple jurisdictions below.)

How large is the audit? When can it end?

There are two basic strategies for determining when a risk-limiting audit can end. The most common one is often called auditing in stages[48] (although usually there is only one stage); the alternative is a sequential audit.

To audit in stages, the auditors select an initial sample size large enough that if that sample shows a sufficiently low error rate, the risk limit will be attained and the audit can stop. This sample size largely depends on the smallest percentage margin (as a proportion of ballots cast: the diluted margin). It can also be adjusted for various degrees of error tolerance: auditing a somewhat larger sample in the first stage may allow the audit to reach the risk limit in one stage

[45] In this example, one could simply divide the random number by 1150, take the remainder, and add 1. For practical purposes, if the random number is sufficiently long, this gives each ballot number an equal chance of selection.
[46] In one simple approach, the error bounds are reckoned in net votes. For instance (in a two-candidate contest), if a batch contains 50 ballots and the apparent winner received 7 more votes in that batch than the apparent loser, the error bound would be 57: the difference between "winning" the batch by 7 votes and losing it by 50 votes.
[47] For instance, if precinct 1A's error bound was 0.12% of the sum of error bounds, and precinct 1B's error bound was 0.26% of the sum, then precinct 1A might be assigned the range from 0 to 0.0012 (because 0.12% = 0.0012), and precinct 1B could be assigned the range from 0.0012 to 0.0038 (because 0.12% + 0.26% = 0.38% = 0.0038). Thus, if the random fraction is 0.002432…, then precinct 1B would be selected.
(Ranges cannot overlap; in this example, if the random fraction is exactly 0.0012, then precinct 1B would be selected.)
[48] Sometimes it is called a "staged audit," but we try to avoid that double entendre!

even if errors are found. These sizes may be determined through simple formulas (for examples, see [1a] and [2] in Appendix 2), or even pre-calculated in tables. After examining the first-stage sample, typically the auditors do a computation with (or apply a rule to) the results to see if the risk limit has been attained and the audit can stop.[49] The details of such computations are outside the scope of this white paper, although many of them are very simple; they depend on details of the method being used.[50]

If the risk limit has *not* been attained after the first stage, then the auditors can proceed with a second stage, and possibly additional stages after that. Proceeding to additional audit stages often is called escalation. At any time (at least in principle) the auditors can switch to a full hand count, if additional stages seem unlikely to end the audit efficiently.

In a sequential audit, instead of choosing an initial sample size, the auditors simply choose one ballot or batch at a time, audit it, and then determine (using an appropriate computation or rule) whether the risk limit has been attained or more counting is needed. Again, the auditors can switch to a full hand count at any time. (One can think of a sequential audit as an audit by stages, where each stage is very small and the number of stages could be large.)

Auditing in stages generally works well in batch and ballot comparison audits, because the necessary sample sizes are relatively predictable when error rates are low. That said, there must be some plan for the possibility of escalation. Current (not risk-limiting) audit requirements often have *ad hoc* escalation rules. For instance, New York's audit regulation (6210.18) calls for up to three audit stages – comprising 3% of voting systems, an additional 5% of systems, and an additional 12% of systems – plus a possible full hand count, based on errors found at each stage. Whatever the merits of such rules, fixed-percentage escalation is not an efficient way to conduct risk-limiting audits. If the first stage comes very close to attaining the risk limit, there is no need to require an equally large or larger second stage. Contrariwise, if the first stage finds large errors that clearly raise doubts about the outcome, it may be more reasonable to conduct a full hand count than to continue checking smaller samples.

It is probably best to allow election officials to exercise discretion about how large a second stage (if any) should be, given the results from the first stage, in order to afford a good chance of attaining the risk limit without undue complication or expense. The crucial rule is that counting cannot stop until either the risk limit has been attained, or all the votes have been counted by hand.

The timing of the audit must allow for the possibility of additional stages, including any necessary coordination among jurisdictions and any public notice requirements. Further discussion comes below.

[49] In some cases no computation is needed; a simple rule suffices. For instance, if the original audit size was chosen to accommodate a certain number of errors, and no more than that number of errors was found in the audit sample, then the audit can stop.

[50] A Gentle Introduction to Risk-Limiting Audits" (see footnote 35) discusses some of these methods.

