Evidence-Based Elections

School of Cybersecurity and Privacy
Georgia Tech
Atlanta, GA

Philip B. Stark

4 February 2022

University of California, Berkeley

Many collaborators including (most recently) Andrew Appel, Josh Benaloh, Michelle Blom, Andrew Conway, Rich DeMillo, Amanda Glazer, Alex Halderman, Harri Hursti, Wojciech Jamroga, Mark Lindeman, Kellie Ottoboni, Aaditya Ramdas, Ron Rivest, Peter Rønne, Peter Ryan, Steve Schneider, Carsten Schürmann, Jake Spertus, Peter Stuckey, Vanessa Teague, Poorvi Vora, Damjan Vukcevic, Dan Wallach, Ian Waudby-Smith

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.

S. Ct. Case No. __________

11th Cir. Case No. 20-14418
N.D. Ga. Case No. 20-cv-04651-SDG

IN THE

SUPREME COURT OF THE UNITED STATES

L. LIN WOOD, JR.

Petitioner,

vs.

BRAD RAFFENSPERGER, et al.,

Respondents.

PETITION FOR WRIT OF CERTIORARI

On Petition for a Writ of Certiorari to the Eleventh Circuit Court of Appeals.

L. Lin Wood, Esq. (lead counsel)
GA Bar No. 774588
L. LIN WOOD, P.C.
P.O. BOX 52584
Atlanta, GA 30305-0584
(404) 891-1402
lwood@linwoodlaw.com

Harry W. MacDougald
Georgia Bar No. 463076
Caldwell, Propst & Deloach, LLP

FIFTH SUPPLEMENTAL DECLARATION OF PHILIP B. STARK

PHILIP B. STARK hereby declares as follows:

1. This statement supplements my declarations of September 9, 2018, September 30, 2018, October 22, 2019, and December 16, 2019. I stand by everything in the previous declarations.

I. False Assertions about the Fulton County Pilot Audit

2. Secretary of State Raffensperger issued the following (undated) press release on approximately June 30, 2020:[1]

> AUDIT SUPPORTS PRIMARY OUTCOME
>
> **(ATLANTA)** – A pilot post-election audit Monday confirmed the outcomes of the presidential preference primaries in Fulton County, Secretary of State Brad Raffensperger announced today.
>
> "This procedure demonstrates once again the validity of the results produced by Georgia's new secure paper-ballot system," [SOS Raffensperger] said. "Auditing

[1] https://sos.ga.gov/index.php/general/audit_supports_primary_outcome last visited 27 July 2020.

IN THE UNITED STATES DISTRICT COURT FOR THE NORTHERN DISTRICT OF GEORGIA ATLANTA DIVISION

DONNA CURLING, et al. **Plaintiff,** **vs.** **BRIAN P. KEMP, et al.** **Defendant.**	**CIVIL ACTION FILE NO.: 1:17-cv-2989-AT**

FIFTH SUPPLEMENTAL DECLARATION OF PHILIP B. STARK

PHILIP B. STARK hereby declares as follows:

1. This statement supplements my declarations of September 9, 2018, September 30, 2018, October 22, 2019, and December 16, 2019. I stand by everything in the previous declarations.

I. False Assertions about the Fulton County Pilot Audit

2. Secretary of State Raffensperger issued the following (undated) press release on approximately June 30, 2020:[1]

> AUDIT SUPPORTS PRIMARY OUTCOME
>
> **(ATLANTA)** – A pilot post-election audit Monday confirmed the outcomes of the presidential preference primaries in Fulton County, Secretary of State Brad Raffensperger announced today.
>
> "This procedure demonstrates once again the validity of the results produced by Georgia's new secure paper-ballot system," [SOS Raffensperger] said. "Auditing

[1] https://sos.ga.gov/index.php/general/audit_supports_primary_outcome last visited 27 July 2020.

POLITICS ELECTION 2020 GEORGIA

Why Georgia's Unscientific Recount 'Horrified' Experts

Observers, including the inventor of the auditing process used by the state, were skeptical of a measure seemingly aimed at placating the GOP.

By Timothy Pratt

NOVEMBER 20, 2020

- US elections neither *tamper evident* nor *resilient*.

- US elections neither *tamper evident* nor *resilient*.
- Need systems/procedures that can provide strong evidence that the reported winners really won.

