EVIDENCE-BASED ELECTIONS:
CREATE A MEANINGFUL PAPER TRAIL, THEN AUDIT

Andrew W. Appel (Princeton University)
Philip B. Stark (University of California, Berkeley)

EVIDENCE-BASED ELECTIONS

There is no perfect, infallible way to count votes. All methods—including optical scan, touchscreen, and hand counting—are subject to errors, procedural lapses, and deliberate manipulation. Almost all U.S. jurisdictions count their votes using computer-based technology, such as touchscreens and optical-scan machines. Computer-based methods are subject to "hacking", that is, the replacement of legitimate vote-counting software with a computer program that changes (some fraction of) the votes in favor of the hacker's preferred party. Hacking can be performed remotely (even if the machines are supposedly "never connected to the Internet") and it is very difficult to detect. Voters and election administrators see nothing out of the ordinary.

The vulnerability of computers to hacking is well understood. Modern computer systems, including voting machines, have many layers of software, comprising millions of lines of computer code; there are thousands of bugs in that code.[1] Some of those bugs are security vulnerabilities that permit attackers to modify or replace the software in the upper layers; so we can never be sure that the legitimate vote-counting software or the vote-marking user interface is actually the software running on election day.[2]

One might think, "our voting machines are never connected to the Internet, so hackers cannot get to them." But all voting machines need to be programmed for each new election: they need a "ballot-definition file" with the contests and candidate names for each election, and lists of the contests different voters are eligible to vote in. This programming is typically done via removable media such as a USB thumbdrive or a memory card. Vote-stealing malware can

[1] Estimates of software defect rates range from one per thousand lines of code (in high quality commercial products) down to 0.1 per thousand lines of code in extremely high-quality products (this is at the 90th percentile for the software industry). MEL LLAGUNO, SYNOPSYS, INC., 2017 COVERITY SCAN REPORT: OPEN-SOURCE SOFTWARE—THE ROAD AHEAD 16 (2017), https://www.synopsys.com/content/dam/synopsys/sig-assets/reports/SCAN-Report-2017.pdf [https://perma.cc/H8L3-SADH]. Modern voting machines contain software components such as an operating system (e.g., Windows 7 is fifty million lines of code; or Linux is twenty-seven million lines). USB drivers (common on voting machines) are quite large software components and are riddled with insecurities. Dave Tian et al., *SoK: "Plug & Pray" Today – Understanding USB Insecurity in Versions 1 through C*, IEEE SYMP. SECURITY AND PRIVACY (2018). Therefore, we can expect 100 to 1000 bugs per million lines of code; some small portion of these are "exploitable vulnerabilities," that is, an adversary can exploit them to take over the computer and install fraudulent software. A software-based product such as a voting machine can be expected to contain, at any given time, one or more exploitable security vulnerabilities.

[2] NAT'L ACADS. SCIS., ENG'G, & MED., SECURING THE VOTE: PROTECTING AMERICAN DEMOCRACY (2018), pages 89-91.

piggyback on removable media and infect voting machines—even machines with no network connection.[3]

There is a way to count votes by computer and still achieve trustworthy election outcomes. A trustworthy paper trail of voter selections can be used to check, or correct, the electoral outcomes of the contests in an election. "Electoral outcome" means the winning candidates or positions,[4] not an exact numerical tally.

The principle of "evidence-based elections"[5] is that local election officials should not only find the true winner(s) of an election, but they should also provide the electorate convincing evidence that they did. Generally, that means that eligible voters must have had the opportunity to vote, the election must have used voter-verified paper ballots, there must be convincing evidence that those ballots were kept inviolate through the audit, and the reported outcomes must be checked against the paper trail by suitable audits or hand counts.

To have affirmative evidence that reported outcomes are correct requires conducting elections using an auditable voting system, then auditing the results appropriately. First, we discuss auditability, or the creation of a trustworthy paper trail. Second, we discuss auditing—the method for efficiently assessing whether the computer-reported election outcomes are correct, based on the paper trail.

VOTER-VERIFIED PAPER BALLOTS

Society wants evidence that election outcomes are correct (e.g., the candidate actually selected by the voters wins the election), even if the computers have been hacked. The only known practical way to have trustworthy ballots to audit, even if the computer software has been hacked, is to have paper ballots, marked with the voters' choices, that are manually interpretable, accountable, auditable, and recountable.

Hand-marked paper ballots (optical scan)

The traditional method of creating this paper trail (since about 1890 in the U.S.) is the use of a preprinted ballot form that lists, for each contest, the names of the candidates. Alongside each candidate is a target (square, oval, etc.) in which the voter indicates a vote. In recent decades, as such ballots are counted by optical scanners, the voter is asked to fill in an oval or complete an arrow to indicate selections. This is a *hand-marked paper ballot*.

With a hand-marked paper ballot, the marks on the ballot necessarily reflect what the voter did, and we can have reasonable assurance that the human-readable mark on the ballot is for the candidate actually intended by the voter. This assurance increases if the ballot follows standard best-practice ballot-design guidelines, such as those published by the U.S. Election Assistance

---

[3] Ariel J. Feldman et al., *Security Analysis of The Diebold Accuvote-TS Voting Machine*, Proc. 2007 USENIX/ACCURATE ELECTRONIC VOTING TECH. WORKSHOP (2007).