For a ballot-polling audit – especially in one jurisdiction – a sequential audit may be best, because ballot-polling audits are less predictable than comparison audits when error rates are low. The auditors could choose an initial sample size that is very likely to attain the risk limit (if the original count was accurate), but this approach may require much more work than is actually needed. However, it is perfectly feasible to conduct ballot-polling audits in a series of stages where very little time is needed between stages. For instance, auditors could choose to audit 100 ballots in a first stage, see that they are getting close to the risk limit, and immediately decide to audit another 20 ballots to see if that suffices – then switch to auditing ballots one at a time as in a sequential audit.

The specific computations or rules for determining whether audits must continue can be handled in various ways. A web page maintained by Philip B. Stark[51] provides various tools for ballot-level comparison audits, including a "calculator" for whether more ballots must be sampled; the calculation is simple enough to be done by hand or with a four-function calculator. (Stark is preparing another web page with similar tools for ballot-polling audits.) For its batch-level comparison audit pilot discussed in the next section, Cuyahoga County, Ohio, created a Microsoft Excel spreadsheet to automate the computations of whether the audit could end. Audits in Boulder County, Colorado in 2008 and 2010 used Neal McBurnett's open-source ElectionAudits software. Many other approaches are possible.

We discourage stating particular computations or escalation rules in law, because these may "lock in" particular risk-limiting audit methods in place of more efficient alternatives developed later. However, the computations should be stated in rule or formal procedure subject to public comment before the audit, both to facilitate public observation and to allow them to be vetted for validity.

Specifying audit procedures

Our previous remark about whether and how to codify computations and escalation rules gets at a fundamental design issue: how to foster predictability, observability, and verification without locking in requirements that do not suit all jurisdictions or that impede innovation and improvement.

In general, we favor a model of partial delegation. For instance, in New York, the legislature enacts laws, and authorizes or explicitly requires the State Board of Elections to create formal regulations consistent with the laws. The state board, in turn, can allow or explicitly require county boards to create written procedures consistent with state law and regulations. Either the legislature or the state board can state outright mandates where necessary; offer "safe harbor" procedures in other cases, allowing county boards to propose alternatives subject to approval; and delegate reasonable discretion to county boards, provided that their procedures are clearly documented and subject to public comment. Like any policymaking model, this one does not work perfectly – and we might change some specifics of New York's implementation. Even so, partial delegation offers the most scope for dedicated public officials and concerned citizens to promote continual improvement in election processes.

[51] "Tools for Comparison Risk-Limiting Election Audits," http://statistics.berkeley.edu/~stark/Vote/auditTools.htm

As an example of "safe harbor," Ohio's current post-election audit directive, promulgated by the Secretary of State, describes procedures for a 5% manual audit, but explicitly allows counties to propose alternative procedures, including risk-limiting procedures.[52] The directive thus enables and tacitly supports innovations at the county level, such as Cuyahoga County's pilot testing of risk-limiting audits in 2011.[53] Although we hope that eventually Ohio will *require* risk-limiting audits at least in some contests, allowing procedural variation and innovation with accountability is a valuable approach.

[52] Directive 2012-12, "Post-Election Audits," February 24, 2012, available at http://www.sos.state.oh.us/SOS/Upload/elections/directives/2012/Dir2012-12.pdf.

[53] For information on Cuyahoga County's pilot, see the Board of Elections' summary at http://boe.cuyahogacounty.us/pdf_boe/en-US/2011/NOV11GENAuditSummary.pdf and the detailed report at http://boe.cuyahogacounty.us/pdf_boe/en-US/2011/NOV11GENAuditReport.pdf.

Timing, Coordination, and Integration with Recounts

As we noted earlier, risk-limiting audits can be tricky to integrate with existing elements of election law. In many cases it may be best to alter existing laws to accommodate and leverage risk-limiting audits. Because election laws vary so much across the nation, we offer points to consider rather than definitive recommendations.

Auditing before finalization of results; coordinating multiple jurisdictions

In order for a risk-limiting audit – or any audit – to correct an election outcome, it must be completed before the official results are finalized. Where candidates have recourse to request recounts or to challenge election results in other ways, audits should be completed before candidates have to decide. In many states, recount requests and other challenges can begin after formal *certification* of the election results. In those states, risk-limiting audits should be completed prior to certification if at all possible.

Conducting a risk-limiting audit before certification may seem to verge on self-contradiction: The audit checks the outcome based on the voting system counts, yet the voting system counts are not officially determined until certification. As a practical matter, it may be easy to conduct a risk-limiting audit before certification – but not so easy if some votes are being counted late, multiple jurisdictions have to coordinate, and the audit may expand into multiple stages. Several approaches can ameliorate those complexities.

