Risk-limiting audits: theory and practice

DEF CON Voting Village
Las Vegas, NV

Philip B. Stark

9 August 2024

University of California, Berkeley

Many collaborators including (most recently):

Andrew Appel, Josh Benaloh, Michelle Blom, Andrew Conway, Rich DeMillo, Alexander Ek, Floyd Everest, Amanda Glazer, Alex Halderman, Zhuoqun Huang, Harri Hursti, Wojciech Jamroga, Dan King, Mark Lindeman, Kellie Ottoboni, Aaditya Ramdas, Ron Rivest, Peter Røenne, Peter Ryan, Laurent Sandrolini, Steve Schneider, Carsten Schürmann, Jake Spertus, Peter Stuckey, Vanessa Teague, Poorvi Vora, Damjan Vukcevic, Dan Wallach, Ian Waudby-Smith, Ran Xie

Half of Republicans say Biden won because of a 'rigged' election: Reuters/Ipsos poll

By Chris Kahn

3 MIN READ

(Reuters) - About half of all Republicans believe President Donald Trump “rightfully won” the U.S. election but that it was stolen from him by widespread voter fraud that favored Democratic President-elect Joe Biden, according to a new Reuters/Ipsos opinion poll.

The Nov. 13-17 opinion poll showed that Trump’s open defiance of Biden’s victory in both the popular vote and Electoral College appears to be affecting the public’s confidence in American democracy, especially among Republicans.

Evidence-Based Elections

P.B. Stark and D.A. Wagner

***Abstract*—We propose an alternative to current requirements for certifying voting equipment and conducting elections. We argue that elections should be structured to provide convincing affirmative evidence that the reported outcomes actually reflect how people voted. This can be accomplished with a combination of software-independent voting systems, compliance audits, and risk-limiting audits. Together, these yield a resilient canvass framework: a fault-tolerant approach to conducting elections that gives strong evidence that the reported outcome is correct or reports that the evidence is not convincing. We argue that, if evidence-based elections are adopted, certification and testing of voting equipment can be relaxed, saving money and time and reducing barriers to innovation in voting systems—and election integrity will benefit. We conclude that there should be more regulation of the evidence trail and less regulation of equipment, and that compliance audits and risk-limiting audits should be required.**

***Keywords*-elections, software-independent voting system, risk-limiting audit, resilient canvass framework EDICS SEC-INTE, APP-CRIM, APP-INTE, APP-OTHE.**

I. Introduction

IDEALLY, what should an election do? Certainly, an election should find out who won, but we believe it also should produce convincing evidence that it found the real winners—or report that it cannot. This is not automatic; it requires thoughtful design of voting equipment, carefully planned and implemented voting and vote counting processes, and rigorous post-election auditing.

While approximately 75% of US voters currently vote on equipment that produces a voter-verifiable paper record of the vote, about 25% vote on paperless electronic voting machines that do not produce such a record [1].

Because paperless electronic voting machines rely upon complex software and hardware, and because there is no feasible way to ensure that the voting software is free of bugs or that the hardware is executing the proper software, there is no guarantee that electronic voting machines record the voter's votes accurately. And, because paperless voting machines preserve only an electronic record of the vote that cannot be directly observed by voters, there is no way to produce convincing evidence that the electronic record accurately reflects the voters' intent. Internet voting shares the shortcomings of paperless electronic voting machines, and has additional vulnerabilities.

Numerous failures of electronic voting equipment have been documented. Paperless voting machines in Carteret County, North Carolina irretrievably lost 4,400 votes; other machines in Mecklenburg, North Carolina recorded 3,955 more votes than the number of people who voted; in Bernalillo County, New Mexico, machines recorded 2,700 more votes than voters; in Mahoning County, Ohio, some machines reported a negative total vote count; and in Fairfax, Virginia, county officials found that for every hundred or so votes cast for one candidate, the electronic voting machines subtracted one vote for her [2]. In short, when elections are conducted on paperless voting

GEORGETOWN LAW TECHNOLOGY REVIEW

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

Andrew W. Appel* & Philip B. Stark**

CITE AS: 4 GEO. L. TECH. REV. 523 (2020)

TABLE OF CONTENTS

1. **Require rigorous physical custody of ballots, and compliance audits, as discussed above.** A 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] This entails a number of things:

   a. **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, printout 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 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. Statutes 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 the defendant is guilty.

   b. **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.

   c. **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.

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

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.**

    a. **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.** This ensures that a malicious opponent will not be able to predict whether any particular race will be audited.

