The Decisive Power of Indecision: Low-Variance Risk-Limiting Audits and Election Contestation via Marginal Mark Recording

Benjamin Fuller, Rashmi Pai, and Alexander Russell,
University of Connecticut - Voting Technology Research Center

https://www.usenix.org/conference/usenixsecurity24/presentation/fuller

This paper is included in the Proceedings of the
33rd USENIX Security Symposium.

August 14–16, 2024 • Philadelphia, PA, USA

978-1-939133-44-1

Open access to the Proceedings of the
33rd USENIX Security Symposium
is sponsored by USENIX.

The Decisive Power of Indecision: Low-Variance Risk-Limiting Audits and Election Contestation via Marginal Mark Recording

Benjamin Fuller, Rashmi Pai, Alexander Russell
University of Connecticut – Voting Technology Research Center
{benjamin.fuller,rashmi.pai,alexander.russell}@uconn.edu

Abstract

Risk-limiting audits (RLAs) are techniques for verifying the outcomes of large elections. While they provide rigorous guarantees of correctness, widespread adoption has been impeded by both efficiency concerns and the fact they offer statistical, rather than absolute, conclusions. We attend to both of these difficulties, defining new families of audits that improve efficiency and offer qualitative advances in statistical power.

Our new audits are enabled by revisiting the standard notion of a cast-vote record so that it can declare multiple possible mark interpretations rather than a single decision; this can reflect the presence of marginal marks, which appear regularly on hand-marked ballots. We show that this simple expedient can offer significant efficiency improvements with only minor changes to existing auditing infrastructure. We consider two ways of representing these marks, both yield risk-limiting comparison audits in the formal sense of Fuller, Harrison, and Russell (IEEE Security & Privacy 2023).

We then define a new type of post-election audit we call a *contested audit*. These permit each candidate to provide a cast-vote record table advancing their own claim to victory. We prove that these audits offer remarkable sample efficiency, yielding control of risk with a constant number of samples (that is independent of margin). This is a first for an audit with provable soundness. These results are formulated in a game-based security model that specify quantitative soundness and completeness guarantees. These audits provide a means to handle contestation of election results affirmed by conventional RLAs.

1 Introduction

Risk-limiting audits (RLAs) are methods for verifying the outcome of a large-scale election. Developed by the academic and election community over that last two decades [1], RLAs offer significant efficiency improvements over the burden of a full hand recount while providing rigorous guarantees of correctness. They can also support a variety of social choice functions and election organizations (see Section 1.3). The framework posits reasonable physical assumptions under which elections—involving diverse interactions among multiple untrusted human and electronic processes—can be authoritatively protected from errors or deliberate falsification in tabulation, vote aggregation, and reporting. Finally, the framework supports *public verification*: third-party observers can certify the outcome of the audit [2].

RLAs are only defined in settings with voter-verified ground truth, typically furnished by hand-marked paper ballots.[1] Such marked ballots determine the *ground-truth outcome of the election*; of course, any particular tabulation of the ballots—generally carried out by electronic tabulators—determines a potentially different outcome [1]. RLAs are designed to detect disagreement between the ground-truth outcome and a given tabulated outcome except with a prescribed, concrete failure probability called the "risk" of the audit. The word "outcome" here refers to the winner(s) of the election rather than the exact vote totals.

RLA adoption in large democracies has been uneven. In the United States, RLAs have been advocated by area experts [2], the US Senate, and the 2014 Presidential Commission on Election Administration. However, only a handful of states run such audits, with most adoption occurring in the last few years. A major obstacle to widespread adoption is efficiency: RLAs require manual interpretation of a sampled set of ballots whose size grows as a function of the margin of the race to be audited. Thus, circumstances with small margins may require examination of many thousands of ballots and, unfortunately, election logistics and planning must prepare for such onerous outcomes. (As an example, a conventional ballot polling audit typically requires over 10,000 ballot samples for a 2% margin and 50,000 samples at 1% margin [3].) A related challenge is that high-efficiency RLAs necessarily provide (only) statistical guarantees; common risk values in practice are on the order of 5% (3% in Colorado, 5% in Georgia, 9% in Rhode Island, 10% in Virginia). This invites an immediate criticism:

[1] Verified Voting summarizes the adoption of such methods in the U.S.

a losing candidate whose loss has been reaffirmed by an RLA may argue that a 1-in-20 chance of error is unsatisfactory. We engage with both of these concerns by defining two new families of post-election audits.

1.1 Our Results in Brief

Our new approaches are enabled by revisiting the conventional understanding of a cast-vote record (CVR), a declaration of how a ballot was counted in an election. CVRs are critical elements of standard "comparison" audits, which are the starting points for our new audits. We redefine CVR semantics to include *explicit indication of marginal marks* [4] by either asserting a collection of possible interpretations for a marginal mark or a predicted probability distribution of interpretations. Marginal marks are those for which voter intent is unclear. Aside from partially filled voting targets, this can include "small lines, atypical ink or marker, and marks outside but near the voting target" [4]. We leverage this declaration of indecision to define two new families of audits.

Ballot-comparison risk-limiting audits with reduced variance and improved efficiency. Recording a *predicted probability distribution* for the interpretation of marginal marks in a CVR and reflecting this appropriately in the audit leads to a new family of *Bayesian risk-limiting audits*, which can offer significant efficiency advantages and improved confidence in the audit. In particular, in settings of practical interest this reduces the standard deviation of completion times for standard sequential statistical testing methods and hence reduces the number of ballots that need to be drawn in the first (and typically only) phase of standard multi-phase audits. For margins of 1%, this reduces the number of ballots sampled by over 10%; see Table 2. We also discuss such audits in a less expressive "conservative" setting where the CVR simply lists a set of possible interpretations. (One could consider interpretations with some minimum probability as shown in Figure 1.) These results are articulated in an adaptation of the formal game-based model of Fuller, Harrison, and Russell [5].

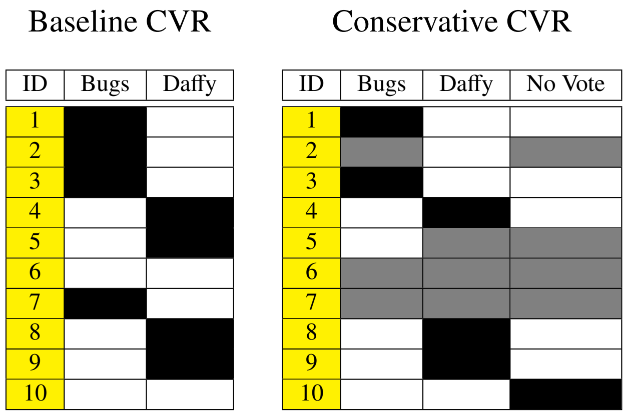
Aside from these efficiency improvements, such CVRs provide an improvement in voter confidence and interpretability by explaining discrepancies. Specifically, conventional CVRs will frequently disagree with marginal marks as, of course, there is no authoritative conclusion for such a mark; this can cast doubt on the CVR and the audit. A CVR permitted to assert the ambiguity of a mark provides an immediate means for separating the common circumstances where the mark is ambiguous from more disturbing discrepancies arising from real failures of the audit (see further discussion in Section 2). We remark that Clearballot publishes mark images with predicted probability distributions and allows citizens to search for marginal marks; thus our method makes use of data that is already public in many jurisdictions (Clearballot indicates that 13 US states use their technology).

***Competitive audits* involving multiple CVRs submitted by multiple interested parties, yielding a natural approach to *election contestation*.** We define a novel notion of a "competitive audit," in which multiple parties submit CVRs to substantiate competing claims of victory. We develop a formal cryptographic game-style framework for defining and analyzing such audits. This modeling permits a strong adversary that may prepare all but one CVR, choose which interpretation to return for a marginal mark, and whether to suppress a sampled ballot. We then show that this model leads to an *extraordinarily efficient audit calling for evaluation of a fixed, constant number of ballots*, independent of margin. Such strong efficiency guarantees are impossible for conventional RLAs. This new auditing framework provides an approach to the problem of election contestation, where the results of an RLA are challenged by a third party. We discuss practical considerations for deploying these audits in Section 2.

1.2 A Detailed Survey of the General Framework, Related Work, and Our Results

An RLA is carried out in the context of an election with voter-verified ground truth. For concreteness, our discussion assumes hand-marked paper ballots, which is the most common instantiation in practice. The audit is intended to detect disagreement between the ground-truth outcome of the election, as determined by the ballots themselves, and the *declared* outcome of the election resulting from a tabulation of the ballots; we use the word "declared" throughout to emphasize that the tabulated outcome may be inconsistent with the ground-truth outcome. Such an audit examines a collection of physical ballots, usually drawn at random according to a convention depending on the audit, and either concludes with CONSISTENT, indicating that the sampled ballots appear consistent with the declared outcome, or INCONCLUSIVE, indicating that the audit has not amassed sufficient evidence of consistency. An audit is said to have *risk* α (a number in the range $[0,1]$) if the probability that it outputs CONSISTENT is no more than α when the declared outcome contradicts the ground-truth outcome. INCONCLUSIVE outcomes usually call for a full hand recount; see the next subsection.