[4] Or, for instance, whether there is a runoff.

[5] Philip B. Start & David Wagner, *Evidence Based Elections*, 10 IEEE SECURITY & PRIVACY 33–41 (2012) (introducing the concept of evidence-based elections).

Commission.[6] Voters are more likely to overlook certain contests on the ballot, to overvote, to undervote, and to make other mistakes if the ballots do not follow these design guidelines.

Hand-marked paper ballots can be quite accurate: in the 2008 Minnesota election for U.S. Senator, of 2.4 million votes cast, only 0.01% (1 in 10,000) was so ambiguous that the State Canvassing Board could not interpret it, and the optical-scan voting machines agreed with the hand-recount totals with an accuracy of 99.99%.[7]

DRE Machines

Direct Recording Electronic (DRE) voting machines have a user interface (typically a touchscreen) and an internal computer. Voters indicate their votes on the touchscreen, and the computer program interprets those indications to add votes to counters in its memory. At the close of the polls, the computer outputs the results by printing them on paper and saving results to a removable-media cartridge.

With a DRE, the record of the vote does not necessarily reflect what the voter did. If the DRE is "hacked," that is, if fraudulent software is installed, then the fraudulent computer program can report arbitrary fraudulent votes. There is no effective paper trail. The close-of-polls printing on paper of the totals is a paper trail that *starts* only when the computer program is reporting the totals. That printout can be effective in auditing the *aggregation* of votes from different precincts, but it cannot serve as a check on the computer program in the voting machine. This is a fatal flaw of paperless DRE voting machines.

In the early 21st century, many states used DRE voting machines. But, because of the widespread recognition of this fatal flaw, only a handful of states use paperless DRE voting machines, and many of those states are transitioning to technologies that have a paper trail starting from the individual voter's ballot.

Voter-verifiable paper audit trail ("VVPAT")

In the 2000s, it was thought that a good solution to the problem of DREs was the VVPAT. For a DRE with VVPAT, the voter indicates choices on a DRE touchscreen. Then, the DRE prints the voter's selections on paper, behind glass. The voter inspects ("verifies") the VVPAT; and the VVPAT serves as the ballot of record in case of recounts or audits.

As we discuss below, VVPAT is not an adequate solution: in practice, the vast majority of voters do not verify the paper printout—it is "voter verifiable" but not "voter verified"; and the few who do inspect the VVPAT cannot safeguard the votes of their fellow voters who do not.

Ballot-marking devices (BMD)

Ballot-marking devices have a user interface (typically a touchscreen) on which voters indicate their selections; then the BMD prints a paper ballot that will be optically scanned. There

---

[6] U.S. ELECTION ASSISTANCE COMM'N, EFFECTIVE DESIGNS FOR THE ADMINISTRATION OF FEDERAL ELECTIONS (2007), https://www.aiga.org/globalassets/migrated-pdfs/eac_effective_election_design [https://perma.cc/G7H3-RDUL].

[7] Andrew Appel, *Optical-scan voting extremely accurate in Minnesota*, FREEDOM TO TINKER (Jan. 21, 2009), https://freedom-to-tinker.com/2009/01/21/optical-scan-voting-extremely-accurate-minnesota/ [https://perma.cc/4FWX-6GUJ].

are many variations of this technology: the paper ballot may have only human-readable marks or the votes may also be encoded in barcodes. The paper ballot might print a summary of the voter's selections, or also contests the voter skipped. The paper ballot may be displayed under glass; it may be ejected for the voter to hold and inspect before feeding back into a slot for scanning; or it may be ejected for the voter to carry to a separate optical scanner.

None of these designs is trustworthy. Just like DRE VVPATs, a BMD vote record might not reflect what the voter did. BMDs print out a paper ballot that is, in principle, voter *verifiable*, but is not, in practice, voter *verified*. In a study of voters using BMDs in an election in Tennessee (2018), DeMillo *et al.* found that 47% of voters did not inspect their BMD-printed ballots at all; the other 53% looked at their paper ballot for an average of 3.9 seconds, not nearly long enough to check that the printout matched what they indicated on the touchscreen for all 18 contests on the ballot.[8] In a controlled experiment with real voters (but not in a real election), Bernhard *et al.* found that when the BMD deliberately misrecorded one vote on each ballot, only 7% of the voters noticed.[9]

If a BMD is hacked and systematically steals 5% of the votes in one contest and only 7% of voters inspect their ballots carefully enough to notice, then the effective rate of vote-theft is $5\% \cdot 93\%$, or 4.65%; this is enough to change the outcome of a moderately close election. The same analysis applies to DRE+VVPAT.