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

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

- Origin of RLAs: service on 2007 California Post Election Audit Standards Working Group

- Origin of RLAs: service on 2007 California Post Election Audit Standards Working Group
- Can audits help protect elections that rely on vulnerable technology?

- Origin of RLAs: service on 2007 California Post Election Audit Standards Working Group
- Can audits help protect elections that rely on vulnerable technology?
- Typical statutory audits: fixed percentage of batches, no consequences
    - Are the machines working properly?

- Origin of RLAs: service on 2007 California Post Election Audit Standards Working Group
- Can audits help protect elections that rely on vulnerable technology?
- Typical statutory audits: fixed percentage of batches, no consequences
  - Are the machines working properly?
- Detection (state-of-the-art 2007):
  - the audit should have big chance of finding at least one error if outcome is wrong
  - but errors often occur even when the outcome is right!

- Origin of RLAs: service on 2007 California Post Election Audit Standards Working Group
- Can audits help protect elections that rely on vulnerable technology?
- Typical statutory audits: fixed percentage of batches, no consequences
  - Are the machines working properly?
- Detection (state-of-the-art 2007):
  - the audit should have big chance of finding at least one error if outcome is wrong
  - but errors often occur even when the outcome is right!
- Affirmative evidence:
  - has the audit given strong evidence that the reported winners really won?
  - if not, collect more evidence or do full count of trustworthy vote record to see who won

- Origin of RLAs: service on 2007 California Post Election Audit Standards Working Group
- Can audits help protect elections that rely on vulnerable technology?
- Typical statutory audits: fixed percentage of batches, no consequences
  - Are the machines working properly?
- Detection (state-of-the-art 2007):
  - the audit should have big chance of finding at least one error if outcome is wrong
  - but errors often occur even when the outcome is right!
- Affirmative evidence:
  - has the audit given strong evidence that the reported winners really won?
  - if not, collect more evidence or do full count of trustworthy vote record to see who won

"Trustworthy" means a complete, accurate count would show who really won.

"Who really won" means who won according to an accurate count of the expressed

Some records born untrustworthy: malleable or vulnerable tech btw voter & record, such as BMDs or client/Internet/server.

Some records born untrustworthy: malleable or vulnerable tech btw voter & record, such as BMDs or client/Internet/server.

To stay trustworthy, need:

- physical inventories of ballots & other materials
- demonstrably secure chain of custody
- appropriate physical security
- eligibility audits
- ballot accounting
- pollbook and participation reconciliation
- comparisons with registration
- trustworthy upper bound on # validly cast cards containing each contest

Some records born untrustworthy: malleable or vulnerable tech btw voter & record, such as BMDs or client/Internet/server.

To stay trustworthy, need:

- physical inventories of ballots & other materials
- demonstrably secure chain of custody
- appropriate physical security
- eligibility audits
- ballot accounting
- pollbook and participation reconciliation
- comparisons with registration
- trustworthy upper bound on # validly cast cards containing each contest

Can't have cyber-resilience w/o some physical security

Risk-Limiting Audit (RLA)

Limit *risk* that an incorrect outcome will be certified.

Corrects wrong reported outcomes w/ high probability.

Never changes correct reported outcomes.

Risk: maximum chance of certifying the outcome if the outcome is in fact wrong.

Risk-Limiting Audit (RLA)

Limit *risk* that an incorrect outcome will be certified.

Corrects wrong reported outcomes w/ high probability.

Never changes correct reported outcomes.

Risk: maximum chance of certifying the outcome if the outcome is in fact wrong.

RLA cannot *restore* trustworthiness to a poorly run election.

Leverages trustworthiness of the vote record in a well-run election to provide affirmative evidence that the reported winners really won–or correct the results if not.

RLA pseudo-algorithm

Input: trustworthy, organized record of all validly cast votes; auxiliary randomness

Output: strong evidence that reported outcome is correct, or correct outcome

```
while (!(full handcount) && !(strong evidence outcome is correct)) {
    examine more ballots
}
```

RLA pseudo-algorithm

Input: trustworthy, organized record of all validly cast votes; auxiliary randomness

Output: strong evidence that reported outcome is correct, or correct outcome

```
while (!(full handcount) && !(strong evidence outcome is correct)) {
    examine more ballots
}

if (full handcount) {
    handcount result replaces reported result
}
```

The National Academies of SCIENCES ENGINEERING MEDICINE

Securing the Vote: Protecting American Democracy, 2018

Home

Elections should be conducted with human-readable paper ballots. Paper ballots form a body of evidence that is not subject to manipulation by faulty software or hardware and that can be used to audit and verify the results of an election. Human-readable paper ballots may be marked by hand or by machine (using a ballot-marking device), and they may be counted by hand or by machine (using an optical scanner), the report says. Voters should have an opportunity to review and confirm their selections before depositing the ballot for tabulation. Voting machines that do not provide the capacity for independent auditing – i.e., machines that do not produce a printout of a voter's selections that can be verified by the voter and used in audits – should be removed from service as soon as possible.