In general, risk-limiting audits depend on unofficial final results – vote totals and subtotals that are believed to be complete and correct, subject to the audit and, perhaps, recounts and other challenges. (However, it is possible for the audit to begin before all ballots are counted.) Many jurisdictions do not presently produce unofficial final results – subtotals, in particular – prior to certification, but could do so with straightforward procedural changes. Thus, adding an intermediate deadline for unofficial final results, prior to certification, will facilitate audits in some jurisdictions.

A risk-limiting audit needs to take all ballots into account, but in some cases it can begin before all ballots have been counted. If almost all the ballots have been counted – particularly if the number of uncounted ballots is small in comparison with the apparent margin – then an audit can begin among all the other ballots, provided that voter privacy is not compromised.[54] The simplest approach is to make a worst-case estimate of how many ballots may be pending, make the worst-case assumption that all of them will contain votes for the apparent runner-up, and subtract them from the apparent margin. In many cases, those worst-case assumptions will have little effect on the workload.

[54] If, for instance, a jurisdiction were to publish unofficial final results by precinct, and then to add small numbers of absentee ballots to the precinct totals, then it could be possible to determine how some absentee voters had voted. Some jurisdictions presently avoid this problem by reporting all absentee votes in one or more batches separate from the precinct results.

If many ballots (for instance, mail ballots) are counted well after Election Day, the worst-case approach just described is not feasible. Nevertheless, if it is desirable to begin auditing the Election Day ballots before the late ballots are counted, this can be done using a stratified audit[55] – perhaps beginning with a modestly conservative estimate of the final margin, and making adjustments if necessary.

When multi-jurisdictional contests are to be audited beginning prior to certification, the counties or other jurisdictions involved must be able to collate their pre-certification results. In the conceptually simplest case, each jurisdiction would provide detailed results (batch-level ballot counts and vote subtotals, or vote totals and ballot manifests as explained earlier) to a state election office; then the state election office would randomly select the first-stage sample, under public observation, and would immediately notify each jurisdiction which ballots or batches to audit. However, if it is difficult to combine the detailed results, or if it is preferred to draw random samples at the jurisdictional level (perhaps so jurisdictions can begin their audits on different days), an alternative is feasible. The jurisdictions only need to provide vote totals; the state election office only needs to tell each jurisdiction how large a first-stage sample to select.[56] Subsequently, of course, the jurisdictions must promptly report their audit results in sufficient detail for state officials to determine whether additional auditing is required.

It may be reasonable to stipulate that only the *first stage* of a risk-limiting audit must be complete before the certification deadline. If voting systems are performing as intended, only the first stage should be necessary; if they are not, then it is appropriate to take additional time to investigate what has gone wrong. Of course, the timeline for continuing the audit must be harmonized with other provisions of election law.

Recounts and election contests

Risk-limiting audits can greatly reduce the need for recounts and election contest litigation. (Strong evidence that the voting system outcome is correct often constitutes strong evidence that there is no point in demanding a recount!) Therefore, state laws and procedures governing recounts and legal actions to contest election results should be modified to allow candidates and the public to wait for completion of the risk-limiting audit before filing for a recount or election contest.

States that set a statutory deadline for filing an election contest or requesting a recount will need to amend statutes accordingly. For example, if the deadline to file is within five days of the completion of the official canvass (California), then the law should be amended to allow for filing within five days of the completion of any risk-limiting audit, but not before the audit is completed.

---

[55] Some of the relevant methods are discussed in Michael J. Higgins, Ronald L. Rivest, and Philip B. Stark, "Sharper *p*-values for Stratified Post-Election Audits," *Statistics, Politics, and Policy* 2, 1 (2011). Elaboration would be needed for the case where the audit begins before all the results are known.

[56] This could be done using either a two-stage sample (where the state election office draws a random sample to determine how large each jurisdiction's sample should be) or a stratified sample where each jurisdiction's sample size is directly based on its vote totals. (The stratified approach involves more complicated mathematics; see footnote 44.)

In states that presently have automatic recounts for close vote margins, it may be desirable to lower some automatic thresholds (making automatic recounts less common) in contests subject to risk-limiting audits. As a separate matter, automatic recounts in these contests might be replaced by *discretionary* recounts (perhaps subject to a fee, perhaps not) that can be requested within a specified period *after* the audit is completed. Judgments about these issues are likely to vary from state to state, depending in part on the size of the state. For instance, a contest with a statewide margin of 0.5% or even 0.25% is relatively easier to audit in a large state than in a small state, so large states may prefer lower automatic recount thresholds.