A trivial audit that always outputs INCONCLUSIVE has zero risk, but is obviously not useful in practice. This points to the importance of *completeness*—the additional guarantee that the audit outputs CONSISTENT with high probability under favorable circumstances; we discuss this in detail below. A straightforward hand recount of an election is itself a RLA audit with zero risk (modulo any errors in manual interpretation). More sophisticated RLAs randomly draw and inspect a collection of ballots to make statistical conclusions about the outcome. A general feature of such statistical audits is that the number of samples scales with the margin of victory; this reflects the natural intuition that a landslide should be easier to statistically verify than a victory by slim margin. In the most familiar setting of a simple, two-candidate, first-past-the-



Bayesian CVR

ID	Bugs	Daffy	No Vote
1	1		
2	.72	.02	.26
3	.99		.01
4		1	
5		.75	.25
6	.12	.38	.5
7	.46	.1	.44
8		1	
9		1	
10		.02	.98

Figure 1: Generalization of conventional CVRs to the conservative setting, where multiple interpretations are declared, and the Bayesian setting, where one associates probabilities with each interpretation. Blanks cells in the Bayesian indicate probability 0; for brevity, the "No Vote" columns reflect both the no-mark case and the overvote case (with marks for both candidates).

post election, the empirical mean of the number of votes for a particular candidate among a collection of $\log(2/\alpha)/(2\epsilon^2)$ uniformly sampled ballots is correct to within ϵ with probability $1-\alpha$. This simple procedure—called "ballot polling" in the literature—yields a risk-limiting audit that inspects

$$O\left(\frac{\log(1/\alpha)}{\mu^2}\right) \tag{1}$$

ballots in expectation, where μ is the actual margin [1].

The most efficient known approach for RLAs, known as a *ballot comparison audit* [1, 6–12], requires a detailed tabulation report called a cast-vote record table (CVR). A CVR declares a unique identifier and interpretation (that is, a determination of the cast votes) for each ballot in the election. This ballot-by-ballot record of the election naturally determines a tabulation and outcome. It also enables high-efficiency audits that proceed by iterating the following experiment: (1) The auditor picks a random identifier ι from the CVR table; (2) the auditor finds a physical ballot with identifier ι; (3) the auditor compares the interpretation of the physical ballot against the associated CVR entry. Intuitively, iterations where the physical ballot matches the CVR provide support for this declared record of the election, while disagreements erode that support. Continuing the discussion in the same two-candidate first-past-the-post setting, observe that if the CVR table declares an incorrect outcome for the election, then at least a $\mu/2$ fraction of entries of the CVR table must disagree with the corresponding physical ballot; as above, μ is the scaled margin of ground-truth victory. (The factor of two arises because a single disagreement can both negate a vote for the declared winner and supply a vote for the declared loser.) It follows that after $2/\mu$ samples we expect to observe a discrepancy if the outcomes differ, suggesting that this approach should yield a risk-limiting audit with sample size

$$O\left(\frac{\log(1/\alpha)}{\mu}\right). \tag{2}$$

(The $\log(1/\alpha)$ term increases the number of samples so that the probability of observing an inconsistency is driven to at least $1-\alpha$.) It is instructive to compare this against (1) above. Due to the improved dependence on μ, in circumstances with small margins ballot comparison beats ballot polling.

The completeness challenge. The efficiency landscape is more complicated than the discussion above (and the asymptotic expression (2)) indicates. The complication is the requirement of completeness—that is, that the audit conclude with CONSISTENT when considering an accurate CVR under favorable conditions. In practice, when a post-election RLA concludes with INCONCLUSIVE, policy calls for either a complete hand count or an additional stage of auditing (with a smaller risk parameter to avoid significant amplification of risk from the composite audit). See for example, Rhode Island Code 17-19-37-4 and Colorado's fact sheet. Thus a full accounting of efficiency must weigh the costs and likelihood of INCONCLUSIVE conclusions in the best-effort settings that arise in practice.

To be more concrete, a ballot-comparison auditor simply wishing to meet a prescribed risk limit α could adopt the framework above with the convention that *any* observed disagreement between the CVR and a physical ballot causes the auditor to immediately terminate the audit and output INCONCLUSIVE. If examination of $\log(1/\alpha)/(2\mu)$ ballots indeed exposes no inconsistencies, the audit can safely return CONSISTENT with only an α probability of error, as desired. For margin $\mu = .01$ and risk $\alpha = .05$, for example, this requires inspection of 499 ballots. Now, if this "strict" auditor has the luxury of unambiguous ground truth, a perfectly correct CVR, and consistently accurate ballot interpretations, then it will indeed conclude the audit with CONSISTENT. However, discrepancies between the CVR and physical ballots arise frequently in practice. (For example, the comparison RLA run by the US state of Colorado for the 2020 presidential election generated hundreds of discrepancies [13].) These discrepancies can result from truly ambiguous voter intent, disagreement among the interpretations of an auditing board, or operational errors such as retrieval of an incorrect ballot. Estimates from prior work (e.g. Stark's audit tools; [14]) suggest discrepancy rates as large as 0.5%, though the actual rate of incidence depends on various features of the election.

These considerations have led to the development and adoption of audits (and associated statistical tests) that can tolerate such anticipated errors. For example, adopting the same parameters (5% risk, 1% margin) but with the introduction of a .01% rate of marks being inadvertently dropped during interpretation, the standard Kaplan-Markov test calls for a sample size of 1220 ballots in order to terminate with probability 95%.[2] This is over twice the sample size arising with no errors (499, as indicated above) and illustrates that even small error rates have a significant impact on required sample size; we remark, additionally, that the incidence of sporadic errors in an otherwise correct CVR or errors in ballot interpretation are the cause of variance in the stopping time for such sequential statistical tests. As audits are typically conducted in phases—drawing a collection of ballots to be inspected at once and calibrated so that each phase is likely to be the last—variance also directly influences efficiency as it forces large phases.

These concerns motivate our exploration of more expressive CVRs. In particular, we note that the party (or tabulator) preparing the CVR can identify marginal marks during tabulation that are likely to lead to errors during a comparison audit. Reflecting these marginal marks on the CVR offers a new dimension of optimization for the auditing rule. As a matter of bookkeeping, we will consider CVRs that may list multiple interpretations for each ballot or a full predicted probability distribution of interpretations. (A traditional CVR corresponds to the case when the size of each set of interpretations is 1.) We apply this to develop both conventional audits with improved efficiency and define a new class of competitive audits, providing both high efficiency and a mechanism for contesting conventional audits.

Enriching CVRs to reflect marginal marks can also improve confidence in the audit, as it can distinguish avoidable discrepancies arising from ambiguity from unexplained auditing failures; see Section 2.

1.2.1 Improved Efficiency Risk-Limiting Audits via Bayesian CVRs

We show how to adapt standard ballot comparison audits to the setting of *Bayesian CVRs* that declare a predicted distribution of ballot interpretations. We indicate how to formulate the standard notions of discrepancy and margin in this setting and observe that this simple change has significant efficiency ramifications. Using discrepancy and marginal rates from practice we show that such audits reduce sample size standard deviation by 14% compared to a traditional audit at a margin of 1% (see Table 2). We also develop a variant of this audit that uses a CVR that merely reflects the *possible* interpretations of each ballot (rather than a full probability distribution). We call this a *conservative CVR*. While this does not provide the full benefits of the Bayesian approach, it may be simpler to deploy in practice since no probability estimation is required. It also considers a stronger adversarial model where the ballot interpretation of the auditor is adversarially controlled.

We adopt to our setting the cryptographic game-style modeling for RLAs developed in [5], and prove that the new audit is risk-limiting. We then carry out simulations to show that the most popular sequential statistical test used for RLAs (the Kaplan-Markov test) can take advantage of the resulting notion of discrepancy to provide more efficient RLAs.

1.2.2 Competitive Audits and Election Contestation

Our second contribution is a new class of post-election audits that we call *competitive RLAs*. In a competitive RLA each candidate has an "advocate" that is allowed to examine ballots (under supervision to ensure ballots are not destroyed or modified). For each candidate $\mathsf{A}_1, \dots, \mathsf{A}_k$, their advocate produces a CVR that may declare different ballot interpretations and winners; the intent is that the advocate for candidate A files a CVR that is favorable to A. In particular, if A won the election, the CVR (filed by the advocate for A) should establish that; in a weaker sense, even if A did not win, the advocate may wish to file a CVR that prevents other candidates that did not win the election from convincingly claiming that they did. For simplicity, we formulate this audit in the "conservative" setting, where CVRs may list multiple interpretations for each ballot but do not record a probability distribution on these various possible interpretations. (It is certainly possible to formulate this in the setting with a Bayesian CVRs, see end of Section 6.) For elections with very small margins, an interpretation of marginal votes may be necessary to determine which candidate receives the most votes. We are only able to prove results about candidates whose winning and losing status is true regardless of interpretation of marginal marks. Definitionally, an election thus at most one winner, and some candidates may neither win nor lose.

Assuming that at least one submitted CVR only lists interpretations that are consistent with the ballots and the election has a winner, the auditor will identify the correct winner using a constant number of ballots. Furthermore, if some submitted CVR lists the exact list of possible interpretations for each ballot, no losing candidate can be identified as the winner. We emphasize that the number of ballot samples required by this approach is a constant that not scale with the margin. (The constant itself is a function of how robust the audit should be in the face of errors by the advocates.)