States should mandate a specific type of audit known as a "risk-limiting" audit prior to the certification of election results. By examining a statistically appropriate random sample of paper ballots, risk-limiting audits can determine with a high level of confidence whether a reported election outcome reflects a correct tabulation

Risk-Limiting Audits

- Endorsed by NASEM, PCEA, ASA, LWV, CC, VV, ...
- ~60 pilot audits in about 17 states and DK
- Laws in ~15 states
- Methods for plurality, multi-winner plurality, supermajority, proportional representation, IRV/RCV, Borda count, all 'scoring rules'

The Annals of Applied Statistics
2008, Vol. 2, No. 2, 550–581
DOI: 10.1214/08-AOAS161

CONSERVATIVE STATISTICAL POST-ELECTION AUDITS

BY PHILIP B. STARK

University of California, Berkeley

There are many sources of error in counting votes: the apparent winner might not be the rightful winner. Hand tallies of the votes in a random sample of precincts can be used to test the hypothesis that a full manual recount would find a different outcome. This paper develops a conservative sequential test based on the vote-counting errors found in a hand tally of a simple or stratified random sample of precincts. The procedure includes a natural escalation: If the hypothesis that the apparent outcome is incorrect is not rejected at stage s, more precincts are audited. Eventually, either the hypothesis is rejected—and the apparent outcome is confirmed—or all precincts have been audited and the true outcome is known. The test uses a priori bounds on the overstatement of the margin that could result from error in each precinct. Such bounds can be derived from the reported counts in each precinct and upper bounds on the number of votes cast in each precinct. The test allows errors in different precincts to be treated differently to reflect voting technology or precinct sizes. It is not optimal, but it is conservative: the chance of erroneously confirming the outcome of a contest if a full manual recount would show a different outcome is no larger than the nominal significance level. The approach also gives a conservative P-value for the hypothesis that a full manual recount would find a different outcome, given the errors found in a fixed size sample. This is illustrated with two contests from November, 2006: the U.S. Senate race in Minnesota and a school board race for the Sausalito Marin City School District in California, a small contest in which voters could vote for up to three candidates.

1. Introduction. Votes can be miscounted because of human error (by voters or election workers), hardware or software "bugs" or deliberate fraud. Post-election audits—manual tallies of votes in individual precincts—are intended to detect miscount, especially miscount large enough to alter the outcome of the election.[1] To the best of my knowledge, eighteen states require or allow post-election audits [National Association of Secretaries of State (2007) and Verified Voting

Received October 2007; revised March 2008.

EVT-WOTE 2009

Implementing Risk-Limiting Post-Election Audits in California

Joseph Lorenzo Hall[1,2,*], Luke W. Miratrix[3], Philip B. Stark[3], Melvin Briones[4], Elaine Ginnold[4], Freddie Oakley[5], Martin Peaden[6], Gail Pellerin[6], Tom Stanionis[5], and Tricia Webber[6]

[1]University of California, Berkeley; School of Information
[2]Princeton University; Center for Information Technology Policy
[3]University of California, Berkeley; Department of Statistics
[4]Marin County, California; Registrar of Voters
[5]Yolo County, California; County Clerk/Recorder
[6]Santa Cruz County, California; County Clerk

Abstract

Risk-limiting post-election audits limit the chance of certifying an electoral outcome if the outcome is not what a full hand count would show. Building on previous work [18, 17, 20, 21, 11], we report pilot risk-limiting audits in four elections during 2008 in three California counties: one during the February 2008 Primary Election in Marin County and three during the November 2008 General Elections in Marin, Santa Cruz and Yolo Counties. We explain what makes an audit *risk-limiting* and how existing and proposed laws fall short. We discuss the differences among our four pilot audits. We identify challenges to practical, efficient risk-limiting audits and conclude that current approaches are too complex to be used routinely on a large scale. One important logistical bottleneck is the difficulty of exporting data from commercial election management systems in a format amenable to audit calculations. Finally, we propose a bare-bones risk-limiting audit that is less efficient than these pilot audits, but avoids many practical problems.