One might think: "not everyone needs to carefully verify their ballots;" if only 7% of voters carefully inspect their ballots, they can serve as a kind of "random audit" of the BMDs. But this sentiment fails to hold up under careful analysis.[10] If and when a voter observes that the BMD-printed ballot is marked with votes that they did not intend, the voter is supposed to alert a poll worker, who is required to void that ballot and allow the voter to mark a fresh ballot. But this situation does not provide usable *evidence* that the BMD was cheating: the voter might be mistaken or lying.[11]

Therefore, in our hypothetical scenario in which a hacked BMD steals 5% of the votes, and 7% of voters carefully inspect their ballots (and know what to do when they see a mistake), then $7\% \times 5\%$ of voters will alert a pollworker; that is, 1 in every 285 voters will claim their paper ballot was mismarked—if the voters do not assume it was their own error. The BMD would successfully steal "only" 4.65% of the votes.

One might think: "but some voters caught the BMD cheating, red-handed." But nothing can be done. It is a rare election official who would invalidate an entire election because 1 out of 285 voters complained.[12]

---

[8] Richard DeMillo et al., *What Voters are Asked to Verify Affects Ballot Verification: A Quantitative Analysis of Voters' Memories of Their Ballots* (last revised Apr. 13, 2019)(unpublished manuscript)(available on SSRN) https://ssrn.com/abstract=3292208.

[9] Matthew Bernhard et al., *Can Voters Detect Malicious Manipulation of Ballot Marking Devices?*, 41 IEEE SYMP. SECURITY AND PRIVACY (2020).

[10] *See* Andrew W. Appel et al., *Ballot-Marking Devices (BMDs) Cannot Assure the Will of the Voters* (last revised Jan 4, 2020)(unpublished manuscript)(available on SSRN) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3375755.

[11] *See* Matthew Bernhard et al., *Public Evidence from Secret Ballots*, *in* 10615 ELECTRONIC VOTING, 84, 86 (Robert Krimmer et al. eds., Springer 2017); Tyler Kaczmarek et al., *Dispute Resolution in Accessible Voting Systems: The Design and Use of Audiotegrity*, *in* 7985 E-VOTING & IDENTITY 127, 131 (James Heather et al. eds., Springer 2013).

[12] *See generally* Bernhard et al., *supra* note 9.

The gap between *voter verifiable* and *voter verified* makes BMDs unacceptable; hacked BMDs can steal the vast majority of the votes they set out to steal, *before* those votes are recorded onto the paper trail. The same analysis applies to DRE+VVPAT.

All-in-one BMDs

All-in-one BMDs combine the ballot-marking functionality of "pure" BMDs with the scanning/tabulating functionality of optical scanners. In various configurations sold by different manufacturers, the "all-in-one" or "hybrid" BMD may eject the ballot for the voter to inspect before feeding it back into the slot from which it was ejected, or the BMD may display the ballot under glass for voter inspection before retracting it past a scanner.

These machines are even less secure—and less acceptable for use in public elections—than pure BMDs. The same paper path contains both the printer (for marking ballots) and the optical scanner (for scanning ballots). The legitimate software (installed by the manufacturer) presumably will not print additional votes onto the ballot after the voter has inspected it, but hacked software could. The software installed on the BMD has complete control over all the physical functions of the paper path: printing, scanning, and paper transport. Therefore, the hacked computer can print votes on the ballot after the voter's last opportunity to inspect the paper. Even those 7% of voters who carefully inspect their ballots are not safe. The same analysis applies to DRE+VVPAT.[13]

Internet voting

Internet voting cannot be secured by any currently known technology.[14] Even if a cryptographic protocol is used to attempt to create an audit trail, the end-user device (phone, computer, or kiosk) is easily hackable. Thus, the voter may indicate a vote for one candidate, but the vote that is encrypted, authenticated, and transmitted may be for another candidate. End-to-end cryptographic paperless voting protocols are an interesting topic for future academic research, but their security and practicality is not mature enough for use in public elections. These scientific facts are well established;[15] we do not discuss them further here.

Software independence, contestability, defensibility

By 2004 it was recognized by most experts that paperless DREs were subject to a massive security hole: if fraudulent software were installed in them, that software could steal votes without any way to detect or correct the fraud, nor a trustworthy way to recount. In 2008, this understanding was framed in the term "software independence":

> An undetected change or error in its software cannot cause an undetectable change or error in an election outcome.[16]

This notion is essential, but still too weak. It very much matters *who* detects the change, and the consequences of this detection. For instance, if an individual voter (amongst the approximately 7% who carefully inspect their BMD-printed ballots) detects an error, that voter has

---

[13] *See also* Appel et al., *supra* note 10.
[14] NAT'L ACADS. SCIS., ENG'G, & MED., SECURING THE VOTE: PROTECTING AMERICAN DEMOCRACY 92 (2018).
[15] *Id.*
[16] Ronald L. Rivest, *On the Notion of 'Software Independence' in Voting Systems*, 366 PHIL. TRANSACTIONS ROYAL SOC'Y A 1881 (2008).

effectively detected a possible error in the election outcome. The election-outcome error is not *undetectable* and the system is software independent.

But in such a case, the election-outcome error is, for all practical purposes, undetectable and uncorrectable by election officials. Voters cannot *prove* that the votes printed on the paper are not the same as the ones they selected on the BMD. Without any such proof, it would be irresponsible to have a do-over election just on the say-so of a few individual voters.