- US elections neither *tamper evident* nor *resilient*.
- Need systems/procedures that can provide strong evidence that the reported winners really won.
- *Every* electronic system is vulnerable to bugs, configuration errors, & hacking.

Security properties of paper

- tangible/accountable
- tamper evident
- human readable
- large alteration/substitution attacks require physical access & many accomplices

Security properties of paper

- tangible/accountable
- tamper evident
- human readable
- large alteration/substitution attacks require physical access & many accomplices

Not all paper is trustworthy

Mother Jones

Smart, Fear

POLITICS ENVIRONMENT CRIME AND JUSTICE FOOD MEDIA INVESTIGATIONS PHO

POLITICS JANUARY 8, 2020

A New Voting System Promises Reliable Paper Records. Security Experts Warn It Can't Be Trusted.

A just-released study says over ninety percent of errors introduced by ballot marking devices go undetected.

AJ VICENS
Reporter

FREEDOM TO TINKER

research and expert commentary on digital technologies in public life

Serious design flaw in ESS ExpressVote touchscreen: "permission to cheat"

SEPTEMBER 14, 2018 BY ANDREW APPEL

Kansas, Delaware, and New Jersey are in the process of purchasing voting machines with a serious design flaw, and they should reconsider while there is still time!

Over the past 15 years, almost all the states have moved away from paperless touchscreen voting systems (DREs) to optical-scan paper ballots. They've done so because if a paperless touchscreen is hacked to give fraudulent results, there's no way to know and no way to correct; but if an optical scanner were hacked to give fraudulent results, the fraud could be detected by a random audit *of the paper ballots that the voters actually marked*, and corrected by a recount of those paper ballots.

Election Law Journal: Rules, Politics, and Policy, VOL. 19, NO. 3 | Original Research Articles normal

Ballot-Marking Devices Cannot Ensure the Will of the Voters

Andrew W. Appel, Richard A. DeMillo, and Philip B. Stark

Published Online: 17 Sep 2020 | https://doi.org/10.1089/elj.2019.0619

Tools Share

Abstract

The complexity of U.S. elections usually requires computers to count ballots—but computers can be hacked, so election integrity requires a voting system in which paper ballots can be recounted by hand. However, paper ballots provide no assurance unless they accurately record the votes as expressed by the voters.

Voters can express their intent by indelibly hand-marking ballots or using computers called ballot-marking devices (BMDs). Voters can make mistakes in expressing their intent in either technology, but only BMDs are also subject to hacking, bugs, and misconfiguration of the software that prints the marked ballots. Most voters do not review BMD-printed ballots, and those who do often fail to notice when the printed vote is not what they expressed on the touchscreen. Furthermore, there is no action a voter can take to demonstrate to election officials that a BMD altered their expressed votes, nor is there a corrective action that election officials can take if notified by voters—there is no way to deter, contain, or correct computer hacking in BMDs. These are the essential security flaws of BMDs.

Risk-limiting audits can ensure that the votes recorded on paper ballots are tabulated correctly, but no audit can ensure that the votes on paper are the ones expressed by the voter on a touchscreen: Elections conducted on current BMDs cannot be confirmed by audits. We identify two properties of voting systems, *contestability* and *defensibility*, necessary for audits to confirm election outcomes. No available BMD certified by the Election Assistance Commission is contestable or defensible.