[2] This estimate uses McBurnett's tool rlacalc with standard rates of errors in the literature drawn from Lindeman and Stark [1]. These are $o_1, u_1 = .001$ and $o_2, u_2 = .0001$, see discussion in Section 5.

The basic idea of the audit is straightforward: Faced with a pair of contradictory CVRs—cvr_1 and cvr_2—there must be at least one ballot on which these CVRs completely disagree: examining this single ballot then provides an immediate strike against one of CVRs. Indeed, a strict auditor can use a generalization of this approach to settle the competing claims made by a collection of k CVRs with just $k-1$ ballots. (That is, the auditor immediately removes from consideration any CVR found to be inconsistent with an inspected ballot.) Such a stringent auditor will not be appropriate in practice, which must account for a small number of errors on the part of the advocates (and the cvr_i) or the possibility that ballots are lost, say, during the audit. Even accounting for such phenomena, our approach yields audits with remarkable efficiency. (As mentioned earlier, the complexity of the audit scales with the desired robustness in the face of such advocacy errors.)

The most natural and practical instantiation of our techniques is providing an efficient adjudication process for disagreements that result after the parallel scanning of ballots, for example, using the OpenScan system [15]. The idea of the OpenScan system is for advocates to place cameras above a physical scanner and create their own CVRs without having physical access to the ballots. Their system does not have a natural adjudication mechanism. Competitive audits provide that adjudication with a constant query ballot complexity.

Election contestation. Competitive RLAs provide an immediate approach to election contestation. Specifically, consider a situation where a candidate wishes to dispute the conclusions of conventional RLA (that is, the candidate claims that an invalid election was not detected by the RLA). In this case, the candidate (or their advocate) can be permitted to examine the ballots and produce a competing CVR for the election (that, presumably, will claim that the candidate is the victor). The CVR originally produced for the election (and reaffirmed by the original RLA) and the candidate's CVR can then be treated by the competitive audit mechanism. This has a number of notable features: (i.) for some integer t, the audit can rigorously settle the dispute with risk $2^{-\theta(t)}$ after examining only t ballots—thus a small constant number of queries, independent of margin, are sufficient; (ii.) this provides an interesting avenue for election policy: for example, the cost and effort of the audit could be borne by the contesting party. See more discussion in Section 2.

Adversarial modeling. We consider a cryptographic game-based model for analysis. The model places the ballots (and, in conservative and competitive audits, the choice of the returned interpretation) in the hands of an adversarial environment, thus reflecting various adverse circumstances that challenge practical audits. Our treatment universally quantifies over the ground-truth results (and, in particular, the marks appearing on the ballots themselves) and tabulated cast-vote records; in this sense they apply to all elections. We remark that in terms of the formal modeling any universally quantified object can be seen as adversarially controlled.

1.3 Other Related Work

Stark [16] defined RLAs in 2008. The framework was generalized and sharpened over the next decade [1, 17–19] leading to a variety of specific methods, often optimized certain practical settings [1, 7, 16, 20–23]. There are four major directions in current research (1) supporting varied social choice functions [10, 12], (2) managing multiple races [6–9, 24], (3) sharpening risk estimates [8, 11, 16, 20, 21, 25–29] and (4) implementation issues [3, 5, 17–19, 30–33].

The most closely related work to our competitive setting is a recent, independent article of Jones et al. [34]. One can view our competitive audits as an analog of a "multi-prover interactive proof system" consisting of competing and potentially malicious "provers" who are attempting to convince a "verifier" of the truth (or falsity) of a statement (reflecting the complexity-theoretic setting of [35, 36]). Jones et al. were inspired by the "cooperative" version of this same framework in which a collection of cooperating, but potentially malicious, provers attempt to convince a verifier of the truth of a statement; in fact, this cooperative version is far more common in the complexity-theoretic literature (reflecting the standard setting of [37]). Jones et al. develop an auditing approach designed to catch inconsistencies across multiple scans used to create (multiple) CVRs. In our auditing context, their techniques focus on detecting circumstances where at least one of the scans (provers) is dishonest. Our techniques focus on deciding who to trust in a situation where at least one scan is honest. The two approaches can thus naturally compose.

Organization. We discuss practical deployment considerations in Section 2. We then cover notational preliminaries and define CVRs and ballots with multiple interpretations in Section 3, cover Bayesian audits in Section 4, and Conservative audits in Section 4.1. Finally, we evaluate efficiency of the Bayesian and conservative methods in Section 5 and describe competitive audits in Section 6. Proofs are relegated to the online version [38].

2 Practical Deployment Considerations

Bayesian and conservative audits The deciding factor in adopting RLA modifications is how they impact the balance of trust in the election outcome and auditing effort. The essential trust assumptions of these new audits—that physical ballots faithfully represent voter intent and are available for interrogation—are identical to classical risk-limiting audits. However, we note that Bayesian and conservative audits have the advantage that marginal marks are directly reflected in the cast-vote record, a benefit for public perception and interpretability. For years, Colorado has published a discrepancy report [13] explaining their guess of each nonzero discrepancy to improve public confidence in the audit. In particular, marginal marks which would have resulted in a consistency failure in a conventional audit are now explicitly indicated and reconciled with ballots; thus they serve to add to public confidence in the audit rather than detract from it. As for public interpretability, conservative audits appear superior to Bayesian audits in the sense that the straightforward role played by a "possible" interpretation in a conservative audit is clearly preferable from the standpoint of intelligibility to the fractional discrepancies arising from "likelihood" estimates. As conservative audits offer most of the statistical improvements of Bayesian audits in our numerical estimates, their relative simplicity may outweigh the small improvements in efficiency. We remark that these new audits can rely on the same statistical tests that underly conventional comparison audits.

As for the challenge of transitioning procedures and equipment to support these audits, the substantive changes pertain to tabulation equipment (responsible for producing the CVRs) and RLA software. In particular, the task of the "audit board" (which evaluates the marks on sampled ballots) is identical to that of a conventional audit: they are still responsible for settling on a single interpretation for each sampled ballot. Both audits demand a richer convention for CVR, and hence changes to tabulator output and RLA administration software. As modern tabulators typically collect full ballots scans, the changes required to generate such CVRs would require only a firmware upgrade (rather than hardware changes). Again, the task here may be simplified by the conservative setting, which requires no explicit likelihood estimation. Changes to RLA administration software are minimal, especially considering that existing statistical tests can be applied. Our simulation [39] builds on Harrison's implementation with minimal changes [40]. We remark that the recent VVSG 2.0 federal voting system guidelines require ballot imprinting, even on voter-facing tabulators. This is an important enabling feature for ballot comparison audits, so adoption of these guidelines will significantly improve the general feasibility of comparison audits in the United States.

Compare the cost of conservative audits against the natural approach that calls for auditors to adjudicate every marginal ballot as a part of the initial tabulation. As an illustrative example, Louisiana is the 25th state by population with roughly 2M votes for president in 2020. Using the .5% marginal mark rate [14], one would have to adjudicate 10K ballots, in addition to what is required for the RLA. At 1% margin ballot comparison requires examining 1K ballots (Table 2).

The competitive setting. Competitive audits have roles and procedures that have no analog in conventional RLAs. Thus, they demand careful consideration before deployment. We outline two different implementations and discuss the trust and efficiency tradeoffs. The first implementation we discuss provides all of the guarantees—including software independence—of a classical RLA while providing an efficient contestation mechanism under two assumptions:

(1) That there is an actual winner (regardless of how marginal marks are interpreted), and
(2) An advocate for the winning party contests an incorrect outcome.

The contestation phase has a risk that reduces exponentially in the number of ballots pulled. We also discuss standalone instances of competitive audits, including an existing framework that supports CVR creation by candidates without directly handling ballots.

Consider the following RLA procedure with an additional contestation phase:

(1) The state's tabulators are run on the ballots, producing $\mathsf{cvr}_{\mathtt{state}}$ that declares candidate W to be the winner.
(2) A conventional ballot comparison RLA is run on $\mathsf{cvr}_{\mathtt{state}}$ with risk α.
(3) If the RLA outputs Inconclusive, existing legal steps are followed such as a recount.
(4) Otherwise, candidates can contest the results of the audit. (Whether such contestations are permitted, or what evidence might be necessary to justify them, is a matter of election law.)
(5) A candidate A permitted to contest the election may inspect the physical ballots (under appropriate supervision to prevent tampering) to assemble $\mathsf{cvr}_{\mathsf{A}}$, an alternate CVR. Various procedures could be used to create $\mathsf{cvr}_{\mathsf{A}}$; for example, (i.) a complete hand count of all ballots, (ii.) use of alternative tabulators trusted by the candidate, (iii.) a hybrid approach, where the candidate uses $\mathsf{cvr}_{\mathtt{state}}$ with the exception of a few precincts of the election that they believe are suspicious; these are hand counted.
(6) Finally, the competitive audit is run on all of the available CVRs (i.e., $\mathsf{cvr}_{\mathtt{state}}$ and $\mathsf{cvr}_{\mathsf{A}}$ for any candidate A that assembled a CVR).

This audit provides the guarantees of a conventional RLA and provides a rigorous approach to contestation. The audit places the vast majority of the work of the competitive audit on the party that believes there is an error and wishes to correct it. The effective risk of the combined audit can be driven to a small constant, e.g., one part in 10,000, with a small number of samples (assuming candidates can create accurate CVRs). We remark that such an audit continues to protect a third party (with only a vested interest in the correct outcome of the election) from an incorrect result with risk α; in particular, when the state's CVR correctly asserts the correct winner, the contestation phase cannot change the outcome. On the other hand, an interested party (wishing to ensure that a candidate of interest is not incorrectly declared a loser of the election) can guarantee risk close to zero. It is a legal matter when to allow a contestation, our method provides a tool to decide when an advocate has gathered sufficient evidence.