On the other hand, states that currently do not have automatic recounts for close margins may wish to adopt them for contests subject to risk-limiting audits. Unavoidably, for unusually close contests, the random sampling required to confirm the election outcomes may amount to, or closely approach, a full hand count – and, in some cases, may actually be more difficult.[57] Moreover, if a state's recount law contains provisions such as ballot challenges that are not also contained in the audit law, then it may not be possible to use the audit results in the recount. Thus, it may be expedient to proceed directly to a recount in the closest contests.

The audit procedures and timing should be integrated with recount law to gain efficiencies where possible. For example, if it is feasible to use the same standards and procedures for evaluating votes in an audit as in a recount, then any votes counted during an audit do not need to be hand counted again in the recount. (However, these votes may be subject to additional candidate challenges during the recount.)

[57] For instance, many jurisdictions might prefer to hand-count 1000 ballots rather than to randomly sample and interpret (say) 900 of the ballots. Similarly, in a batch-level comparison audit, a jurisdiction may prefer to hand-count ballots from 1000 precincts rather than randomly sample and count ballots from 900 precincts.

Moving Ahead with Risk-Limiting Audits

Risk-limiting audits offer an important step forward in election verification. In combination with other best practices, they provide a rigorous, flexible, efficient approach to confirming election outcomes. They help election officials and citizens to check how well the voting systems are functioning, and to make informed improvements. As with other innovations, careful planning and implementation – with judicious use of pilots and/or phased implementation – will help to reap the full benefits of risk-limiting audits and to minimize costs.

Risk-limiting audits have some outright requirements, and some desiderata that can make them much simpler and/or more effective. A jurisdiction should consider these criteria in deciding how quickly and how widely to implement risk-limiting audits.

- Paper ballots or a voter-verifiable audit trail. An audit (or recount) that lacks such an audit trail cannot provide strong evidence that election outcomes are consonant with voters' intentions. [Pages 20-21.]

- Evidence that the audit trail was preserved inviolate. The reliability of the audit trail should be demonstrated, not assumed. [Pages 21ff.]

- (For batch-level comparison audits:) Timely, convenient reporting of auditable "batch" subtotals. It may be necessary to upgrade election management systems in order to report vote totals for auditable batches[58], in a data format convenient for auditing (especially when audits are coordinated across jurisdictions). In some cases, ballot handling and storage procedures should be revised.

- (For ballot-level comparison audits:) Cast Vote Records matchable to individual ballots. The following may be necessary or helpful:

    - Upgrade voting hardware to add identifying numbers to ballots as they are cast, so that ballots can be matched to Cast Vote Records without compromising voter privacy. (Alternatively, implement a transitive audit using a secondary system that achieves the same goal [page 10].)

    - Upgrade election management systems to report Cast Vote Records, including ballot identifiers. (The Cast Vote Record may be published, or handled in some other way to address the voter privacy concerns discussed earlier)

    - Revise ballot handling and storage procedures to facilitate retrieving the particular ballots selected for audit.

[58] "Auditability" is relative. Some jurisdictions presently report the subtotals for many thousands of absentee votes in just one absentee "batch." Certainly it is possible to audit the absentee ballots by hand-counting them all, but dividing the absentee ballots into smaller batches can greatly reduce the auditing burden.

- An appropriate timeline for completing audits. For instance, it may be reasonable to add a new deadline for reporting unofficial totals, and/or to delay certification deadlines to provide more time to finish audits. The timeline should appropriately accommodate the possibility that the audit will uncover problems that require additional auditing or counting. [Pages 36ff.]

- New procedures for reporting and combining unofficial totals and subtotals. In order to coordinate audits over multiple local jurisdictions, it may be useful for local jurisdictions to send unofficial final results to central election officials in a common data format.

- Procedures for managing the "math" of risk-limiting audits and determining when an audit can end. As described above, individual jurisdictions can use various software tools – or, in some cases, even hand calculations – to determine whether an audit can end [page 34]. If a central election office is coordinating audits across multiple jurisdictions, then jurisdictions must supply audit results to the central office (in an appropriate format) so that it can make that determination.

- Procedures for collating and reporting audit results: The complete results of an audit should be reported and interpreted [page 25].