Testing Cannot Tell Whether Ballot-Marking Devices Alter Election Outcomes

Philip B. Stark and Ran Xie

[1] University of California, Berkeley
[2] University of California, Berkeley

29 July 2020

Abstract. Like all computerized systems, ballot-marking devices (BMDs) can be hacked, misprogrammed, and misconfigured. BMD printout might not reflect what the BMD screen or audio conveyed to the voter. If voters complain that BMDs misbehaved, officials have no way to tell whether BMDs malfunctioned, the voters erred, or the voters are attempting to cast doubt on the election. Several approaches to testing BMDs have been proposed. In pre-election *logic and accuracy* (*L&A*) tests, trusted agents input known test patterns into the BMD and check whether the printout matches. In *parallel* or *live* testing, trusted agents use the BMDs on election day, emulating voters. In *passive* testing, trusted agents monitor the rate at which voters "spoil" ballots and request another opportunity to mark a ballot: an anomalously high rate might result from BMD malfunctions. In practice, none of these methods can protect against outcome-altering problems. L&A testing is ineffective against malware in part because BMDs "know" the time and date of the test and the election. Neither L&A nor parallel testing can probe even a small fraction of the combinations of voter preferences, device settings, ballot language, duration of voter interaction, input and output interfaces, and other variables that could comprise enough votes to change outcomes. Under mild assumptions, to develop a model of voter interactions with BMDs accurate enough to ensure that parallel tests could reliably detect changes to 5% of the votes (which could change margins by 10% or more) would require monitoring the behavior of more than a million voters in each jurisdiction in minute detail—but the median turnout by jurisdiction in the U.S. is under 3000 voters, and 2/3 of U.S. jurisdictions have fewer than 43,000 active voters. Moreover, all voter privacy would be lost. Given an accurate model of voter behavior, the number of tests required is still larger than the turnout in a typical U.S. jurisdiction. Even if less testing sufficed, it would require extra BMDs, new infrastructure for creating test interactions and reporting test results, additional polling-place staff, and more training. Under optimistic assumptions, passive testing that has a 99% chance of detecting a 1% change to the margin with a 1% false alarm rate is impossible in jurisdictions with fewer than about 1 million voters, even if the "normal" spoiled ballot rate were known exactly and did not vary from election to election and place to place. Passive testing would also require training and infrastructure to monitor the spoiled ballot rate in real time. And if parallel or passive testing discovers a problem, the only remedy is a new election: there is no way to reconstruct the correct election result from an untrustworthy paper trail. Minimizing the number of votes cast using BMDs is prudent election administration.

- Hand-marked paper ballots are a record of what the voter did.
- Machine-marked paper ballots are a record of what the machine did.
- BMDs make voters responsible for catching & correcting machine errors/bugs/hacks.
- Experiments & polling-place observations show few voters check BMD printout; fewer notice errors.

Madison, Oconee, and Oglethorpe. The study, dated January 22, 2021, was not published; its existence was discovered through a Georgia Open Records Act request by *The Atlanta Journal Constitution*.[2] Dr. Gilbert does not mention this study.

7. The results of the Haynes and Hood (2021) study are summarized in the table below. Less than 19 percent of voters looked at the BMD printout for 5 seconds or more.

Duration of glance	Percentage of voters
did not look at all	20.0 percent
less than one second	31.3 percent
one to five seconds	29.9 percent
five seconds or more	18.8 percent

County	Contests	Minimum estimated time required to *read* 4 words per contest (seconds)	included in Haynes & Hood (2021)
Barrow	21	17–37	yes
Clarke	16	13–28	yes
Jackson	24	19–42	yes
Madison	23	18–40	yes
Oconee	27	22–47	yes
Oglethorpe	19	15–33	yes
Dougherty	19	15–33	no
Fulton	20	16–35	no

Did the reported winner really win?

- Procedure-based vs. evidence-based elections
  - sterile scalpel v. patient's condition

Did the reported winner really win?

- Procedure-based vs. evidence-based elections
  - sterile scalpel v. patient's condition
- *Any* way of counting votes can make mistakes
- *Every* electronic system is vulnerable to bugs, configuration errors, & hacking

Did the reported winner really win?

- Procedure-based vs. evidence-based elections
  - sterile scalpel v. patient's condition
- *Any* way of counting votes can make mistakes
- *Every* electronic system is vulnerable to bugs, configuration errors, & hacking
- **Did error/bugs/hacking cause the wrong candidate(s) to appear to win?**

IEEE SECURITY AND PRIVACY, SPECIAL ISSUE ON ELECTRONIC VOTING, 2012. LAST EDITED 8 MAY 2012. 1

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

Voting system properties needed to justify public trust

- (Strong) Software Independence
- Contestability
- Defensibility

Voting system properties needed to justify public trust

- (Strong) Software Independence
- Contestability
- Defensibility

DREs, BMDs, online voting have none of these properties.

Risk-Limiting Audits (RLAs, Stark, 2008)

- If there's a trustworthy paper record of votes, can check whether reported winner really won.
  - Can manually count
  - If you accept a controlled *risk* of not correcting a wrong reported outcome, can save effort

A *risk-limiting audit* has a known maximum chance of not correcting the reported outcome if it's wrong & never changes correct outcomes.

A *risk-limiting audit* has a known maximum chance of not correcting the reported outcome if it's wrong & never changes correct outcomes.

Risk limit: largest possible chance of *not* correcting a wrong reported outcome, no matter where or how errors/problems occurred.

A *risk-limiting audit* has a known maximum chance of not correcting the reported outcome if it's wrong & never changes correct outcomes.