One can also consider a standalone competitive RLA. The procedure follows the prescription above without the conventional RLA. (As with the audit above, supervision is required to protect ballot integrity.) If candidate A is the winner and indeed submits a correct $\mathsf{cvr}_{\mathsf{A}}$, A will win the competitive audit regardless of the procedures or software adopted by other candidates (or state) to assemble their CVRs. We remark that such tabulations significantly change the guarantees offered to a third party with a vested interest only in the correct outcome of the election. With this audit, such a third party can only be guaranteed a correct outcome if the state's tabulation is correct or the true winning candidate participates in the competitive audit.

Such competitive audits provide a rigorous approach to reconciling the results generated by the existing Openscan [15] system. This system permits each candidate to install their own camera over a scanning bed on which each ballot is rested for data collection. Adapting slightly to our framework, advocates do not touch ballots, but can be in the room to ensure each ballot is scanned exactly once. This has the advantage that all data is collected in one pass, and advocates use their own hardware and software to collect and process ballot images. Of course, the state itself may have an installed camera that produces an official outcome. The competitive audit now provides a rigorous guarantee to each candidate: if they won the election (and themselves generate a faithful CVR), they can be guaranteed victory independent of the software used by other candidates or the state. This assumes, of course, that each physical ballot is processed by their camera.

The above deployments require physical access to ballots—this is unavoidable for software-independent guarantees. In either deployment, candidates shoulder auditing responsibilities (borne entirely by the official auditors in a traditional RLA). This is one advantage of the first deployment scenario: it limits candidate effort to only those circumstances where they suspect that a traditional RLA has failed. However, competitive audits introduce a potential inequity: Candidates with fewer financial resources may have less ability to muster a full retabulation for the competitive audit.

Our results assume a low rate of disagreement between the "honest" CVR and the set of interpretations that could be reasonably inferred by an audit board. Preparing a CVR with a low error rate may require knowledge on mark adjudication and audit board behavior.

Audit	Notation	Meaning
All	$\mathbf{b}$	Physical ballot
	$\mathbf{B}$	Ballot family
	$\mathbf{B}_\iota$	Ballots with identifier ι
	S	Number of ballots in election
	C	Candidates in the election
	α	Risk of audit
	γ	Inflation factor of Kaplan-Markov when discrepancy encountered
	E	Election
Bayes.	$\mathcal{T}_{\mathbf{b}}$	Probability distribution over auditor interpretation of ballot
	$\mathrm{I}_{\mathbf{b}}$	R.V. with distribution $\mathcal{T}_{\mathbf{b}}$
	$\mathsf{T}_{\mathbf{b}}^{\LEFTcircle}(\mathsf{A})$	Expected number of votes for ballot $\mathbf{b}$ received by candidate A if audited
	$E^{\LEFTcircle}(\mathsf{A})$	Expected number of votes across election received by candidate A if hand counted
	$\mathcal{P}_r$	Predicted distribution of interpretations for ballot with identifier r
	$\mathsf{P}_r^{\LEFTcircle}(\mathsf{A})$	Number of votes predicted for candidate A on ballot with identifier r
	$\mu_{\mathsf{A,B}}$	Bayesian diluted margin of A with respect to candidate B
	μ_{A}	Minimum of Bayesian diluted margin across candidates
	cvr	Cast vote record containing identifiers and predicted interpretations for each ballot
	$\mathsf{cvr}(\mathsf{A})$	Predicted number of votes for A on cvr
	μ_{cvr}	Bayesian margin declared by cvr
	$\mathsf{D}_{\mathsf{cvr}}$	Discrepancy of cvr with respect to election E
Cons.	$T_{\mathbf{b}}$	Set of possible interpretations on $\mathbf{b}$
	$T_{\mathbf{b}}^{\CIRCLE}(\mathsf{A})$	Maximum interpretation for A on $\mathbf{b}$
	$T_{\mathbf{b}}^{\Circle}(\mathsf{A})$	Minumum interpretation for A on $\mathbf{b}$
	$T^{\CIRCLE}(\mathsf{A})$	Most possible votes for A in election
	$T^{\Circle}(\mathsf{A})$	Least possible votes for A in election
	$\mu^+_{\mathsf{A,B}}$	Conservative diluted margin of A with respect to candidate B
	μ^+_{A}	Minimum of Conservative diluted margin across candidates
	μ^+_{cvr}	Conservative diluted margin in cvr^+
	cvr^+	Conservative cvr declaring set of possible interpretations
	$\mathsf{P}_r^{\Circle}(\mathsf{A})$	Minimum predicted interpretation for A
	$\mathsf{P}_r^{\CIRCLE}(\mathsf{A})$	Maximum predicted interpretation for A
	$\mathsf{cvr}^{\Circle}(\mathsf{A})$	Minimum possible votes for A on cvr
	$\mathsf{cvr}^{\CIRCLE}(\mathsf{A})$	Maximum possible votes for A on cvr
	$\mathsf{D}^+_{\mathsf{cvr}}$	Conservative discrepancy of cvr with respect to election E

Table 1: Summary of Notation

3 Preliminaries; the Bayesian Setting

We use boldface to refer to "physical" objects, such as individual ballots (typically denoted $\mathbf{b}$) or groups of ballots (typically $\mathbf{B}$). For a natural number k, we define $[k] = \{1,\dots,k\}$ (and $[0] = \emptyset$). We let $\Sigma = [-2,2]$, a set of particular significance as standard single-ballot discrepancy takes values in this set; this is discussed in detail later in the paper. For a set X, let X^* be the set of all finite-length sequences over X; that is, $X^* = \{(x_1,\dots,x_k) \mid k \geq 0, x_i \in X\}$. In particular, we let $\{0,1\}^*$ denote the collection of finite-length bitstrings. Finally, we define $X^{\mathbb{N}}$ to be the set of all sequences $\{(x_0,x_1,\dots) \mid x_i \in X\}$. We typically use script letters, e.g., $\mathcal{P}$, to denote probability distributions. Making multiple interpretations explicit requires many new definitions. A summary is in Table 1.

Formal definitions for ballots, elections, and cast-vote records. For simplicity, we consider only audits of a single first-past-the-post race with a set C of candidates. See Stark [7, 10] for a detailed account of how more complex elections can be reduced to this canonical setting. We now set down the elementary definitions of elections, manifests, and CVRs. **The main definitional change is that a physical ballot induces a probability distribution of possible outcomes;**

this models the important situation where a physical ballot may have ambiguous markings that could reasonably be interpreted in various ways by an auditing board or a tabulation device. We explore two conventions for cast vote records to reflect this uncertainty: the first is the *Bayesian* framework, which associates with each ballot a predicted probability distribution of possible readings (depending on the anticipated reading a particular audit board, say). We then formulate a simpler, *conservative* model that posits a collection of interpretations for each ballot without assigning them probabilities. This reduces the complexity of producing cast-vote records, as one only needs to identify the *possible* interpretations of a ballot rather than a comprehensive probability distribution. In this conservative model, we consider quite strong adversarial behavior that assumes that the audit board always returns the "worst-case" interpretation allowed by the ballot. (In fact, this is equivalent to replacing the probabilities listed on the CVR with a probability of 1 for some interpretation that "minimizes" the margin for a declared winner.) To move as swiftly as possible to our results on conventional audits with marginal marks, we first lay out the definitions for the Bayesian case; we return to the definitions for the conservative case in Section 4.1.

Definition 1 (Interpretations, ballots, and ballot families)**.** *Let C be a set of candidates. An* interpretation *is a function $I : C \to \{0,1\}$; when C is understood from context, we write $\mathcal{I} = \{I : C \to \{0,1\}\}$ for the collection of all interpretations. Such interpretations to indicate the votes appearing for candidates on individual ballots. A* ballot *is a physical object* **b** *with two properties:*

(1) The ballot **b** *is* labeled *with an indelible* identifier $\mathrm{id}_{\mathbf{b}} \in \{0,1\}^*$.
(2) The ballot **b** *determines a "ground truth" probability distribution $\mathcal{T}_{\mathbf{b}}$ on the set of interpretations $\mathcal{I}$. Thus $\mathcal{T}_{\mathbf{b}} : \mathcal{I} \to \mathbb{R}$ is a non-negative function on $\mathcal{I}$ for which* $\sum_{I\in\mathcal{I}} \mathcal{T}_{\mathbf{b}}(I) = 1$.