Sets of Half-Average Nulls Generate Risk-Limiting Audits: SHANGRLA

ESORICS Voting 20, LNCS

Philip B. Stark

University of California, Berkeley

31 January 2020

Abstract. Risk-limiting audits (RLAs) for many social choice functions can be reduced to testing sets of null hypotheses of the form "the average of this list is not greater than 1/2" for a collection of finite lists of non-negative numbers. Such social choice functions include majority, super-majority, plurality, multi-winner plurality, Instant Runoff Voting (IRV), Borda count, approval voting, and STAR-Voting, among others. The audit stops without a full hand count iff all the null hypotheses are rejected. The nulls can be tested in many ways. Ballot polling is particularly simple; two new ballot-polling risk-measuring functions for sampling without replacement are given. Ballot-level comparison audits transform each null into an equivalent assertion that the mean of re-scaled tabulation errors is not greater than 1/2. In turn, that null can then be tested using the same statistical methods used for ballot polling—applied to different finite lists of non-negative numbers: the SHANGRLA approach reduces auditing different social choice functions and different audit methods to the same simple statistical problem. Moreover, SHANGRLA comparison audits are more efficient than previous comparison audits for two reasons: (i) for most social choice functions, the conditions tested are both necessary and sufficient for the reported outcome to be correct, while previous methods tested conditions that were sufficient but not necessary, and (ii) the tests avoid a conservative approximation. The SHANGRLA abstraction simplifies stratified audits, including audits that combine ballot polling with ballot-level comparisons, producing sharper audits than the "SUITE" approach. SHANGRLA works with the "phantoms to evil zombies" strategy to treat missing ballot cards and missing or redacted cast vote records. That also facilitates sampling from "ballot-style manifests," which can dramatically improve efficiency when the audited contests do not appear on every ballot card. Open-source software implementing SHANGRLA ballot-level comparison audits is available. SHANGRLA was tested in a pilot audit of an instant-runoff contest in San Francisco, CA, in November, 2019.

Keywords: sequential tests, martingales, Kolmogorov's inequality

Acknowledgments: I am grateful to Andrew Conway, Steven N. Evans, Kellie Ottoboni, Ronald L. Rivest, Vanessa Teague, and Poorvi Vora for helpful conversations and comments on earlier drafts. The SHANGRLA software was a collaborative effort that included Michelle Blom, Andrew Conway, Dan King, Laurent Sandrolini, Peter Stuckey, and Vanessa Teague.

The Annals of Applied Statistics
2023, Vol. 17, No. 1, 641–679
https://doi.org/10.1214/22-AOAS1646

ALPHA: AUDIT THAT LEARNS FROM PREVIOUSLY HAND-AUDITED BALLOTS

BY PHILIP B. STARK[a]

Department of Statistics, University of California, Berkeley, [a]*stark@stat.berkeley.edu*

A risk-limiting election audit (RLA) offers a statistical guarantee: if the reported electoral outcome is incorrect, the audit has a known maximum chance (the risk limit) of not correcting it before it becomes final. BRAVO (Lindeman, Stark and Yates (In *Proceedings of the 2011 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE'11)* (2012) USENIX)), based on Wald's sequential probability ratio test for the Bernoulli parameter, is the simplest and most widely tried method for RLAs, but it has limitations. It cannot accommodate sampling without replacement or stratified sampling which can improve efficiency and are sometimes required by law. It applies only to ballot-polling audits which are less efficient than comparison audits. It applies to plurality, majority, supermajority, proportional representation, and instant-runoff voting (IRV, using RAIRE (Blom, Stuckey and Teague (In *Electronic Voting* (2018) 17–34 Springer))) but not to other social choice functions for which there are RLA methods. And while BRAVO has the smallest expected sample size among sequentially valid ballot-polling-with-replacement methods when the reported vote shares are exactly correct, it can require arbitrarily large samples when the reported reported winner(s) really won but the reported vote shares are incorrect. ALPHA is a simple generalization of BRAVO that: (i) works for sampling with and without replacement, with and without weights, with and without stratification, and for Bernoulli sampling; (ii) works not only for ballot polling but also for ballot-level comparison, batch polling, and batch-level comparison audits; (iii) works for all social choice functions covered by SHANGRLA (Stark (In *Financial Cryptography and Data Security* (2020) Springer)), including approval voting, STAR-Voting, proportional representation schemes, such as D'Hondt and Hamilton, IRV, Borda count, and all scoring rules, and (iv) in situations where both ALPHA and BRAVO apply, requires smaller samples than BRAVO when the reported vote shares are wrong but the outcome is correct—five orders of magnitude in some examples. ALPHA includes the family of betting martingale tests in RiLACS (Waudby-Smith, Stark and Ramdas (In *Electronic Voting. E-Vote-ID 2021* (2021) Springer)) with a different betting strategy parametrized as an estimator of the population mean and explicit flexibility to accommodate sampling weights and population bounds that change with each draw. A Python implementation is provided.