Risk limit: largest possible chance of *not* correcting a wrong reported outcome, no matter where or how errors/problems occurred.

Establishing whether paper trail is trustworthy involves other processes, generically, *compliance audits* along w/ thorough canvass, ballot accounting, pollbook/participation reconciliation, eligibility verification, demonstrably secure chain of custody, etc.

A *risk-limiting audit* has a known maximum chance of not correcting the reported outcome if it's wrong & never changes correct outcomes.

Risk limit: largest possible chance of *not* correcting a wrong reported outcome, no matter where or how errors/problems occurred.

Establishing whether paper trail is trustworthy involves other processes, generically, *compliance audits* along w/ thorough canvass, ballot accounting, pollbook/participation reconciliation, eligibility verification, demonstrably secure chain of custody, etc.

DRE & BMD printout is not trustworthy, no matter how well it's protected.

RLA pseudo-algorithm

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

RLA pseudo-algorithm

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

if (full handcount) {
    handcount result is final
}
```

The National Academies of SCIENCES ENGINEERING MEDICINE

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 AK, CA, CO, GA, IN, KS, MI, MT, NJ, OH, OR, PA, RI, WA, WY, VA, DK.
- CA counties: Alameda, El Dorado, Humboldt, Inyo, Madera, Marin, Merced, Monterey, Napa, Orange, San Francisco, San Luis Obispo, Santa Clara, Santa Cruz, Stanislaus, Ventura, Yolo.
- Routine statewide in CO since 2017. Statewide audits in AK, KS, WY in 2020.
- Laws in CA, CO, GA, NV, NJ, OH, OR, RI, TX, VA, WA

Role of math/stat

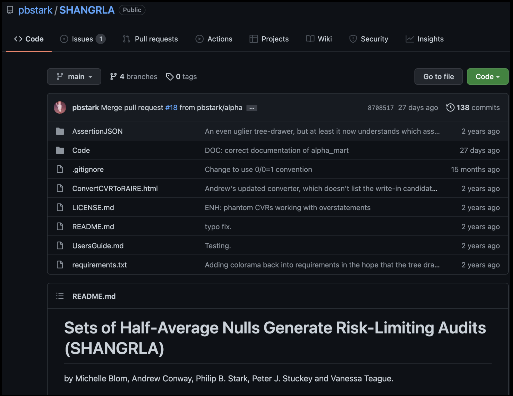
- Guarantee a large chance of correcting wrong outcomes; minimize work if reported outcome is correct.
- When can you stop inspecting ballots?
  - When there's strong enough evidence that continuing is pointless
- Frame audits as sequential hypothesis tests
  - Null hypothesis: one or more reported outcomes is *wrong*.
  - Significance level: risk limit
  - Frame hypothesis quantitatively
  - Can reduce to canonical problem: test whether mean of finite, bounded population is $\leq 1/2$

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

Philip B. Stark(✉)

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

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 nonnegative 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 nonnegative numbers. The SHANGRLA approach thus 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.



b_i is ith ballot card, N cards in all.

$$1_{\text{candidate}}(b_i) := \begin{cases} 1, & \text{ballot } i \text{ has a mark for candidate} \\ 0, & \text{otherwise.} \end{cases}$$

$$A_{\text{Alice,Bob}}(b_i) := \frac{1_{\text{Alice}}(b_i) - 1_{\text{Bob}}(b_i) + 1}{2} \in [0, 1].$$

mark for Alice but not Bob, $A_{\text{Alice,Bob}}(b_i) = 1$.

mark for Bob but not Alice, $A_{\text{Alice,Bob}}(b_i) = 0$.

marks for both (overvote) or neither (undervote) or doesn't contain contest, $A_{\text{Alice,Bob}}(b_i) = 1/2$.

$$\bar{A}^b_{\mathrm{Alice,Bob}} := \frac{1}{N}\sum_{i=1}^{N} A_{\mathrm{Alice,Bob}}(b_i).$$

Mean of a finite list of N bounded numbers.

Alice won iff $\bar{A}^b_{\mathrm{Alice,Bob}} > 1/2$.

Plurality & Approval Voting

$K \geq 1$ winners, $C > K$ candidates in all.