We let $\mathsf{T}^{\LEFTcircle}_{\mathbf{b}} : C \to \mathbb{R}$ denote the expected interpretation over the distribution $\mathcal{T}_{\mathbf{b}}$: that is, for each $\mathsf{A} \in C$, $\mathsf{T}^{\LEFTcircle}_{\mathbf{b}}(\mathsf{A}) = \mathbb{E}[\mathrm{I}_{\mathbf{b}}(\mathsf{A})]$, *where $\mathrm{I}_{\mathbf{b}}$ is a random variable distributed according to $\mathcal{T}_{\mathbf{b}}$.*

A ballot family **B** *is a collection of ballots with the same candidates. We let $\mathbf{B}_\iota = \{\mathbf{b} \in \mathbf{B} \mid \mathrm{id}_{\mathbf{b}} = \iota\}$ denote the subset of ballots with identifier ι. When ballots identifiers are distinct across* **B***, so that $|\mathbf{B}_\iota| \leq 1$ for each ι, we say that the family is* uniquely labeled*.*

Ballots are typically intended to have an unambiguous interpretation. This corresponds to the case where the distribution $\mathcal{T}_{\mathbf{b}}$ is supported on a single interpretation (taking the value 1 at that interpretation and zero elsewhere).

Definition 2. *An* election *E is a tuple $(C, \mathbf{B}, \mathsf{S})$ where C is a set of candidates,* **B** *is a ballot family, and $\mathsf{S} = |\mathbf{B}|$ is the total number of ballots.*

Definition 3 (Election winners, losers, and margin)**.** *Let $E = (C, \mathbf{B}, \mathsf{S})$ be an election. For a candidate $\mathsf{A} \in C$, define*

$$E^{\LEFTcircle}(\mathsf{A}) = \sum_{\mathbf{b}\in\mathbf{B}} \mathsf{T}^{\LEFTcircle}_{\mathbf{b}}(A).$$

The margin of A with respect to B *is defined to be*

$$\mu_{\mathsf{A},\mathsf{B}} = \frac{E^{\LEFTcircle}(\mathsf{A}) - E^{\LEFTcircle}(\mathsf{B})}{\mathsf{S}}.$$

When we wish to emphasize that this notion of margin is defined with respect to the expected values $\mathsf{T}^{\LEFTcircle}_{\mathbf{b}}(\cdot)$ we refer to it as Bayesian margin*. A candidate $\mathsf{W} \in C$ is the* winner *of the election E if $\forall \mathsf{A} \in C \setminus \{\mathsf{W}\}, \mu_{\mathsf{W},\mathsf{A}} > 0$. In this case, we define the* margin*, denoted μ, of the election to be the minimum of these quantities:*

$$\mu_E = \min_{\mathsf{A}\in C\setminus\{\mathsf{W}\}} \mu_{\mathsf{W},\mathsf{A}}.$$

Otherwise, the election does not determine a winner and we define $\mu_E := 0$. If a winner exists it is unique. A candidate L is called a loser *if there is a candidate A for which $E^{\LEFTcircle}(\mathsf{L}) < E^{\LEFTcircle}(\mathsf{A})$.*

Cast-vote record (CVR). A cast-vote record table (CVR) is an (untrusted) declaration of both the ballots appearing in an election and the interpretations of the ballots.

Definition 4 (Bayesian Cast-Vote Record Table)**.** *Let C be a set of candidates. A* Bayesian Cast-Vote Record table (CVR) *is a sequence of pairs $\mathsf{cvr} = ((\iota_1, \mathcal{P}_1), \ldots, (\iota_t, \mathcal{P}_{\mathsf{S}_{\mathsf{cvr}}}))$ where the ι_r are distinct indetifier bitstrings in $\{0,1\}^*$ and each $\mathcal{P}_r : \mathcal{I} \to \mathbb{R}$ is a probability distribution on $\mathcal{I}$ which we refer to as a "prediction." Intuitively, such a cvr declares that, for each row r, the interpretation of the ballot labeled ι_r is given by the distribution $\mathcal{P}_r$.*

For each candidate $\mathsf{A} \in C$ and $r \in \{1, \ldots, \mathsf{S}_{\mathsf{cvr}}\}$, we define $\mathsf{P}^{\LEFTcircle}_r(\mathsf{A}) = \mathbb{E}[I(\mathsf{A})]$, the expected value of the vote for A when I is drawn according to the distribution $\mathcal{P}_r$. We then define the total cvr declaration for candidate A as

$$\mathsf{cvr}(\mathsf{A}) = \sum_{r=1}^{\mathsf{S}_{\mathsf{cvr}}} \mathsf{P}^{\LEFTcircle}_r(\mathsf{A}).$$

Declared winners and losers. *If there is a candidate $\mathsf{A} \in C$ with the property that $\mathsf{cvr}(\mathsf{A}) > \mathsf{cvr}(\mathsf{B})$ for every $\mathsf{B} \in C \setminus \{\mathsf{A}\}$, then we say that A is the* declared winner *according to cvr. A candidate A for which there exists some B such that $\mathsf{cvr}(\mathsf{A}) < \mathsf{cvr}(\mathsf{B})$ is called a* declared loser.

We use the following general language when referring to CVRs:

(1) The bitstrings ι_r are identifiers *and we let* `Idents`(cvr) *denote this set.*
(2) The number $\mathsf{S}_{\mathsf{cvr}}$ is the size *of the CVR.*
(3) The rth row*, denoted $\mathsf{cvr}(r)$, refers to the tuple $(\iota_r, \mathcal{P}_r)$.*

(4) *Identifiers appearing in the CVR are distinct, so when convenient we reference $\mathcal{P}_r$ and P_r by identifier rather than row: that is, for an identifier $\iota = \iota_r$ appearing in the CVR, we define $\mathcal{P}_\iota := \mathcal{P}_r$ and $\mathsf{P}^{\LEFTcircle}_\iota := \mathsf{P}^{\LEFTcircle}_r$.*

Definition 5 (Bayesian declared margin)**.** *Let* cvr *be a Bayesian CVR with declared winner* W *and size* $\mathsf{S}_{\mathsf{cvr}}$*. The* declared margin *of* cvr *is the quantity*

$$\mu_{\mathsf{cvr}} := \min_{\mathsf{A} \in C \setminus \mathsf{W}} \frac{\mathsf{cvr}(\mathsf{W}) - \mathsf{cvr}(\mathsf{A})}{\mathsf{S}_{\mathsf{cvr}}} \geq 0.$$

If cvr *has no declared winner, we define* $\mu_{\mathsf{cvr}} := 0$.

Definition 6 (Bayesian validity)**.** *Let $E = (C, \mathbf{B}, \mathsf{S})$ be an election and* cvr *be a Bayesian CVR. We say that* cvr *is* valid *for E if E has a winner and* cvr *declares that winner. We say that* cvr *is invalid for E if it declares a winner* W *that does not win E.*

Discrepancy is a standard measure of how much a CVR overcounts W's margin over other candidates $\mathsf{A} \neq \mathsf{W}$. Discrepancy naturally extends to the Bayesian case by considering the discrepancy of the expected interpretation of the auditor; this is defined precisely below. It is convenient to generalize the notion to yield a default result appropriate for cases when no ballot matches the identifier declared in a CVR; this is the $\perp$ case in the definition below.

Definition 7 (Bayesian discrepancy)**.** *Let $E = (C, \mathbf{B}, \mathsf{S})$ be an election and let $\mathsf{cvr} = ((\iota_1, \mathcal{P}_1), \ldots, (\iota_t, \mathcal{P}_{\mathsf{S}_{\mathsf{cvr}}}))$ be a Bayesian CVR with declared winner* W*. For a ballot $\mathbf{b} \in \mathbf{B}$ and a distribution $\mathcal{P} : \mathcal{I} \to \mathbb{R}$, the* discrepancy *is*

$$\mathsf{D}[\mathcal{P}, \mathbf{b}; \mathsf{W}] = \max_{\mathsf{A} \in C \setminus \mathsf{W}} \left((\mathsf{P}^{\LEFTcircle}(\mathsf{W}) - \mathsf{P}^{\LEFTcircle}(\mathsf{A})) - (\mathsf{T}^{\LEFTcircle}_{\mathbf{b}}(\mathsf{W}) - \mathsf{T}^{\LEFTcircle}_{\mathbf{b}}(\mathsf{A})) \right)$$

where, as above, $\mathsf{P}^{\LEFTcircle}(\mathsf{A}) = \mathbb{E}[I(\mathsf{A})]$ with I distributed according to $\mathcal{P}$. We expand this definition to apply, additionally, to a special symbol $\perp$:

$$\mathsf{D}[\mathcal{P}, \perp; \mathsf{W}] = \max_{\mathsf{A} \in C \setminus \mathsf{W}} \left(\mathsf{P}^{\LEFTcircle}(\mathsf{W}) - \mathsf{P}^{\LEFTcircle}(\mathsf{A}) \right) + 1\,.$$

For an identifier ι define

$$\mathsf{D}_{\mathsf{cvr},\iota} = \begin{cases} \min\{\mathsf{D}[\mathcal{P}_\iota, \mathbf{b}; \mathsf{W}] \mid \mathbf{b} \in \mathbf{B}_\iota\} & \text{if } \mathbf{B}_\iota \neq \emptyset, \\ \mathsf{D}[\mathcal{P}_\iota, \perp; \mathsf{W}] & \text{otherwise.} \end{cases}$$

The discrepancy of the cast vote record table cvr *is:* $\mathsf{D}_{\mathsf{cvr}} = \sum_{i=1}^{\mathsf{S}_{\mathsf{cvr}}} \mathsf{D}_{\mathsf{cvr},\iota_i}$.

We begin by establishing the fundamental relationship between discrepancy and margin for invalid CVRs.