Conceptually, none of these points is insuperably complicated or difficult. Nevertheless, jurisdictions may find it impractical to address some of them in the near future. Moreover, depending on the scope of the changes to be made, it may be prudent to spread them over time, rather than implement many changes within one election cycle.

Pilot audits, such as the program authorized in California under AB2023 (see page 8), can confer several advantages. Pilots allow for relatively inexpensive experimentation with audit procedures. This hands-on experience can reveal what procedures work well and what should be revised or improved before audits are implemented more widely. Pilots can effectively integrate local election officials into the process of audit design, promoting better design decisions and stronger buy-in from the officials.

Phased implementation of risk-limiting audits – gradually broadening their scope – offers further advantages. Risk-limiting audits in selected statewide contests would complement pilot audits, which typically start on a smaller scale. As we saw earlier, risk-limiting audits – batch-level comparison audits in particular – are *relatively* more efficient (i.e., typically require smaller percentages of ballots to be audited) in large contests than in small contests. Auditing one or a few statewide contests allows local jurisdictions to start risk-limiting audits with minimal burden, and allows the state to experiment with coordination across jurisdictions, without waiting upon changes that may be needed for ballot-level comparison audits. Moreover, because statewide contests often are especially important to voters, risk-limiting audits of statewide contests can immediately bolster public confidence and generate support for more extensive audits.

Prudent initiatives to implement risk-limiting audits now will offer immediate benefits for election administration now, and will yield even greater returns in the future.

Appendix 1: Schematic summary of risk-limiting audits

Comparison audits and ballot-polling audits follow somewhat different paths, as sketched below. They share a basic iterative logic:

- Randomly select and examine some ballots.

- Determine whether those ballots (plus any ballots previously examined) provide strong evidence that the originally reported outcome is correct. If so, stop the audit and accept that outcome.

- If the ballots examined so far do *not* provide strong evidence that the originally reported outcome is correct, either return to the first step – i.e., randomly select and examine more ballots – or proceed to a full hand count. The audit continues until it finds strong evidence for the original outcome, or until all ballots have been counted by hand.

All audits

- o Before the election, determine which contests will be audited, at what risk limit(s). (Some contests may be selected for auditing after the election, as discussed at page 19.) Establish the rules used to interpret valid votes, and the statistical method(s) and rules used to determine whether an audit can stop as well as whether it should proceed to a full hand count.

- o After the election but before the audit begins, produce a ballot manifest that details where all the ballots (or audit records) can be found.

- o Through a compliance audit or similar procedure(s), check that the audit trail has been preserved inviolate. If the audit trail has been compromised, respond appropriately (see page 21).

Comparison audits

1. Commit to vote totals (vote counts for each candidate or position), and to vote subtotals for each batch or ballot. The subtotals must sum to the totals. Publishing the totals and subtotals is the conceptually simplest commitment method. For ballot subtotals (i.e., Cast Vote Records), it is possible to use an alternative that permits public observation of the audit (see page 28).

2. Typically, determine an initial sample size. Generally this initial size is such that, if no more than an expected small number or rate of errors is observed in the sample, the risk limit is attained and the audit can stop, certifying the outcome.[59]

---

[59] It may be feasible, instead, to conduct the audit sequentially – randomly selecting and auditing one batch or ballot, then deciding whether to select and audit another.

3. Under public observation, select a random sample of batches or ballots – the initial sample, or an additional sample to be audited as determined at step 5. In a sequential audit, the sample typically comprises just one batch or ballot. (If the audit uses a pseudo-random number generator [PRNG, see page 31], the random seed used with the PRNG only needs to be chosen once; additional samples can be selected by using the PRNG to generate more numbers from that seed.)[60]

4. Retrieve the batch(es) or ballot(s) in the sample selected in step 3; inspect the ballot(s) and determine the valid votes. For each batch or ballot, compare the vote subtotals from the audit to the corresponding voting system subtotals. (The specific comparison depends on the statistical method used, but generally is based on the number and kind of discrepancies between the voting system subtotals and the audit subtotals.)

5. Based on the comparisons in step 4, determine whether (a) the risk limit has been attained and the audit can stop, accepting the voting system outcome; (b) additional batches or ballots selected at random should be audited; or (c) a complete hand count should be undertaken. If additional random auditing is needed (b), then repeat steps 3 through 5 until either the risk limit has been attained or all the ballots have been counted by hand. (The size of any additional sample depends on the audit procedure being used.)