Candidates $\{w_k\}_{k=1}^{K}$ are reported winners.

Candidates $\{\ell_j\}_{j=1}^{C-K}$ reported losers.

Plurality & Approval Voting

$K \geq 1$ winners, $C > K$ candidates in all.

Candidates $\{w_k\}_{k=1}^{K}$ are reported winners.

Candidates $\{\ell_j\}_{j=1}^{C-K}$ reported losers.

Outcome correct iff

$$\bar{A}^b_{\mathrm{w_k},\ell_\mathrm{j}} > 1/2, \quad \text{for all } 1 \leq k \leq K, \;\; 1 \leq j \leq C - K$$

$K(C-K)$ inequalities.

Plurality & Approval Voting

$K \geq 1$ winners, $C > K$ candidates in all.

Candidates $\{w_k\}_{k=1}^{K}$ are reported winners.

Candidates $\{\ell_j\}_{j=1}^{C-K}$ reported losers.

Outcome correct iff

$$\bar{A}^b_{\mathrm{w_k},\ell_\mathrm{j}} > 1/2, \quad \text{for all } 1 \leq k \leq K, \;\; 1 \leq j \leq C-K$$

$K(C-K)$ inequalities.

Same approach works for D'Hondt & other proportional representation schemes. (Stark & Teague 2015)

Super-majority

$f \in (0, 1]$.

Alice won iff

$$\text{(votes for Alice)} > f \times \text{(valid votes for anyone)}$$

Set

$$A(b_i) := \begin{cases} \frac{1}{2f}, & b_i \text{ has a mark for Alice and no one else} \\ 0, & b_i \text{ has a mark for exactly one candidate, not Alice} \\ \frac{1}{2}, & \text{otherwise.} \end{cases}$$

Alice won iff

$$\bar{A}^b > 1/2.$$

Borda count, STAR-Voting, & other additive weighted schemes

Winner is the candidate who gets most "points" in total.

$s_{\mathrm{Alice}}(b_i)$: Alice's score on ballot i.

$s_{\mathrm{cand}}(b_i)$: another candidate's score on ballot i.

s^+: upper bound on the score any candidate can get on a ballot.

Alice beat the other candidate iff Alice's total score is bigger than theirs:

$$A_{\mathrm{Alice,c}}(b_i) := \frac{s_{\mathrm{Alice}}(b_i) - s_{\mathrm{c}}(b_i) + s^+}{2s^+}.$$

Alice won iff $\bar{A}^b_{\mathrm{Alice,c}} > 1/2$ for every other candidate c.

Ranked-Choice Voting, Instant-Runoff Voting (RCV/IRV)

2 types of assertions (Blom et al. 2018):

1. Candidate *i* has more first-place ranks than candidate *j* has total mentions.
2. After a set of candidates *E* have been eliminated from consideration, candidate *i* is ranked higher than candidate *j* on more ballots than *vice versa*.

Both can be written $\bar{A}^b > 1/2$.

Finite set of such assertions implies reported outcome is right.

More than one set suffices; can optimize expected workload.

Auditing assertions

Test *complementary null hypothesis* $\bar{A}^b \leq 1/2$ sequentially.

- Audit until either all complementary null hypotheses about a contest are rejected at significance level α or until all ballots have been tabulated by hand.
- Yields a RLA of the contest in question at risk limit α.
- No multiplicity adjustment needed.

Sequential tests (Wald, 1945) and martingales

Key object: nonnegative (super)martingale

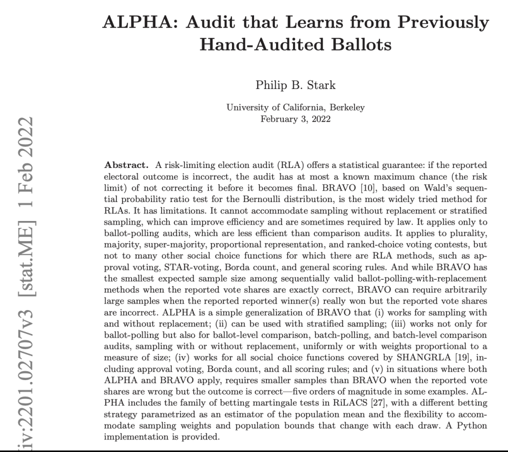
Sequence of rvs (Z_j), $j = 1, \ldots$ s.t.

- $\mathbb{E}|Z_j| < \infty$
- $\mathbb{E}(Z_{j+1}|Z_1, \ldots, Z_j) = (\leq) Z_j$
- $\mathbb{P}(Z_j \geq 0) = 1$

Ville's inequality (1939)

If (Z_j) is a nonnegative supermartingale, then for any $\alpha \in (0, 1]$ and all $J \in \{1, \ldots, N\}$,

$$\Pr\left(\max_{1 \le j \le J} Z_j \ge 1/\alpha\right) \le \alpha \, \mathbb{E}|Z_J|.$$



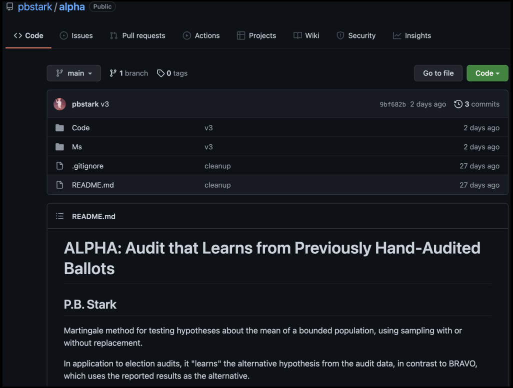
ALPHA: Audit that Learns from Previously Hand-Audited Ballots

Philip B. Stark

University of California, Berkeley