Non(c)esuch Ballot-Level Comparison Risk-Limiting Audits

ESORICS, Voting 2022, LNCS

Philip B. Stark(✉)

University of California, Berkeley, CA, USA
stark@stat.berkeley.edu

Abstract. Risk-limiting audits (RLAs) guarantee a high probability of correcting incorrect reported electoral outcomes before the outcomes are certified. The most efficient are *ballot-level comparison audits* (BLCAs), which compare the voting system's interpretation of randomly selected individual ballot cards (*cast-vote records*, CVRs) from a trustworthy paper trail to a human interpretation of the same cards. BLCAs have logistical and privacy hurdles: Individual randomly selected cards must be retrieved for manual inspection; the voting system must export CVRs; and the CVRs must be linked to the corresponding physical cards, to compare the two. In practice, such links have been made by keeping cards in the order in which they are scanned or by printing serial numbers on cards as they are scanned. Both methods may compromise voter privacy. Cards selected for audit have been retrieved by manually counting into stacks or by looking for cards with particular serial numbers. The methods are time-consuming; the first is also error-prone. Connecting CVRs to cards using a unique pseudo-random number ("cryptographic nonce") printed on each card after the voter last sees it could reduce privacy risks, but retrieving the card imprinted with a particular random number may be harder than counting into a stack or finding the card with a given serial number. And what if the system does not in fact print a unique number on each ballot or does not accurately report the numbers it printed? This paper presents a method for conducting BLCAs that maintains the risk limit even if the system does not print a genuine nonce on each ballot or misreports the identifiers it used. The method also allows untrusted technology to be used to retrieve the cards selected for audit—automation that may reduce audit workload even if cards are imprinted with serial numbers rather than putative nonces. The method limits the risk rigorously, even if the imprinting or retrieval technology misbehaves. If the imprinting and retrieval systems behave properly, this protection does not increase the number of cards the RLA has to inspect to confirm or correct the outcome.

Keywords: Risk-limiting audit · Voter privacy

Overstatement-Net-Equivalent Risk-Limiting Audit: ONEAudit

ESORICS Voting 23: LNCS

Philip B. Stark[1][0000–0002–3771–9604]

University of California, Berkeley, CA USA stark@stat.berkeley.edu

Abstract. A procedure is a risk-limiting audit (RLA) with risk limit α if it has probability at least $1 - \alpha$ of correcting each wrong reported outcome and never alters correct outcomes. One efficient RLA method, card-level comparison (CLCA), compares human interpretation of individual ballot cards randomly selected from a trustworthy paper trail to the voting system's interpretation of the same cards (cast vote records, CVRs). CLCAs heretofore required a CVR for each cast card and a "link" identifying which CVR is for which card—which many voting systems cannot provide. This paper shows that every set of CVRs that produces the same aggregate results overstates contest margins by the same amount: they are *overstatement-net-equivalent* (ONE). CLCA can therefore use CVRs from the voting system for any number of cards and ONE CVRs created *ad lib* for the rest. In particular:

- Ballot-polling RLA is algebraically equivalent to CLCA using ONE CVRs derived from the overall contest results.
- CLCA can be based on batch-level results (e.g., precinct subtotals) by constructing ONE CVRs for each batch. In contrast to batch-level comparison auditing (BLCA), this avoids manually tabulating entire batches and works even when reporting batches do not correspond to physically identifiable batches of cards, when BLCA is impractical.
- If the voting system can export linked CVRs for only some ballot cards, auditors can still use CLCA by constructing ONE CVRs for the rest of the cards from contest results or batch subtotals.