Ballot-polling audits

In a *sequential* ballot-polling audit:

1. Randomly select one ballot; examine it to determine which candidate or position, if any, it bears a valid vote for.

2. Based on the determination in step 1 (and previous determinations), compute a test statistic. (Typically, this test statistic initially equals 1.0; each time through step 2, the current value is multiplied by some number, which depends on the valid vote if any, to find the new value of the statistic.)

3. Based on the statistic in step 2, determine whether (a) the risk limit has been attained and the audit can stop, accepting the voting system outcome; (b) at least one additional ballot selected at random should be audited; or (c) a complete hand count should be undertaken. (The audit procedure may specify some maximum number of ballots to examine before proceeding to a complete hand count.) If additional random auditing is needed, then repeat steps 1 through 3 until either the risk limit has been attained or all the ballots have been counted by hand.

A *batch* ballot-polling audit has essentially the same steps as a sequential ballot-polling audit, except that each random sample (step 1) may contain arbitrarily many ballots. The audit procedure specifies how to decide the number of ballots in each sample.

[60] If a contest involves multiple jurisdictions, several approaches to random sampling may be possible (see page 34, including footnote 56).

Appendix 2: Methods used in estimates of audit counting

The batch-level estimates are based on data from California's 2008 election archived at the Statewide Database hosted by the University of California at Berkeley, at http://swdb.berkeley.edu/d00/g08.html. Specifically, the analysis here uses the statewide "SOV by svprec" file at http://swdb.berkeley.edu/pub/data/G08/state/state_g08_sov_data_by_g08_svprec.zip (in DBF format). Most jurisdictions report each precinct's Election Day and vote-by-mail counts as separate "svprec" returns.[61]

The batch-level estimates consider only the number of votes cast in each batch, plus the reported margin of victory in a given scenario (ranging from 0.5% to 30%). The maximum possible net miscount (error bound) in each batch is imputed from the margin. For instance, for a 10% margin, it is assumed that each batch reported vote shares of 55% for the apparent winner and 45% for the apparent loser, so the error bound would equal 110% of votes cast in the batch. (That is the difference between winning the batch by 10% and losing it by 100%.) Thus, the error bound *proportion* is simply 100% plus the reported margin. (The error bound is measured in votes, rounding up if necessary.) The number of draws in the sample is estimated using a formula from Aslam, Popa, and Rivest[62]:

$$\text{ceiling}(\ln \alpha / \ln (1 - M / E)) \qquad [1]$$

where α is the risk limit (here 0.10), M is the reported margin in votes, and E is the sum of error bounds (the total maximum possible net miscount). ("ln" means natural logarithm, a common function available on scientific calculators and in spreadsheets; "ceiling" means always to round up.) Because the sum of error bounds depends on the margin, this formula can be rewritten as

$$\text{ceiling}(\ln \alpha / \ln (1 - m / (1 + m))) \qquad [1a]$$

where m is the reported *proportional* margin (for instance, 10% = 0.1 in the example above). Expression [1a] gives the expected number of draws if the audit finds no miscounts, or if the miscounts are small. Then the number of ballots in the sample is estimated by computing each batch's probability of inclusion in the sample, and multiplying by the number of ballots in each batch.[63]

[61] In this data file, data from Siskiyou County are erroneously reported, such that some units have implausible total vote counts. These were replaced with the total number of valid presidential votes, resulting in a discrepancy of 95 votes between the statewide totals used in this analysis and the official statewide totals. With over 13 million votes cast in the election, the handling of Siskiyou has very little impact on the results.

[62] Javed A. Aslam, Raluca A. Popa and Ronald L. Rivest, "On Auditing Elections When Precincts Have Different Sizes," paper presented at the 2008 EVT/WOTE conference, available at http://static.usenix.org/event/evt08/tech/full_papers/aslam/aslam.pdf (eq. 13)

[63] Batch i's probability of selection in one draw is u_i / U, where u_i is the error bound in batch i and U is the sum of error bounds. Thus, if d is the number of draws in the sample, as given by formula [1a], batch i's probability of inclusion in the sample is $p_i = 1 - (1 - u_i / U)^d$. The expected number of ballots to audit is therefore $\sum b_i p_i = \sum b_i (1 - (1 - u_i / U)^d)$.