February 3, 2022

Abstract. A risk-limiting election audit (RLA) offers a statistical guarantee: if the reported electoral outcome is incorrect, the audit has at most a known maximum chance (the risk limit) of not correcting it before it becomes final. BRAVO [10], based on Wald's sequential probability ratio test for the Bernoulli distribution, is the most widely tried method for RLAs. 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, super-majority, proportional representation, and ranked-choice voting contests, but not to many other social choice functions for which there are RLA methods, such as approval voting, STAR-voting, Borda count, and general scoring rules. 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, BRAVO 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; (ii) can be used with stratified sampling; (iii) works not only for ballot-polling but also for ballot-level comparison, batch-polling, and batch-level comparison audits, sampling with or without replacement, uniformly or with weights proportional to a measure of size; (iv) works for all social choice functions covered by SHANGRLA [19], including approval voting, Borda count, and all scoring rules; and (v) 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 [27], with a different betting strategy parametrized as an estimator of the population mean and the flexibility to accommodate sampling weights and population bounds that change with each draw. A Python implementation is provided.



ALPHA (Stark, 2022)

Test $\theta \le \mu$ against the alternative $\theta > \mu$.

- $X^j := (X_1, \ldots, X_j)$; $X_i \in [0, u_i]$.
- $\mu_j := \mathbb{E}(X_j | X^{j-1})$ computed under the null $\theta = \mu$.
- $\eta_j = \eta_j(X^{j-1})$, $j = 1, \ldots$, a *predictable sequence*: can't depend on X_k for $k \ge j$.

$T_0 := 1$;

$$T_j := T_{j-1} u_j^{-1} \left(X_j \frac{\eta_j}{\mu_j} + (u_j - X_j) \frac{u_j - \eta_j}{u_j - \mu_j} \right), \quad j = 1, \ldots . \tag{1}$$

(T_j) is a nonnegative supermartingale w/ expected value ≤ 1 if $\theta \le \mu$.

Thus if $\theta \le \mu$,

$$\mathbb{P}\{\max_j T_j \ge 1/\alpha\} \le \alpha.$$

- Set audit parameters
  - risk limit $\alpha \in (0, 1)$; # cards N, sampling method, u_i, η_0
  - Pick $\eta(i, X^{i-1}) \in (\mu_i, u_i]$, where $\mu_i := \mathbb{E}(X_i | X^{i-1})$ is computed under the null.
- Initialize variables
  - $j \leftarrow 0$: sample number
  - $T \leftarrow 1$: test statistic
  - $S \leftarrow 0$: sample sum
  - $m = 1/2$: population mean under the null
- While $T < 1/\alpha$ and not all ballot cards have been audited:
  - draw a ballot card at random
  - $j \leftarrow j + 1$
  - determine X_j by applying assorter to selected card
  - if $m < 0$, $T \leftarrow \infty$; else $T \leftarrow T u_j^{-1} \left(X_j \frac{\eta(j,S)}{m} + (u_j - X_j) \frac{u - \eta(j,S)}{u_j - m} \right)$;
  - $S \leftarrow S + X_j$
  - if sampling w/o replacement, $m \leftarrow (N/2 - S)/(N - j + 1)$
  - if desired, break & conduct a full hand count

Comparison audits

Use system's interpretation of individual ballots or batches of ballots.

Like checking an expense report.

b_i is ith ballot, c_i is cast-vote record for ith ballot.

A an assorter.

overstatement error for ith ballot is

$$\omega_i := A(c_i) - A(b_i) \leq A(c_i) \leq u,$$

where u is an upper bound on the value A assigns to any ballot card or CVR.