Appel, DeMillo, and Stark propose the terms "contestable" and "defensible" as more useful in the analysis of voting-system security:[17]

> A voting system is contestable if, when an undetected change or error in its software causes a change or error in an election outcome, the system can always produce *public* evidence that the outcome is untrustworthy.
>
> A voting system is defensible if, when the reported electoral outcome is correct, it is possible to generate convincing public evidence that the reported electoral outcome is correct—despite any malfunctions, software errors, or software alterations that might have occurred.

A voting system based on BMD-marked ballots is neither contestable nor defensible. A voting system based on hand-marked paper ballots, counted by optical scanners and recountable (and auditable) by humans, are both contestable and defensible—provided careful procedures are practiced to check administrative processes, physical chain of custody of the ballots, and other physical security measures. Such procedures are called *compliance audits*.

RISK-LIMITING AUDITS

If there is a trustworthy paper record of the votes—meaning that a full, accurate hand tabulation of the recorded votes would show the true winners—there is a way to check whether the computers misbehaved: count the votes by hand.

That is an expensive prospect, so some states mandate looking at a sample of ballots instead, i.e., auditing. Generally, statutory audits provide no assurance that, if a reported outcome is wrong, the error will be detected, much less corrected.[18]

In contrast, a "risk-limiting audit" (RLA) is any post-election procedure that offers the following statistical guarantee:[19]

> If the reported electoral outcome is wrong, there is a known, pre-determined minimum chance that the procedure will correct the reported outcome.

---

[17] Appel et al., *supra* note 10.

[18] Some officials claim that the statutory audits check whether the machines are working correctly. But machines never work perfectly. The question is whether they worked well enough, in this election, to find the true winner(s). That is the question a risk-limiting audit answers.

[19] Risk-limiting audits have been endorsed by the Presidential Commission on Election Administration, the American Statistical Association, the League of Women Voters, Common Cause, Verified Voting Foundation, and many other organizations concerned with election integrity. RLAs have been piloted dozens of times in eleven U.S. states and in Denmark. They are required by statute in Colorado, Nevada, Rhode Island, and Virginia, and authorized by statute in California and Washington. RLAs were developed in 2007; the first publication is P.B.

The maximum chance that the procedure will *not* correct the outcome, if the outcome is wrong, is the "risk limit." For instance, an RLA with a risk limit of 5% has at least a 95% chance of correcting the reported outcome if the reported outcome is wrong (and no chance of altering a correct reported outcome).

The only possible touchstone for determining the correct outcome and correcting wrong outcomes is the paper trail: an RLA corrects the outcome by conducting a careful, full manual tally of the paper trail. The result of that tally replaces the reported outcome if the two differ.

If the paper trail is trustworthy—i.e., if a full hand tabulation would show who really won—the replacement outcome is the correct electoral outcome, and the overall procedure limits the risk that an incorrect reported outcome will become official. If the paper trail is not trustworthy (for instance, if it has not been kept secure or if it was generated by BMDs), no procedure can limit the risk that an incorrect reported outcome will become official. Indeed, applying an RLA procedure to an untrustworthy paper trail could even replace a correct reported outcome with an incorrect outcome. At best, applying an RLA procedure to an untrustworthy paper trail can check whether tabulation error altered the outcome reflected in the untrustworthy paper trail.

There are many methods for conducting risk-limiting audits. For instance, a full hand count is a risk-limiting audit, with a risk limit of zero. But, by inspecting randomly selected ballots and using appropriate statistical methods, it is possible to conduct risk-limiting audits much more efficiently—when the reported electoral outcome is correct.[20]

COMPLIANCE AUDITS

An RLA procedure that relies on an untrustworthy paper trail, or any audit that purports to ascertain voter intent from an electronic record or from an artifact that the voter did not have the opportunity to check, is "security theater." There is little reason to believe that a full manual tally of such records would reveal the true winner(s). It is therefore crucial to base audits on voter-verified paper records; to ensure that those records include every validly cast vote exactly once, and no others (checking the determination of eligibility, in particular); to ensure that those records remain complete and intact from the moment they are cast through the audit; and to assess the evidence that they are trustworthy. Absent affirmative evidence that the paper trail is a trustworthy record of voter intent—that tabulating it accurately would show who won, according to the intent of every voter who legitimately cast a ballot in the contests under audit, and no others—the audit might be likely to confirm the incorrect outcome or to change a correct outcome into an incorrect outcome.