The number of draws in the ballot-level comparison audit is estimated simply as

$$\text{ceiling}(6.2 / m) \qquad [2]$$

where *m* is the reported proportional margin, as in [1a]. The value 6.2 (which assumes a risk limit of 0.10) would allow the audit to terminate if the initial sample finds no more than one one-vote overstatement, i.e., an error whose correction reduces the margin by one vote. ("A Gentle Introduction to Risk-limiting Audits," cited in footnote 35, provides further explanation, and shows how to easily calculate values other than 6.2 to accommodate other kinds and amounts of errors.) The number of *ballots* in the sample is estimated in the same way as for batch-level audits, except that each ballot is a batch. The number of ballots differs from the number of draws only for the smallest margins in Amador County.

In the ballot-polling audit analysis, which follows the approach in "A Gentle Introduction" (AGI) (but is slightly more conservative), the two vote shares are set to total 1 (i.e., 100%) given the reported margin. Where the apparent winner's vote share is *W* (equivalent to *s* in AGI) and the tolerance is set at 0.001, the estimated number of draws is

$$\frac{\ln 9.9 + \ln\left(\frac{W-0.001}{0.5}\right)/2}{\ln\left[\left(\frac{W-0.001}{0.5}\right)^{W}\left(\frac{1-(W-0.001)}{0.5}\right)^{1-W}\right]} \qquad [3]$$

where 9.9 is a threshold value corresponding to $\alpha = 0.1$ and $\beta = 0.01$, as explained in AGI.[64] The number of ballots (and, therefore, votes) in the sample is computed in the same way as for the ballot-level comparison audit.

Table 1 on the next page provides some selected results for the statewide analysis (corresponding with Figures 1 and 2).

[64] The formula here, with the second ln term in the numerator, is slightly more conservative than the approach in *AGI.*

Table 1: estimated audit burden, California statewide elections

% margin	batch-level comparison audit		ballot-level comparison audit	ballot-polling audit
	batches	ballots	ballots	ballots
0.5%	462	225,900	1,240	216,751
1.0%	232	114,002	620	47,760
1.5%	155	76,292	414	20,789
2.0%	117	57,636	310	11,618
2.5%	94	46,328	248	7,417
3.0%	78	38,456	207	5,146
3.5%	67	33,041	178	3,780
4.0%	59	29,101	155	2,895
4.5%	53	26,145	138	2,289
5.0%	48	23,681	125	1,855
6.0%	40	19,737	104	1,290
7.0%	35	17,272	89	949
8.0%	30	14,806	78	728
9.0%	27	13,326	69	576
10.0%	25	12,340	62	467
12.0%	21	10,366	52	325
15.0%	17	8,393	42	209
20.0%	13	6,418	31	118
25.0%	11	5,431	25	76
30.0%	9	4,444	21	53

Definitions

Auditing in stages: Auditing an initial sample, with the possibility of escalation to one or more additional audit stages depending on the results of the initial sample. Cf. a sequential audit, which has no predetermined "initial sample."

Ballot-level comparison audit: An audit in which individual ballots are sampled and the audit interpretation of each sampled ballot is compared with the voting system interpretation of the ballot.

Ballot manifest: A list that indicates how the ballots (or VVPRs) in an election are organized and where they can be found. For instance, a simple ballot manifest might list the ballot bags used for an election, and the number of ballots in each bag.

Ballot-level audit: An audit in which individual ballots are sampled and inspected. Ballot-level comparison audits and ballot-polling audits are ballot-level audits.

Ballot-polling audit: An audit in which individual ballots are sampled and interpreted, much as in a public opinion poll, seeking strong evidence that the original outcome is correct (e.g., that the reported winner received the most votes). Ballot-polling audits do not depend upon (or use) the voting system interpretations of the ballots in the sample.

Cast Vote Record (CVR): A record of the machine interpretation of the votes cast by a particular voter, possibly containing additional information identifying the ballot corresponding to the CVR, such as a ballot identification number.

Commit(ment) to vote subtotals: The act of reporting vote subtotals indelibly, i.e., so that the subtotals cannot be altered undetectably later. Publishing the subtotals is the simplest commitment mechanism, but may undermine voter privacy, especially if subtotals are reported for all contests on individual ballots.

Comparison audit: An audit in which the voting system count of a batch (or the voting system interpretation of a ballot) is compared with the audit's count/interpretation of the same batch or ballot. Cf. ballot-polling audits, which are *not* comparison audits.

Compliance audit: In general, an audit designed to check whether required procedures have been followed; in this context, a procedure (or set of procedures) designed to check the integrity of the audit trail as a prerequisite for the validity of a risk-limiting audit.