$v := 2\bar{A}^c - 1$, *reported assorter margin*.

$B(b_i, c) := (1 - \omega_i/u)/(2 - v/u) > 0$, $i = 1, \ldots, N$.

B assigns non-negative numbers to ballots.

Reported outcome correct iff

$$\bar{B} > 1/2.$$

Stratified sampling

Cast ballots are partitioned into $S \geq 2$ *strata*.

Stratum s contains N_s cast ballots.

Let $\bar{A}_s^b$ denote the mean of the assorter applied to just the ballot cards in stratum s. Then

$$\bar{A}^b = \frac{1}{N} \sum_{s=1}^{S} N_s \bar{A}_s^b = \sum_{s=1}^{S} \frac{N_s}{N} \bar{A}_s^b.$$

Can reject the hypothesis $\bar{A}^b \leq 1/2$ if we can reject the hypothesis

$$\bigcap_{s \in S} \left\{ \frac{N_s}{N} \bar{A}_s^b \leq \beta_s \right\}$$

for all $(\beta_s)_{s=1}^S$ s.t. $\sum_{s=1}^S \beta_s \leq 1/2$.

Union-Intersection Test

Fisher's Combining Function

$\{P_s(\beta_s)\}_{s=1}^S$ are independent random variables.

If $\bigcap_{s \in S} \left\{ \frac{N_s}{N} \bar{A}_s^b \leq \beta_s \right\}$, distribution of

$$-2 \sum_{s=1}^{S} \ln P_s(\beta_s)$$

is dominated by chi-square distribution with $2S$ degrees of freedom.

Low-dimensional optimization problem to maximize P-value over $(\beta_s)_{s=1}^S$.

Sample design

- individual ballots?
- groups of ballots?
- stratify? (law, logistics, equipment capabilities, ...)
- sampling probabilities?
- w/ replacement? w/o replacement? Bernoulli?
- fully sequential? escalation schedule?

Open research questions

- What is the class of social choice functions that can be audited with SHANGRLA?
- If there are sufficient conditions, are there also necessary and sufficient conditions?
- Are all sets of necessary and sufficient conditions equally expensive to audit?
- Can "round-by-round" sampling reduce sample sizes?

Wrinkles

- $\sim$20% of U.S. voters don't vote on paper
- States adopting universal-use BMDs: paper trail hackable/untrustworthy
- inadequate rules for chain of custody, ballot accounting, pollbook reconciliation, eligibility verification, ...
- need transparent high-quality randomness
  - public ceremony of die rolls, published crypto-quality PRNG
- missing ballots; imperfect manifests (Bañuelos & Stark 2012)
- producing CVRs linked to ballots while preserving vote anonymity; redacted CVRs
- preserve privacy but ensure the public can confirm audit didn't stop too soon

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

PHILIP B. STARK and FREE SPEECH FOR PEOPLE, *Plaintiffs*, v. UNITED STATES ELECTION ASSISTANCE COMMISSION, *Defendant*.	Civil Action No. 1:21-cv-01864 (CKK)

PLAINTIFFS' MEMORANDUM OF POINTS AND AUTHORITIES IN OPPOSITION TO DEFENDANT'S MOTION TO DISMISS OR, IN THE ALTERNATIVE, FOR PARTIAL SUMMARY JUDGMENT

Open-source software

- auditTools
- ballotPollTools
- SUITE
- SHANGRLA
- ALPHA
- Arlo

Evidence-Based Elections: 3 C's

- Voters *CREATE* complete, durable, verified audit trail.

Evidence-Based Elections: 3 C's

- Voters *CREATE* complete, durable, verified audit trail.
- LEO *CARES FOR* the audit trail adequately to ensure it remains complete and accurate.

Evidence-Based Elections: 3 C's

- Voters *CREATE* complete, durable, verified audit trail.
- LEO *CARES FOR* the audit trail adequately to ensure it remains complete and accurate.
- Verifiable audit *CHECKS* reported results against the paper


411412
511221
411413
521225
411418
561341
421224
421422
411221
421222
511211
421225
421226
421426