The process of assessing the trustworthiness of the paper trail is called a "compliance audit." Compliance audits should include the following steps, among others:

Stark, *Conservative Statistical Post-Election Audits*, 2 ANN. APPL. STATISTICS 550 (2008). Since then, there have been extensions for other social choice functions (e.g., proportional representation, *see* P.B. Stark, & V. Teague, *Verifiable European Elections: Risk-limiting Audits for D'Hondt and Its Relatives*, 3 JETS: USENIX J. ELECTION TECH. & SYS. 18 (2014) (for auditing any number of contests simultaneously, for different types of voting equipment, etc.); *see also* P.B. Stark & M. Lindeman, *A Gentle Introduction to Risk-Limiting Audits*, 10 IEEE SECURITY & PRIVACY 42, (2012) (for a general but still somewhat technical introduction); P.B. Stark, *Sets of Half-Average Nulls Generate Risk-Limiting Audits: SHANGRLA* VOTING '20 (forthcoming 2020)[hereinafter Stark, *Half-Average Nulls*] (for the most recent and efficient methods for RLAs).

[20] When the reported outcome is incorrect, the audit is *intended* to have a large probability of requiring a full manual tally, so it generally will not save labor then.

Ballot Accounting. Check that the number of ballots sent to polling places equals the number returned voted, plus the number returned spoiled, plus the number returned unvoted. For systems that print ballots on demand, check that the paper stock (sheets cast, spoiled, and still blank) adds up to the number of sheets sent to the polling place or vote center. Using accountable ballot stock, rather than plain paper, is an important security measure. Check that the number of ballots returned from each polling place does not exceed the number of voters registered at that polling place or the number of pollbook signatures at the polling place. Check that the number of ballots of each style corresponds to the number of ballots of each style reported by the voting system. Ballot counts for this purpose should be based on the physical paper, not on the voting system: the audit needs external touchstones to check the voting system.

Eligibility. Check signature verification on vote-by-mail ballots. Check the disposition of provisional ballots to ensure that all that were validly cast (and no others) were included in the results. Check that each voter received the correct ballot style based on her eligibility. For vote-by-mail ballots, there should be a record of the ballot style mailed to the voter; for in-person voting, this might require recording (e.g., in pollbooks) the ballot style given to the voter. For provisionally cast ballots, this might be more complicated.

Physical chain of custody. Adopt a formal seal-use protocol[21] for the tamper-evident seals on ballot boxes and other important records: use numbered, tamper-evident seals that are hard to forge or bypass, train staff in assessing evidence of tampering, record seal numbers when seals are applied, check seal numbers against records, and much more. Review custody logs.

Check that at least two staff members accompanied the ballots whenever ballots were not locked securely or under surveillance. Review surveillance video of the secure ballot storage facility to ensure there was no unauthorized access to ballots.

Due diligence regarding processes, equipment, etc. Review voting equipment event logs. Review any complaints made by voters or anomalies or problems noted by pollworkers.

Some of these steps are formally or informally part of the canvass procedure in some jurisdictions. Ideally, the Secretary of State would require these steps (and others) to be conducted in a way that is publicly verifiable and would require jurisdictions to publish the results. Protocols around physical seals and physical chain of custody are uneven at best. Before the election, voter registration databases should be scrutinized, and changelogs included. Pre-election "logic and accuracy testing" should include compliance review of the ballot design against EAC usability guidelines,[22] to ensure that voters will understand the ballot and will not inadvertently overlook some contests or mark ballots incorrectly.

---

[21] S*ee generally* Andrew W. Appel, *Security Seals on Voting Machines: A Case Study*, 14(2), ACM Transactions Info. & Sys. Security (TISSEC) 18:1(2011).

[22] *See generally* U.S. Election Assistance Comm'n, *Effective Designs for the Administration of Federal Elections* (2007), https://www.aiga.org/globalassets/migrated-pdfs/eac_effective_election_design [https://perma.cc/W7UW-S5CQ]

Compliance audits should be a standard part of any recount, and not just a precursor to risk-limiting audits. Absent a compliance audit, there is little reason for the public to trust that a recount will find the true winner(s).

EFFICIENT RISK-LIMITING AUDITS

The basic strategy behind current methods for risk-limiting audits begins by acknowledging that the reported electoral outcome might be incorrect, then examines randomly selected ballots until either (a) the evidence is convincing that a full manual tally would confirm the reported outcome, or (b) there has been a full manual tally.

There is more than one way to do this. Two basic building blocks are *ballot-polling* and *comparison*. Both can be conducted by randomly selecting either groups of ballots (*batch-level audits*) or individual ballots (*ballot-level audits*).[23]

Ballot-polling audits are like exit polls, but instead of asking voters how they voted, the audit manually examines randomly selected ballots.[24] If a sufficiently large sample of ballots shows a sufficiently large margin in favor of the reported winner, that is evidence that the reported winner really won.[25] Ballot-polling audits have the advantage of requiring very little of the voting system: just the reported winners and access to the ballots. They also require local election officials to organize the ballots well enough to draw a random sample of ballots.

Comparison audits compare how the voting system tallied groups of ballots to how humans tally the same physical group of ballots. A group might be, for instance, all ballots tallied in a given precinct or by a given machine, which yields a *batch-level comparison audit*. The most efficient comparison audits use groups consisting of individual ballots, which yield *ballot-level comparison audits*. To conduct a ballot-level comparison audit, the voting system must report how it interpreted individual ballots in a way that allows the corresponding physical ballot to be identified and retrieved for manual inspection. Such interpretations are called "cast-vote records" or CVRs. The CVR for a ballot lists the voting system's interpretation of voter intent for each contest on the ballot. Most legacy voting systems cannot report CVRs in a way that the corresponding ballot can be identified and retrieved, but some newer systems have this capability.