This works for every social choice function for which there is a known RLA method, including IRV. Sample sizes for BPA and CLCA using ONE CVRs based on contest totals are comparable. With ONE CVRs from batch subtotals, sample sizes are smaller than than for BPA when batches are homogeneous, approaching those of CLCA using CVRs from the voting system, and much smaller than for BLCA: A CLCA of the 2022 presidential election in California at risk limit 5% using ONE CVRs for precinct-level results would sample approximately 70 ballots statewide, if the reported results are accurate, compared to about 26,700 for BLCA. The 2022 Georgia audit tabulated more than 231,000 cards (the expected BLCA sample size was ≈103,000 cards); ONEAudit would have audited ≈1,300 cards. For data from a pilot hybrid RLA in Kalamazoo, MI, in 2018, ONEAudit gives a risk of 2%, substantially lower than the 3.7% measured risk for SUITE, the "hybrid" method the pilot used.

Keywords: Risk-limiting audit, BPA, card-level comparison audit, batch-level comparison audit

You can do RLAs for IRV

The Process Pilot of Risk-Limiting Audits for the San Francisco District Attorney 2019 Instant Runoff Vote

Michelle Blom*, Andrew Conway†, Dan King‡, Laurent Sandrolini§, Philip B. Stark¶, Peter J. Stuckey‖ and Vanessa Teague**

April 2, 2020

The City and County of San Francisco, CA, has used Instant Runoff Voting (IRV) for some elections since 2004. This report describes the first ever process pilot of Risk Limiting Audits for IRV, for the San Francisco District Attorney's race in November, 2019. We found that the vote-by-mail outcome could be efficiently audited to well under the 0.05 risk limit given a sample of only 200 ballots. All the software we developed for the pilot is open source.

1. Introduction

Post-election audits test a reported election result by randomly sampling paper ballots.[1] A *Risk Limiting Audit (RLA)* of a trustworthy paper trail of votes either finds strong statistical evidence that the reported outcome is correct, or reverts to a full manual tabulation to set the record straight.[2] (The outcome is the political result—i.e., who won—not the exact vote counts.) The maximum chance that a RLA will fail to correct the reported outcome if the reported outcome is wrong is the *risk limit*. RLAs are

*School of Computing and Information Systems, University of Melbourne. michelle.blom@unimelb.edu.au

†Silicon Econometrics Pty. Ltd., andrewelections@greatcactus.org

Adaptively Weighted Audits of Instant-Runoff Voting Elections: AWAIRE

2023 E-Vote-ID

Alexander Ek[1], Philip B. Stark[2], Peter J. Stuckey[3], and Damjan Vukcevic[1](✉)

[1] Department of Econometrics and Business Statistics, Monash University, Clayton, Australia
damjan.vukcevic@monash.edu

[2] Department of Statistics, University of California, Berkeley, CA, USA

[3] Department of Data Science and AI, Monash University, Clayton, Australia

Abstract. An election audit is *risk-limiting* if the audit limits (to a pre-specified threshold) the chance that an erroneous electoral outcome will be certified. Extant methods for auditing instant-runoff voting (IRV) elections are either not risk-limiting or require cast vote records (CVRs), the voting system's electronic record of the votes on each ballot. CVRs are not always available, for instance, in jurisdictions that tabulate IRV contests manually.

We develop an RLA method (AWAIRE) that uses adaptively weighted averages of test supermartingales to efficiently audit IRV elections when CVRs are not available. The adaptive weighting 'learns' an efficient set of hypotheses to test to confirm the election outcome. When accurate CVRs are available, AWAIRE can use them to increase the efficiency to match the performance of existing methods that require CVRs.

We provide an open-source prototype implementation that can handle elections with up to six candidates. Simulations using data from real elections show that AWAIRE is likely to be efficient in practice. We discuss how to extend the computational approach to handle elections with more candidates.

Adaptively weighted averages of test supermartingales are a general tool, useful beyond election audits to test collections of hypotheses sequentially while rigorously controlling the familywise error rate.

Stylish Risk-Limiting Audits in Practice

Amanda K. Glazer[1,2], Jacob V. Spertus[3], Philip B. Stark[4]

Abstract: Risk-limiting audits (RLAs) can use information about which ballot cards contain which contests (*card-style data*, CSD) to ensure that each contest receives adequate scrutiny, without examining more cards than necessary. RLAs using CSD in this way can be substantially more efficient than RLAs that sample indiscriminately from all cast cards. We describe an open-source Python implementation of RLAs using CSD for the Hart InterCivic Verity voting system and the Dominion Democracy Suite® voting system. The software is demonstrated using all 181 contests in the 2020 general election and all 214 contests in the 2022 general election in Orange County, CA, USA, the fifth-largest election jurisdiction in the U.S., with over 1.8 million active voters. (Orange County uses the Hart Verity system.) To audit the 181 contests in 2020 to a risk limit of 5% without using CSD would have required a complete hand tally of all 3,094,308 cast ballot cards. With CSD, the estimated sample size is about 20,100 cards, 0.65% of the cards cast—including one tied contest that required a complete hand count. To audit the 214 contests in 2022 to a risk limit of 5% without using CSD would have required a complete hand tally of all 1,989,416 cast cards. With CSD, the estimated sample size is about 62,250 ballots, 3.1% of cards cast—including three contests with margins below 0.1% and 9 with margins below 0.5%.