Lemma 1. *Let $E = (C, \mathbf{B}, \mathsf{S})$ be an election and let* cvr *be an invalid CVR for E with declared winner* W *and size $\mathsf{S}_{\mathsf{cvr}} = \mathsf{S}$. Then $\mathsf{D}_{\mathsf{cvr}} \geq (\mu_{\mathsf{cvr}} + \mu_E) \cdot \mathsf{S} \geq \mu_{\mathsf{cvr}} \cdot \mathsf{S}$.*

As a reminder, all proofs are deferred to the online version [38].

3.1 Adaptive Statistical Tests

The standard approach for ballot comparison audits determines an experiment that appropriately samples ballots in order to obtain discrepancy samples; the procedure concludes with Inconclusive unless one can statistically reject the hypothesis that discrepancy is at least μ_{cvr}, as this is true for all invalid elections (Lemma 1). Such hypothesis testing requires a one-sided statistical test. In fact, as we consider environments that need not respond independently during each ballot request, we require a slightly richer family of tests, formulated in this setting by [5]. We briefly lay out the definitions here. The parameter δ will ultimately be set to μ in our application.

Definition 8. *[5, Definition 8] A sequence of bounded (real-valued) random variables $X_1, \ldots$ are said to be δ-*dominating *if, for each $t \geq 0$, $\mathbb{E}[X_t \mid X_1, \ldots, X_{t-1}] \geq \delta$. We apply this terminology to the distribution $\mathcal{D}$ corresponding to the random variables, writing $\delta \trianglelefteq \mathcal{D}$.*

Definition 9. *[5, Definition 9] Let $\Sigma = [-2, 2]$. A* stopping time *is a function $\mathsf{Stop} : \Sigma^* \to \{0, 1\}$ so that for any sequence $x_1, x_2, \ldots$ of values in Σ there is a finite prefix $x_1, \ldots, x_k$ for which $\mathsf{Stop}(x_1, \ldots, x_k) = 1$.*

For a sequence of random variables $X_1, \ldots$ taking values in Σ, let $\tau_{\mathsf{Stop}}(X_1, \ldots)$ be the random variable given by the smallest t for which $\mathsf{Stop}(X_1, \ldots, X_t) = 1$. This naturally determines the random variable $X_1, \ldots, X_{\tau_{\mathsf{Stop}}}$, the prefix of the X_i given by the first time $\mathsf{Stop}() = 1$.

Definition 10. *[5, Definition 10] An* adaptive audit test*, denoted $(\mathsf{Stop}, \mathsf{R})$, is described by two families of functions, Stop_δ and R_δ. For each $0 < \delta \leq 1$,*

(1) Stop_δ is a stopping time, as in Definition 9, and
(2) $\mathsf{R}_\delta : \Sigma^* \to \{0, 1\}$ is the rejection criterion.

Let $\mathcal{D}$ be a probability distribution on $\Sigma^{\mathbb{N}}$; for such a distribution, define $\alpha_{\delta,\mathcal{D}} = \mathbb{E}[\mathsf{R}_\delta(X_1, \ldots, X_\tau)]$ where $X_1, \ldots$ are random variables distributed according to $\mathcal{D}$ and τ is determined by Stop_δ. The risk *of the test $(\mathsf{Stop}, \mathsf{R})$ is*

$$\alpha = \sup_{\substack{0 < \delta \leq 2 \\ \delta \trianglelefteq \mathcal{D}}} \alpha_{\delta,\mathcal{D}}\,. \tag{3}$$

4 Bayesian Comparison Audits; Leveraging Marginal Marks

We present our auditor and analysis using a cryptographic game between the *auditor* and the *environment*, in the spirit of [5]. The auditor naturally reflects the role played by an auditing team in a conventional audit; as such it is provided with a CVR, the actual size of the election (in real-world settings this is provided by a trusted *ballot manifest*), and the ability to request ballots by identifier in order to carry out the audit. When the auditor has completed the audit, it returns either the

Auditor (`MAudit`)–Environment (`Env`) game for election $E = (\mathbf{B}, \mathcal{C}, \mathsf{S})$ and CVR cvr

(1) **Setup**.

  (a) **Ballot and tabulation delivery (to `Env`).** The physical ballots $\mathbf{B}$ is given to `Env`.

  (b) **Ballot manifest and CVR delivery (to `MAudit`).** The size S and the cvr are given to the auditor `MAudit`.

(2) **Audit**. `MAudit` repeatedly makes a ballot request to `Env`, or chooses to conclude the audit:
A ballot request. `MAudit` requests a ballot from `Env` with identifier $\iota \in \{0,1\}^*$.
`Env` does one of two things:

  (a) Responds with a ballot $\mathbf{b}$ with identifier ι'; in this case, `MAudit` is given ι' and I, an interpretation drawn from $\mathcal{T}_{\mathbf{b}}$.

  (b) Responds with **No ballot**; this is forwarded to `MAudit`.

(3) **Conclusion**. `MAudit` returns one of the two values: Consistent or Inconclusive.

Figure 2: The $\mathsf{RLA}_{\texttt{MAudit},\texttt{Env}}(E, \mathsf{cvr})$ auditing game for comparison audits with multiple interpretations.

token Inconclusive or Consistent. The environment, on the other hand, is responsible for providing physical ballots when they are requested. Note, in particular, that the auditor's access to ballots is entirely mediated by the environment, which may choose not to return ballots when they are requested, make choices about which ballot to return when multiple ballots share an identifier, etc. The power of this modeling framework is that the critical concept of risk can be defined *in the worst case over all possible behaviors of the environment*. This reflects a wide variety of failures—inadvertent or malicious—that occur in practical audits.

The formal Auditor-Environment game is described in Figure 2. We remark that each ballot "delivered" by the environment in the game generates a single draw from the associated ground truth distribution of interpretations; multiple requests for the same ballot are permitted by the convention, which generate new independent samples from the ground truth distribution (modeling a situation where an audit board evaluates the ballot afresh each time it is sampled).

The formal auditor. The auditor we consider is a standard "comparison auditor" [1] adapted to the setting where discrepancy is computed with respect to the expected value of the predicted distribution in the CVR rather than a single interpretation. Fortunately, once discrepancy is redefined in this way, the remaining analysis of the standard auditor requires no significant changes.

The formal auditor for the Bayesian setting is presented in Figure 3. We now define the risk of a Bayesian audit: this is an upper bound on the probability (taken over all elections and environments) that the audit outputs Consistent if the CVR is in fact invalid.

`MAudit`$[\mathsf{Stop}, \mathsf{R}](E = (\mathcal{C}, \mathbf{B}, \mathsf{S}), \mathsf{cvr})$

(1) Receive $\mathsf{cvr} = ((\iota_1, \mathcal{P}_1), \ldots, (\iota_t, \mathcal{P}_{\mathsf{S}_{\mathsf{cvr}}}))$.
(2) If cvr has repeated identifiers or $\mathsf{S} \neq \mathsf{S}_{\mathsf{cvr}}$, return Inconclusive.
(3) If cvr declares a winner W, define $\mu := \mu_{\mathsf{cvr}}$; otherwise, return Inconclusive.
(4) Initialize $\mathsf{iter} = 0$.
(5) Repeat
  (a) Increment $\mathsf{iter} := \mathsf{iter} + 1$.
  (b) Perform $\mathsf{D}_{\mathsf{iter}}$:= `BasicExperiment`
  until $\mathsf{Stop}_\mu(\mathsf{D}_1, ..., \mathsf{D}_{\mathsf{iter}}) = 1$
(6) If $\mathsf{R}_\mu(\mathsf{D}_1, ..., \mathsf{D}_{\mathsf{iter}}) = 1$ return `Consistent`; otherwise return Inconclusive.

`BasicExperiment`:

(1) `RowSelect`: Select a row $r \in [\mathsf{S}]$ uniformly.
(2) Let ι be the identifier in row r; request delivery of ι.
(3) If a ballot $\mathbf{b}$ was delivered with identifier $\iota' = \iota$ and interpretation I^*, return

$$\text{act} := \max_{\mathsf{A} \in \mathcal{C} \setminus \mathsf{W}} \left(\mathsf{P}_r^{◐}(\mathsf{W}) - \mathsf{P}_r^{◐}(\mathsf{A}) - (I^*(\mathsf{W}) - I^*(\mathsf{A})) \right).$$

Else, return $\text{act} := \max_{\mathsf{A} \in \mathcal{C} \setminus \mathsf{W}} \left(\mathsf{P}_r^{◐}(\mathsf{W}) - \mathsf{P}_r^{◐}(\mathsf{A}) \right) + 1.$

Figure 3: The auditor `MAudit`$[(\mathsf{Stop}, \mathsf{R})]$.

Definition 11 (Risk of Bayesian audit)**.** *Let* `MAudit` *be an Auditor. For election E and environment* `Env` *let* $\mathsf{RLA}_{\texttt{MAudit},\texttt{Env}}(E)$ *denote the random variable equal to the conclusion of the audit as described in Figure 2. An auditor* `MAudit` *has* α-risk *(or* α-soundness*) if, for all elections E, all invalid* cvr*s and all environments* `Env`*,*

$$\Pr_{\texttt{MAudit},\texttt{Env}} [\mathsf{RLA}_{\texttt{MAudit},\texttt{Env}}(E, \mathsf{cvr}) = \text{Consistent}] \leq \alpha\,.$$

Theorem 1. *Let E be an election and* cvr *be an invalid CVR. For any environment* `Env`*, let* $\mathsf{D}^{\text{obs}}_{i,\texttt{Env}}$ *denote the random variable computed in* `BasicExperiment` *during iteration i; then* $\mathbb{E}[\mathsf{D}^{\text{obs}}_{i,\texttt{Env}}] \geq \mathsf{D}_{\mathsf{cvr}}/\mathsf{S} \geq \mu_{\mathsf{cvr}}$.