One method for conducting a ballot-level comparison audit with a 5% risk limit requires manually inspecting approximately 7/(diluted margin) ballots, unless the audit finds errors in the CVRs. The "diluted margin" is the margin of victory in votes, divided by the total number of ballot cards[26] in the population from which the sample is drawn (which must include all ballot cards cast in the contest, and may include others). For instance, in the 2018 gubernatorial primary in California, Newsom and Cox advanced to the general election. The margin of Cox over Villaraigosa, the runner-up, was 618,215 votes out of 7,060,646 ballots cast, including undervotes.

---

[23] Ballot-level audits tend to require examining fewer ballots in all than audits based on larger batches. Roughly speaking, the number of batches one needs to examine to confirm a contest with a given margin of victory at a given risk limit is about the same, regardless of the batch size. Hence, to attain a given risk limit, an audit that uses batches the size of precincts (say, 500 ballots per batch on average) requires examining about 500 times as many ballots as an audit that uses batches consisting of a single ballot (i.e., a ballot-level audit).

[24] Unlike voters, ballots have to reply, and have to reply truthfully, so ballot-polling audits give strong statistical evidence while exit polls generally suffer from large biases.

[25] How to quantify the strength of the evidence depends on how the sample is drawn, among other things.

[26] A "ballot" often consists of two or more "ballot cards" that contain different contests. Sorting the physical ballot cards into homogeneous groups can greatly reduce the number of cards that must be inspected at random to yield a given number of cards that contain a particular contest.

The diluted margin is thus 618,215/7,060,646 = 8.75%. A ballot-level comparison audit with a risk limit of 5% would have required inspecting approximately 7/0.0875 = 80 ballots selected at random from the entire state (assuming the audit did not find any errors). A ballot-polling audit with a risk limit of 5% would have been expected to examine 443 ballots (assuming that the reported results are correct). For either approach, the amount of work required to justify public confidence in the outcome is *de minimis*.

Most ways of conducting RLAs require a "ballot manifest" describing how ballots are stored. For example, "There are 913 boxes of ballots, numbered 1 through 913. Box 1 contains 301 ballots. Box 2 contains 199 ballots . . . ." It is reasonable to require local election officials to construct ballot manifests routinely—if election officials cannot keep track of how much paper there is and where it is, they are not doing their job. Some counties might not currently organize their paper flow in a way that makes constructing ballot manifests possible.

Ballot manifests should be constructed without relying on the voting system to count the paper; otherwise, we are trusting the voting system to check itself.[27] [28]

RESOURCES FOR RISK-LIMITING AUDITS

Ballot polling requires a ballot manifest and the reported results—the hardware and software requirements are minimal, open-source code exists for all the computations.[29] Batch-level comparison RLAs using precincts as batches generally do not save effort compared to ballot-polling RLAs for typical margins and precinct sizes, but require substantially more "data wrangling." Ballot-level comparison audits require voting systems that can report cast-vote records for individual ballots in a way that allows the corresponding physical ballot to be retrieved, and vice versa; however, most current voting systems do not have this ability. Ballot-level comparison audits also require exporting those CVRs and "committing" to them in a publicly verifiable way.

RLA methods exist for all common social choice functions used in the U.S., including plurality, vote-for-n plurality (e.g., school boards), super-majority, and instant-runoff voting (IRV, aka ranked-choice voting or RCV), as well as proportional representation.[30]

There is a variety of open-source software to select random samples of ballots and perform risk calculations.[31] The most difficult aspect of auditing is logistical: coordinating audits of contests that cross jurisdictional lines. That can be facilitated by well designed software.

---

[27] However, ballot manifests can be augmented by data from the voting system to facilitate audits, provided the audit is designed to take into account the possibility that the voting system data are incorrect. For instance, there are ways to combine cast-vote records with ballot manifests to make it easier to sample ballots that contain specific contests and still ensure that the procedure is an RLA. *See* Stark, *Half-Average Nulls*, *supra* note 19; Michelle Blom et al., *Sets of Half-Average Nulls Generate Risk-Limiting Audits (SHANGRLA)*, GITHUB (last visited Mar. 17, 2020), https://github.com/pbstark/SHANGRLA [https://perma.cc/WN2U-FJGU].

[28] Moreover, common human errors include scanning the same box of ballots twice and failing to scan a box of ballots. Scanner mis-picks and errors resulting from clearing scanner paper jams can also cause the number of actual ballots to differ from the number according to the voting system. Relying on the voting system to construct a manifest would miss such errors.

[29] *See, e.g.*, P.B. Stark, *Tools for Ballot-Polling Risk-Limiting Election Audits*, U.C., BERKLEY: DEP'T STAT. (last modified Feb. 16, 2017), https://www.stat.berkeley.edu/~stark/Vote/ballotPollTools.htm [https://perma.cc/DBS2-8LXB].