Quantifying Evidence

Heuristic: in a series of fair or unfavorable games, you are unlikely ever to win a large multiple of your initial stake.

Example: betting on a fair coin.

- Start with \$1 stake.
- Bet on the outcome of the next coin toss:
  - If coin lands heads, you get back your bet, doubled.
  - If coin lands tails, you lose your bet.
  - Can bet up to your current fortune, but can't borrow.
  - Can use any betting scheme whatsoever
- If you go broke, you're out.

The chance your fortune ever reaches \$10 is at most 10% (1/10).
The chance it ever reaches \$20 is at most 5% (1/20).
The chance it ever reaches \$k is at most 1/k.

Quantifying Evidence

Heuristic: in a series of fair or unfavorable games, you are unlikely ever to win a large multiple of your initial stake.

Example: betting on a fair coin.

- Start with \$1 stake.
- Bet on the outcome of the next coin toss:
  - If coin lands heads, you get back your bet, doubled.
  - If coin lands tails, you lose your bet.
  - Can bet up to your current fortune, but can't borrow.
  - Can use any betting scheme whatsoever
- If you go broke, you're out.

The chance your fortune ever reaches \$10 is at most 10% (1/10).
The chance it ever reaches \$20 is at most 5% (1/20).
The chance it ever reaches \$k is at most 1/k.

Betting on ballots

Create set of repeated gambling games, at least one of which is fair or unfavorable if any reported winner didn't really win. (SHANGRLA + ALPHA)

- Each game involves betting on the next number sampled at random from a list.
- Each game involves a different list.
- If the outcome is wrong, the mean of at least one of the lists is $\leq 1/2$.
- Start with a stake of $1 in each game.
- Bet using any strategy you want (can't peek into the future).
- If your fortune gets to $20 in every game, audit stops.
- If you go broke in any game, do a full hand count.
- If you don't get to $20 in every game (or get bored), full hand count of trustworthy vote record.

Ville (1939): In a sequence of fair or sub-fair wagers in which you aren't allowed to borrow money, the chance you ever multiply your bankroll by k is at most 1/k

Ville (1939): In a sequence of fair or sub-fair wagers in which you aren't allowed to borrow money, the chance you ever multiply your bankroll by k is at most 1/k

At most 5% chance you get to $20 in every game if any reported winner didn't really win.

Thus, RLA with risk limit 5%.

Better betting strategies → more efficient audits: current research

Common misconceptions

- RLAs are "tabulation audits"
- RLAs work by checking whether the vote shares in the sample "match" the reported vote shares.
- RLAs assume that errors/malfunctions/problems are distributed randomly.
- RLAs are worthwhile even if the paper trail isn't trustworthy.

RLA isn’t “tabulation audit.” Doesn’t check tabulation: checks who won.

RLA isn't "tabulation audit." Doesn't check tabulation: checks who won.

	ballot 1	ballot 2	ballot 3	ballot 4	ballot 5	ballot 6	ballot 7	ballot 8
machine	Alice	Alice	Alice	Alice	Bob	Bob	Bob	none

RLA isn't "tabulation audit." Doesn't check tabulation: checks whether accurate tabulation would find the same winners.

	ballot 1	ballot 2	ballot 3	ballot 4	ballot 5	ballot 6	ballot 7	ballot 8
machine	Alice	Alice	Alice	Alice	Bob	Bob	Bob	none
reality	Bob	none	none	none	none	none	Alice	Alice

RLA isn't "tabulation audit." Doesn't check tabulation: checks whether accurate tabulation would find the same winners.

	ballot 1	ballot 2	ballot 3	ballot 4	ballot 5	ballot 6	ballot 7	ballot 8
machine	Alice	Alice	Alice	Alice	Bob	Bob	Bob	none
reality	Bob	none	none	none	none	none	Alice	Alice

Every ballot was mistabulated, and the totals are wrong:

tally	Alice	Bob	non-vote
machine	4	3	1
reality	2	1	5

RLA isn't "tabulation audit." Doesn't check tabulation: checks whether accurate tabulation would find the same winners.

	ballot 1	ballot 2	ballot 3	ballot 4	ballot 5	ballot 6	ballot 7	ballot 8
machine	Alice	Alice	Alice	Alice	Bob	Bob	Bob	none
reality	Bob	none	none	none	none	none	Alice	Alice

Every ballot was mistabulated, and the totals are wrong:

tally	Alice	Bob	non-vote
machine	4	3	1
reality	2	1	5

But Alice really won, so it's appropriate for an RLA to stop without a full handcount.