Let $(\mathsf{Stop}, \mathsf{R})$ *be an adaptive audit test with risk α. Let* `MAudit` *be as in Figure 3,* `MAudit`$[\mathsf{Stop}, \mathsf{R}]$ *has risk α.*

4.1 Conservative Comparison Audits

Our second approach is called a *conservative comparison audit*. At a high level, this can be seen as replacing the probability distribution from the previous section with a set of possible interpretations. For a declared winner W, the auditor runs the audit as though the interpretation that minimizes μ_{W} occurs with probability 1. When accurate prediction of the auditor distribution is possible this model has worse results than Bayesian interpretation but doesn't suffer from the poor efficiency when one overshoots the probability of the auditor (see results in the next subsection). As a second advantage, for our modeling in the conservative setting, the environment is allowed to choose the interpretation of the ballot selected by the audit board among the possible interpretations of the ballot. This is opposed to the previous section where this interpretation was i.i.d. sampled by the game.

We continue to use the common definitions and terminology for ballots and interpretations. The major change is rather than calling for a ground truth distribution $\mathcal{T}_{\mathbf{b}}$, we merely define a set $T_{\mathbf{b}}$ of *possible interpretations* for this ballot. (That is, interpretations that could reasonably be the conclusion of an audit board's inspection of the ballot.) In the setting with a ground-truth distribution, the set $T_{\mathbf{b}}$ would naturally correspond to the *support* of the distribution $T_{\mathbf{b}} = \{I \in \mathcal{I} \mid \mathcal{T}_{\mathbf{b}}(I) > 0\}$.

Definition 12 (Conservative Interpretation limits)**.** *Let $E = (C, \mathbf{B}, \mathsf{S})$ be an election. For a ballot $\mathbf{b} \in \mathbf{B}$ and a candidate $\mathsf{A} \in C$, define*

$$\mathsf{T}^{\circ}_{\mathbf{b}}(\mathsf{A}) = \min\{I(\mathsf{A}) \mid I \in T_{\mathbf{b}}\},$$
$$\mathsf{T}^{\bullet}_{\mathbf{b}}(\mathsf{A}) = \max\{I(\mathsf{A}) \mid I \in T_{\mathbf{b}}\}.$$

Thus $\mathsf{T}^{\bullet}_{\mathbf{b}}(\mathsf{A})$ and $\mathsf{T}^{\circ}_{\mathbf{b}}(\mathsf{A})$ indicate the most and least favorable interpretations of the ballot for the candidate A. We additionally define

$$E^{\circ}(\mathsf{A}) = \sum_{\mathbf{b} \in \mathbf{B}} \mathsf{T}^{\circ}_{\mathbf{b}}(A), \qquad \text{and} \qquad E^{\bullet}(\mathsf{A}) = \sum_{\mathbf{b} \in \mathbf{B}} \mathsf{T}^{\bullet}_{\mathbf{b}}(A).$$

Definition 13 (Conservative margin and winner)**.** *The* conservative margin of A with respect to B *compares the worst and best case interpretations for two candidates:*

$$\mu^{+}_{\mathsf{A},\mathsf{B}} = \frac{E^{\circ}(\mathsf{A}) - E^{\bullet}(\mathsf{B})}{\mathsf{S}}.$$

We then define $\mu^{+}_{\mathsf{A}} = \min_{\mathsf{B} \in C \setminus \{\mathsf{A}\}} \mu^{+}_{\mathsf{A},\mathsf{B}}$. If there exists some candidate W such that $\mu^{+}_{\mathsf{W}} > 0$ we call this candidate the (conservative) winner and denote this quantity by μ^{+}. If a conservative winner exists, then it is unique. If there is no conservative winner, the election is called indeterminate. *We call a candidate L a conservative loser if there exists some other candidate A such that $\mu^{+}_{\mathsf{L},\mathsf{A}} < 0$.*

Observe that losing candidates may exist even in indeterminate elections and that, in general, a losing candidate candidate cannot tie the election even if every ballot is given the interpretation most favorable to the candidate. As mentioned above, conservative CVRs specify a set of declared interpretations for each ballot.

Definition 14 (Conservative Cast-Vote Record Table (CVR))**.** *Let C be a set of candidates. A* conservative Cast-Vote Record table (CVR) *is a sequence of pairs*

$$\mathsf{cvr}^{+} = ((\iota_1, P_1), \ldots, (\iota_t, P_{\mathsf{S}_{\mathsf{cvr}}})),$$

where the ι_r are distinct bitstrings in $\{0,1\}^*$ and each $P_r \subset \mathcal{I}$ is a subset of interpretations. The CVR declares that for each row r the interpretation of the ballot labeled ι_r lies in the set P_r. For each candidate $\mathsf{A} \in C$, we define

$$\mathsf{P}^{\circ}_{r}(\mathsf{A}) = \min\{I(A) \mid I \in P_r\},$$
$$\mathsf{P}^{\bullet}_{r}(\mathsf{A}) = \max\{I(A) \mid I \in P_r\},$$

$$\mathsf{cvr}^{\circ}(\mathsf{A}) = \sum_{r=1}^{\mathsf{S}_{\mathsf{cvr}}} \mathsf{P}^{\circ}_{r}(\mathsf{A}), \qquad \text{and} \qquad \mathsf{cvr}^{\bullet}(\mathsf{A}) = \sum_{r=1}^{\mathsf{S}_{\mathsf{cvr}}} \mathsf{P}^{\bullet}_{r}(\mathsf{A}).$$

Definition 15 (Declared winners, losers, and contradictory CVRs.)**.** *If there is a candidate $\mathsf{A} \in C$ with the property that $\mathsf{cvr}^{\circ}(\mathsf{A}) > \mathsf{cvr}^{\bullet}(\mathsf{B})$ for every $\mathsf{B} \in C \setminus \{\mathsf{A}\}$, then we say that A is the* declared conservative winner *according to cvr^{+}. If there is no such candidate, we say that cvr^{+} is* indeterminate. *Any candidate $\mathsf{L} \in C$ for which $\mathsf{cvr}^{\bullet}(\mathsf{L}) < \mathsf{cvr}^{\circ}(\mathsf{A})$ for some candidate $\mathsf{A} \in C$ is a* declared conservative loser. *We adopt the same naming conventions for such conservative CVRs as we do with Bayesian CVRs (that is, for the set of identifiers, size, and treating identifiers as indices for P_r and P_r).*

Finally, we say that two CVRs are contradictory *if there is a candidate declared as a winner in one and a loser in the other.*

An election or CVR can be indeterminate without a tie: e.g., there may be two candidates A and B for which $\mathsf{cvr}^{\bullet}(\mathsf{A}) > \mathsf{cvr}^{\circ}(\mathsf{B})$ and $\mathsf{cvr}^{\bullet}(\mathsf{B}) > \mathsf{cvr}^{\circ}(\mathsf{A})$.

Definition 16 (Conservative Declared Margin)**.** *We call a candidate L a conservative loser if there exists some other candidate A such that $\mu^{+}_{\mathsf{A},\mathsf{L}} > 0$. The* declared margin *of a conservative* cvr *with* declared winner W *is the quantity*

$$\mu^{+}_{\mathsf{cvr}} := \min_{\mathsf{A} \in C \setminus \mathsf{W}} \frac{\mathsf{cvr}^{\circ}(\mathsf{W}) - \mathsf{cvr}^{\bullet}(\mathsf{A})}{\mathsf{S}}.$$

Definition 17 (Conservative discrepancy)**.** *Let $E = (C, \mathbf{B}, \mathsf{S})$ be an election and let $\mathsf{cvr} = ((\iota_1, P_1), \ldots, (\iota_t, P_t))$ be a conservative CVR with declared winner W. For a ballot $\mathbf{b} \in \mathbf{B}$ and interpretation I, the* conservative discrepancy *of I with respect to $\mathbf{b}$ is*

$$\mathsf{D}^{+}[I, \mathbf{b}; \mathsf{W}] = \max_{\mathsf{A} \in C \setminus \mathsf{W}} \Big((I(\mathsf{W}) - I(\mathsf{A})) - (T^{\bullet}_{\mathbf{b}}(\mathsf{W}) - T^{\circ}_{\mathbf{b}}(\mathsf{A})) \Big).$$

We expand this definition to apply, additionally, to a special symbol $\perp$:

$$\mathsf{D}^{+}[I, \perp; \mathsf{W}] = \max_{\mathsf{A} \in C \setminus \mathsf{W}} (I(\mathsf{W}) - I(\mathsf{A})) + 1.$$

The discrepancy of a cast vote record table cvr *is:*

$$\mathsf{D}^{+}_{\mathsf{cvr}} = \sum_{i=1}^{t} \min\{\mathsf{D}[I, \mathbf{b}; \mathsf{W}] \;\big|\; I \in \mathtt{I}^{\mathsf{cvr}}_{i}, \mathbf{b} = \perp \text{ or } \mathrm{id}_{\mathbf{b}} = \iota_i\}.$$

Auditor (MAudit)–Environment (Env) game for election $E = (\mathbf{B}, C, \mathsf{S})$ and CVR cvr

(1) **Setup**.
   (a) **Ballot and tabulation delivery (to Env).** The physical ballots $\mathbf{B}$ is given to Env.
   (b) **Ballot manifest and CVR delivery (to MAudit).** The size S and the cvr are given to the auditor MAudit.