[30] *See* Stark, *Half-Average Nulls*, *supra* note 19.

[31] *See, e.g.*, Blom, *supra* note 27; Stark, *supra* note 29; P.B. Stark, *Tools for Comparison Risk-Limiting Election Audits*, U.C., BERKLEY: DEP'T STAT. (last modified Feb. 26, 2020),

In our experience, it takes about two minutes to retrieve a particular randomly selected ballot and transcribe the votes for two or three contests.[32] Additional contests take on the order of ten seconds each per audited ballot. The cost of conducting RLAs seems to be very small compared to the overall cost of holding an election. In Colorado, some local election officials report that RLAs are easier than the statutory audits that RLAs replaced, even though the previous audits had little evidentiary value.

Audit the digital images?

Some vendors are promoting systems that create digital images of ballots. These vendors claim that the images make RLAs easier to perform because fewer (or no) paper ballots need to be inspected. That is incorrect: if a risk-limiting audit relies on images of ballots, it must check that the error in making the images from the voter-verified paper ballots *plus* the error the system made interpreting those images to make cast-vote records is not large enough to cause the electoral outcome to be wrong. It is a mathematical fact that this requires examining at least as many physical ballots as an audit that just compares CVRs to a human reading of the paper ballots, without relying on the digital images.[33]

PRINCIPLES FOR ELECTION INTEGRITY LEGISLATION

Laws to ensure that election results are trustworthy should satisfy a number of principles:

1. Require rigorous physical custody of ballots, and compliance audits, as discussed above. An RLA that relies on an untrustworthy paper record accomplishes little.

2. **Require genuine RLAs**: the procedures and calculations should ensure that whenever an outcome is incorrect, the audit has the requisite chance of leading to a full hand count.[34] That in turn entails a number of things:

   – **The audit must ascertain voter intent manually, directly from the human-readable marks on the paper ballots** the voters had the opportunity to verify. It is not adequate to rely on digital images of ballots, paper printed from an electronic record, barcodes, or other artifacts that are not verifiable by the voter or are not tamper evident; nor is it adequate to re-tabulate the votes electronically, either from

---

https://www.stat.berkeley.edu/users/stark/Vote/auditTools.htm [https://perma.cc/4FUT-8C8X]; pbstark, *auditTools*, GITHUB (last visited Mar. 17, 2020), https://github.com/pbstark/auditTools [https://perma.cc/V8KH-TN4V]; pbstark, *Tools for SUITE Risk-Limiting Election Audits*, GITHUB (last visited Mar. 17, 2020), https://github.com/pbstark/CORLA18/blob/master/code/suite_toolkit.ipynb [https://perma.cc/G6MJ-PA5S]; VotingWorks, *ARLO: Open-Source Risk Limiting Audit Software by VotingWorks*, GITHUB (last visited Mar. 17, 2020), https://github.com/votingworks/arlo [https://perma.cc/9T8S-PZ99].

[32] The process is much faster if serial numbers are printed on the ballots (after the voted ballot has been dissociated from the voter's identity).

[33] *See supra* note 28 for some errors that could result in missing or duplicated images. Moreover, there are demonstrations that scanners can inadvertently alter images in ways that would change the appearance of voter intent, including erasing votes. Expecting digital images to accurately reflect voter intent from every validly cast ballot, exactly once, is wishful thinking, even in the absence of hacking. Of course, hacking the scanners or the image processing software is within the technical ability of many undergraduate computer science students.

[34] The statute should not dictate methods or calculations, only principles. That makes it possible to use improved methods as they are developed or as voting systems are replaced.

images of the ballots or from the original paper. BMD printouts, digital images of ballots, re-printed ballots, and other computer data are not reliable records of voter intent: they can be incomplete, fabricated, or altered, accidentally or maliciously, by software bugs, procedural lapses, or hacking. Statues should prohibit relying on such things for the determination of voter intent. Making this prohibition explicit is important because, as mentioned above, voting system vendors are marketing technology that purports to facilitate RLAs by allowing auditors to examine digital images of ballots instead of paper ballots. Relying on an electronic record created by the voting system to accurately reflect voter intent amounts to asking a defendant whether he is guilty.

- **The audit must take all validly cast ballots into account.** If ballots are omitted from consideration, for instance, vote-by-mail ballots that did not arrive by election night or provisionally cast ballots, the audit cannot be a genuine RLA. Still, there are ways to *begin* an RLA before all ballots are available.

- **The audit must have the ability to correct incorrect outcomes.** This might mean that the audit must take place before results are certified, or that the audit can revise already-certified results.

3. **Set the risk limit in statute.** Allowing the Secretary of State or local election official to choose the risk limits may create a real or apparent conflict of interest.

4. **Specify how the contests to be audited are selected.**

    - If not every contest will be audited in every election, the selection of contests to audit should involve a random element to ensure that every contest has some chance of being selected, to ensure that a malicious opponent would not be able to predict whether any particular race will be audited.