Cryptographic hash: A digital "signature" calculated by applying a cryptographic hash function to a set of data. Small alterations to the data radically alter the hash value; it is extremely difficult to find two sets of data with the same hash. Thus, hashes can be used to provide strong evidence that data such as vote subtotals have not been altered.

CVR: see Cast Vote Record.

Diluted margin: The smallest margin (between any winning and any losing candidate; in the case of a simultaneous audit, in any of the contests being audited) as a fraction of all the ballots subject to the audit. This is called the "diluted" margin because, unlike typical margin calculations, it includes ballots not counted as votes for any candidate. Including those ballots accounts for the fact that the audit might find valid votes where the voting system reported overvotes or undervotes.

Escalation: In an audit in stages, expansion of the audit beyond the first stage (or, more generally, to an additional stage or a full hand count). A decision whether to escalate generally depends on the audit results so far.

Fixed-percentage audit: An audit whose initial sample size is a predetermined percentage of precincts, voting systems, or some other unit.

Fixed-size audit: An audit whose initial sample size is a predetermined number of precincts, voting systems, or some other units.

Outcome: The winner(s) in an election and/or other substantive conclusions (for instance, whether a runoff is necessary and, if so, which candidates will participate).

Pattern voting: Casting votes across multiple contests in a predetermined pattern in order to reveal one's identity. The possibility of pattern voting is a pitfall of publishing Cast Vote Records because it compromises voter privacy.

Risk limit: The largest chance that a risk-limiting audit will stop short of a full hand count when the original outcome is incorrect. The risk limit is chosen before the audit is conducted.

Risk-limiting audit: A procedure for checking a sample of ballots (or voter-verifiable records) that is guaranteed to have a large, pre-specified chance of correcting the reported outcome if the reported outcome is wrong (i.e., if a full hand count would reveal an outcome different from the reported outcome).

Sequential audit: An audit that inspects one ballot or batch at a time, then determines whether another ballot or batch must be audited. In contrast to auditing in stages, sequential risk-limiting audits minimize the number of ballots or batches inspected in order to attain the predetermined risk limit.

Simultaneous audit: An audit of more than one contest using the same audit sample.

SOBA (Secrecy-preserving Observable Ballot-level Audit): A ballot-level comparison audit method in which records of individual votes – not Cast Vote Records for complete ballots – are published, along with cryptographic hashes that allow election officials selectively to demonstrate that particular vote records belong to a particular ballot. SOBA is designed to allow ballot-level comparison audits while protecting voter privacy. For details, see the reference in footnote 38.

Stage: see auditing in stages.

Stratified audit: An audit in which samples are drawn separately from non-overlapping "strata." For instance, each county could comprise a stratum, or ballots cast on Election Day and those cast by mail could comprise strata.

Tiered audit: An audit whose initial sample size is determined from a table based on the reported margin. (For instance, in Oregon, the initial sample size is a percentage of precincts or batches that depends upon whether the margin is less than 1%; at least 1% but less than 2%; or at least 2%.)

Transitive audit: An audit that checks voting system outcomes indirectly, using a secondary system (for instance, software that independently interprets the original ballot scans). If the secondary system outcome matches the voting system outcome, and a risk-limiting audit of the secondary system confirms the secondary system outcome at some risk limit, then it confirms the voting system outcome at the same risk limit. (For background, see footnote 14.)

Unofficial final results: Vote totals and subtotals that are believed to be complete and correct, subject to an audit and, perhaps, recounts and other challenges.

Vote tabulation audit: A post-election audit that checks vote tabulation – i.e., how votes have been interpreted and aggregated into vote totals.

Voter privacy: The attribute that a voter's choices cannot be linked to the voter. Voter privacy can be compromised if ballots can be traced to voters using identifying marks, or if Cast Vote Records can be traced to voters (perhaps via pattern voting).

Voter-verifiable paper record (VVPR): A paper record of a voter's selections that the voter had an opportunity to verify before casting her vote. Ordinarily the term "VVPR" refers to records produced by voting machines, as distinct from hand-marked paper ballots. (Some sources call these "voter-verified" paper records, although the records may or may not have been verified.)

Voting system interpretation: How the voting system – including electronic voting equipment and any other methods and procedures prior to the audit – interpreted the votes on a particular ballot or set of ballots.

VVPAT: A voter-verifiable paper audit trail, typically consisting of voter-verifiable paper records and/or hand-marked paper ballots. (Confusingly, some sources use "VVPAT" to refer to individual voter-verifiable paper records.)

VVPR: See voter-verifiable paper record.