Addressing other misconceptions

- RLAs do *not* work by comparing the vote shares in the sample to the reported vote shares: they assess the evidence that the reported winners won.
- RLAs *do* work whether errors/misbehavior is "random" or adversarial: the risk calculations assume "worst-case"
- Applying RLA procedures to untrustworthy vote records *cannot* provide evidence that reported outcomes are correct: security theater.

Abuse

RLAs increasingly used to distract from a "game-over" problem: no trustworthy, organized, complete record of expressed preferences of eligible voters who validly cast ballots.

E.g., GA SoS Raffensperger claimed that a (deeply flawed) audit of one contest in 2020 based on untrustworthy paper trail "reaffirmed that the state's new secure paper ballot voting system accurately counted and reported results."
https://sos.ga.gov/news/historic-first-statewide-audit-paper-ballots-upholds-result-presidential-race

When Audits and Recounts Distract from Election Integrity: The 2020 U.S. Presidential Election in Georgia

Philip B. Stark[1][0000−0002−3771−9604]

University of California, Berkeley, CA USA pbstark@berkeley.edu

Abstract. The U.S. state of Georgia was central to efforts to overturn the results of the 2020 Presidential election, including a phone call from then-president Donald Trump to Georgia Secretary of State Brad Raffensperger asking Raffensperger to 'find' 11,780 votes. Raffensperger has maintained that a '100% full-count risk-limiting audit' and a machine recount agreed with the initial machine-count results, which proved that the reported election results were accurate and that 'no votes were flipped.' While there is no indication of widespread fraud, there is reason to distrust the election outcome: the two machine counts and the manual 'audit' tallies disagree substantially, even about the number of ballots cast. Some ballots in Fulton County, Georgia, were included in the original count at least twice; some were included in the machine recount at least thrice. Audit handcount results for some tally batches were omitted from the reported audit totals: reported audit results do not include every vote the auditors counted. In short, the two machine counts and the audit were not probative of who won because of poor processes and controls: a lack of secure physical chain of custody, ballot accounting, pollbook reconciliation, and accounting for other election materials such as memory cards. Moreover, most voters voted with demonstrably untrustworthy ballot-marking devices, so even a perfect handcount or audit would not necessarily reveal who really won. True risk-limiting audits (RLAs) and rigorous recounts can limit the risk that an incorrect electoral outcome will be certified rather than being corrected. But no procedure can limit that risk without a trustworthy record of the vote. And even a properly conducted RLA of some contests in an election does not show that any other contests in that election were decided correctly. The 2020 U.S. Presidential election in Georgia illustrates unrecoverable errors that can render recounts and audits 'security theater' that distract from the more serious problems rather than justifying trust.

2 The 2020 audit

Secretary of State Brad Raffensperger claimed, "Georgia's historic first statewide audit reaffirmed that the state's new secure paper ballot voting system accurately counted and reported results."[2] Moreover, "[] we did a 100 percent risk-limiting audit with a hand recount which proved the accuracy of the count and also proved that the machines were accurately counting it, and that no votes were flipped."[3] VotingWorks Executive Director Ben Adida claimed "Georgia's first statewide audit successfully confirmed the winner of the chosen contest and should give voters increased confidence in the results."[4] Per the official report of the audit, "[t]he audit confirmed the original result of the election, namely that Joe Biden won the Presidential Contest in the State of Georgia. The audit [] provides sufficient evidence that the correct winner was reported."[5]

Secretary Raffensperger has also used the recount and audit in his defense against a lawsuit that seeks to provide all Georgia voters the option to hand-mark paper ballots in person, rather than being compelled to use BMDs (Curling et al. v. Raffensperger et al., Civil Action No. 1:17-CV-2989-AT, U.S. District Court for the Northern District of Georgia, Atlanta Division). Raffensperger


411412
511221
411413
521225
411418
561341
421224
421422
411221
421222
421225
421226