    - Every contest not audited with an RLA should be audited using a *risk-measuring audit* instead.[35]

    - **Statute must require RLAs on cross-jurisdictional contests, including statewide contests.** Because the point of an RLA is to ensure that reported contest outcomes are correct, every county involved in a particular contest must examine ballots in such a way that the overall cross-jurisdictional procedure is an RLA of that contest. Operationally, auditing cross-jurisdictional contests requires coordination between (e.g.) counties, so that each county knows when its portion of the audit can stop. For example, the Secretary of State can tell each jurisdiction how many ballots it needs

---

[35] *Risk-measuring* audits are related to *risk-limiting* audits, but they do not have a pre-specified minimum chance of requiring a full manual tabulation when that tabulation would show a different result. In statistical terminology, a risk-measuring audit reports a *P*-value for the hypothesis that a full count would yield a different electoral outcome, based on the audit data. Equivalently, it reports the smallest value for which a risk-limiting audit conducted using that value as its risk limit would have stopped without examining more ballots.

to draw from each cross-jurisdictional contest, in light of the margin and what the audit reveals as it progresses.

5. **The audit sample must not be predictable before the audit starts**—otherwise any hacked software would know in which precincts it's safe to cheat. Audits in Colorado, California, Rhode Island, and elsewhere have initialized a random number generator by rolling dice in a public ceremony, to ensure that the sample is unknown until that time. [36]

   The sample from any collection of ballots should not be selected before election officials have "committed" to the tally of those ballots. For example, nobody should be able to know whether precinct 207 will be audited until the election official has published the tally for precinct 207.[37]

6. **The public must be able to verify that the RLA did not stop prematurely**, not merely "observe" the RLA. Among other things, this requires election officials to: Disclose the algorithms used to select the sample, to calculate the risk, and to determine when the audit can stop; provide public opportunity to observe the selection of the "seed" for drawing the sample; provide adequate evidence that the paper trail of cast ballots is complete and intact (evidence generated in part by the *compliance audit*); provide public opportunity to verify that the correct ballots were inspected during the audit; provide public opportunity to observe the voters' marks on the ballots that were inspected by the audit;[38] and, in "ballot-level comparison audits," the public also needs to see the cast-vote record for each audited ballot and proof that the full set of cast-vote records yields the reported contest results.

CONCLUSIONS

Electronic records of ballots are easy to manipulate by computer hacking. Therefore, voter-verified paper ballots must serve as the auditable evidence that connects the voters' selections with the election outcome.

Optical scan voting systems, using hand-marked paper ballots designed with usability in mind, have proved to be reliable and highly accurate. These voting systems should be used with compliance-auditable ballot accounting and chain-of-custody procedures, coupled with risk-limiting audits of election tallies, to achieve reliable and trustworthy evidence-based elections.

---

[36] Colorado's public ceremony is a good model. *See* Colorado Secretary of State's Office, *Colorado's risk-limiting audit*, YOUTUBE (Jun. 15, 2018), https://youtu.be/ysG4pFFmQ-E.

[37] There are examples (notably, in Cuyahoga County, OH) where election officials altered tallies in precincts selected for recount after the sample was selected, to ensure that the inspection would not find any discrepancies *See* Kim Zetter, *The Mysterious Case of Ohio's Voting Machines*, WIRED (Mar. 26, 2008, 5:51 PM), https://www.wired.com/2008/03/the-mysterious/ [https://perma.cc/2388-JC6G]

[38] It is important to have published rules governing how marks on ballots are to be interpreted in audits and recounts. For instance, if a voter makes a writes-in vote for a candidate who is also listed on the ballot, is that a valid vote? If a voter marks a vote for a listed candidate and also writes in that candidate's name, is that a valid vote? If a voter marks a vote for a candidate, crosses through the mark, and marks a vote for a second candidate, is that a valid vote for the second candidate? If a voter makes a stray mark on the ballot that is distinctive enough to identify the ballot, is the ballot valid?

Ballot-marking devices were originally envisioned as assistive devices for voters with disabilities who are unable to mark a paper ballot with a pen. Such BMDs have touchscreens, audio interfaces, and ports for other assistive interfaces for, e.g., voters with motor disabilities.

Only recently, some states and counties have adopted voting systems that use BMDs for all voters. In light of the insecurity of BMDs—the chasm between *voter-verifiable* and *voter-verified* BMD ballots—hand-marked paper ballots should be the default option presented to all voters, with BMDs available to voters who wish to use them.

Most states already use paper ballots; what we now need to conduct evidence-based elections is better procedures for safeguarding ballots, compliance audits, and risk-limiting audits. These procedures should be enacted in statutes, so they have sufficient force of law to truly safeguard our elections against software hacking, insider manipulation, and other threats.

Deploying RLAs (and associated compliance audits) involves the coordination of statistical methods, administrative procedures, paper handling, etc., by election administrators across towns, counties, and statewide (in each state). This cannot be done overnight: it requires developing methods appropriate to the election procedures in each state, training officials, educating citizens, practice, and experience. For this reason, the National Academies of Sciences report recommends that states and local jurisdictions begin with pilot programs and work toward full implementation.[39]

[39] NAT'L ACADS. SCIS., ENG'G, & MED., *supra* note 2.