(2) **Audit**. MAudit repeatedly makes a ballot request to Env, or chooses to conclude the audit:
   A ballot request. MAudit requests a ballot from Env with identifier $\iota \in \{0,1\}^*$.
   Env either responds a ballot $\mathbf{b}$ with single identifier ι' and **an interpretation** $I \in T_{\mathbf{b}}$ or responds with **No ballot.**

(3) **Conclusion**. MAudit returns one of the two values: CONSISTENT or INCONCLUSIVE.

Figure 4: The $\mathsf{RLA}_{\mathtt{MAudit},\mathtt{Env}}(E, \mathsf{cvr})$ auditing game for comparison audits with multiple interpretations.

Definition 18 (Conservative consistency and validity)**.** *A* $\mathsf{cvr}^+ = ((\iota_1, P_1), \ldots, (\iota_t, P_t))$ *is* consistent *with E if the ballots* $\mathbf{B}$ *are uniquely labeled, the identifiers appearing in* cvr *are identical to those appearing on the ballots of E, and for each ballot* $\mathbf{b} \in \mathbf{B}$*,* $T_{\mathbf{b}} \subset P_{\iota_{\mathbf{b}}}$*. When* $T_{\mathbf{b}} = P_{\mathrm{id}_{\mathbf{b}}}$ *for each ballot, we say that* cvr *is the* canonical *CVR for the election E. We only define these notions for uniquely labeled elections.*

For election E with winner W*, a* cvr *is invalid if it declares a winner other than* W*.*

4.1.1 Conservative Game and Auditor

We consider the following adaption of the auditing game where the environment is allowed to specify the interpretation returned to the auditor in Figure 4 with the change in boldface. The auditor is identical to that in Figure 3 except the Bayesian cvr margin in Step (3) is replaced with μ^+_{cvr}.

Theorem 2. *Let E be an election and* cvr *be an invalid CVR. For any environment* Env *engaging in Figure 4, let* $\mathsf{D}^{\mathrm{obs}}_{i,\mathtt{Env}}$ *denote the random variable computed in* BasicExperiment *then* $\mathbb{E}[\mathsf{D}^{\mathrm{obs}}_{i,\mathtt{Env}}] \geq \mathsf{D}_{\mathsf{cvr}}/\mathsf{S} \geq \mu$*. where* μ *is the value computed in Step (3).*

Let $(\mathsf{Stop}, \mathsf{R})$ *be an adaptive audit test with risk* α*. Let* MAudit *be as in Figure 3 with* μ_{cvr} *replaced with* μ^+_{cvr}*,* $\mathtt{MAudit}[\mathsf{Stop}, \mathsf{R}]$ *has risk* α*.*

5 Completeness of Bayesian and Conservative Audits; Efficiency Analysis

The previous subsection established that one can recover a natural RLA from CVRs that declare multiple interpretations for ballots. In the case when all ballots and CVRs have a single interpretation the audit is a traditional ballot comparison audit. We also note that the auditor receives a single interpretation of the ballot. This models the fact that an audit board has to produce a final adjudication of a ballot. As described above, the purpose of the test is to decide whether $\mathbb{E}[\mathsf{D}_{\mathsf{cvr}}/\mathsf{S}] \geq \mu_{\mathsf{cvr}}$.

Perhaps the most widely deployed test meeting the demands outlined in the previous section is the Kaplan-Markov test (though there are other natural choices [10, 11, 26, 41]); in particular, this is the test featured in the Arlo RLA administration system. The basic Kaplan-Markov test (omitting minimum and maximum sample sizes) is as follows: for a parameter $\gamma > 1$, define the value

$$\mathsf{Risk}^{(\gamma)}_{\delta}(\mathsf{D}_1, \ldots, \mathsf{D}_\ell) = \prod_{\mathrm{iter}=1}^{\ell} \left(\frac{1 - \frac{\delta}{2\gamma}}{1 - \frac{\mathsf{D}_{\mathrm{iter}}}{2\gamma}} \right).$$

Intuitively, the "inflation" parameter γ determines the relationship between the rate of ballot comparison inconsistencies and sample sizes. Common values in the literature are approximately 1.1, see RLACalc. Moreover, the stopping time for risk limit α is determined by the test $\mathsf{Risk}_\delta \overset{?}{\leq} \alpha$; when this occurs, the test rejects the hypothesis (that the mean is larger than μ) and, in our setting, the auditor would output CONSISTENT.

Simulation parameters. We consider a two candidate election with 100,000 ballots with $m = .5\%$ marginal ballots and underlying error rates of $o_1 = .1\%$, $u_1 = .1\%$, $o_2 = .01\%$, and $u_2 = .01\%$ [1]; Errors o_1 and o_2 represent ballots with discrepancies of +1 and +2 respectively. If the CVR lists a vote for the winner W and the actual ballot shows a vote for a loser this is an o_2 error. The actual ballot showing no vote results in an o_1, this is called an undervote. The errors u_1 and u_2 are defined similarly for discrepancies of $-1, -2$. For marginal ballots, we assume there are two possible interpretations: an interpretation for W and an interpretation of an undervote. Throughout the simulations, we assign a discrepancy of 1 and 0 with equal probability in these cases. For the math, the important thing is the distribution of discrepancy, not the votes appearing on the ballots. We compare three approaches in Table 2:

(1) In the *Baseline* approach, half of the marginal ballots are counted as votes for the reported winner on the CVR. When a marginal ballot is selected for audit it is determined as a vote for W with probability 50%. Note this can result in discrepancy values of $\{-1, 0, 1\}$.
(2) In the *Bayesian* approach, the margin receives $\mathsf{P}^{\text{◐}}_r(\mathsf{W}) = .5$ "votes" for each marginal ballot. As above, when a marginal ballot is selected for audit it is determined as a vote for W with probability 50%. However, since $\mathsf{P}^{\text{◐}}_r(\mathsf{A}) = .5$ the possible discrepancy values are $\{-.5, .5\}$
(3) In the *Conservative* approach, no marginal ballots are included in the margin and only negative discrepancy values are possible if the auditor interprets the ballot as a vote for W.

All results perform 5000 simulations for each parameter setting. A simulation pulls a uniform ballot until the risk limit is met. Each time a marginal ballot is pulled its interpretation is independently sampled (50% probability of vote for W and 50% of blank). RLAs require the most work for small margins which is where our improvement is evident. We report on the mean, standard deviation, median, and 95% threshold of ballots retrieved in Table 2. In practice, the 95% mark is most relevant; this reflects the fact that the work to conduct an additional round of an RLA is often prohibitive—requiring a distributed process across a whole region—so termination within the specified sample size with high probability is desirable. The Bayesian approach reduces the 95% percentile of ballots sampled by roughly 10.5%. Roughly, this improvement arises because the variance of discrepancy for "marginal" ballots is reduced from $1/2$ using the baseline approach to $1/4$ using the Bayesian approach.

In the above, both the ground truth (observed by the auditor) and CVR prediction probability of a marginal mark being interpreted as a mark is $p_m = .5$. This is the setting that has the largest improvement on the variance of discrepancy. However, for varying values of p_m the approach still provides efficiency improvements over the baseline approach. This is shown in Table 3. Note for $p_m \in \{0,1\}$ the baseline and Bayesian approach are identical as marginal ballots never result in nonzero discrepancy. In Table 4, we consider two separate parameters for the probability that the CVR marks each marginal ballot as a vote for W, denoted p_{cvr} and the probability that the auditor does the same denoted $p_{\mathtt{MAudit}}$. Interestingly, the baseline approach outperforms the Bayesian approach when $p_{\mathsf{cvr}} > p_{\mathtt{MAudit}}$ by $\approx 20\%$; however Bayesian can outperform Baseline by $\approx 30\%$ when $p_{\mathsf{cvr}} < p_{\mathtt{MAudit}}$. This is part of the reason we introduce the conservative approach where one does not have to estimate probabilities.

6 Election Contestation and Competitive Ballot Comparison Audits

We consider a new class of post-election audits we call *competitive audits*. These audits can efficiently distinguish CVRs that are consistent with the ballots from invalid CVRs and, more generally, reconcile competing claims of victory made by two CVRs. This novel auditing technique also provides, as described earlier, a direct approach to handling election contestation.

To simplify our treatment of these audits, we focus on the conservative setting, in which CVRs declare, for each ballot, a subset of possible interpretations without assigning a probability distribution to the interpretations in the subset.

An intuitive survey of the framework. After an election, we assume that each candidate $\mathsf{A} \in C$ is represented by an *advocate*. In a preliminary "rescan phase," the advocate for the candidate A may inspect the ballots $\mathbf{B}$ and submit to the audit a conservative CVR, denoted cvr_A; the idea is that the advocate will submit a CVR that is favorable for the candidate. If the declared conservative winner of cvr_A is A, we say that this is a "declaration of victory." (Advocates may also be allowed to announce the presence of duplicate ballot labels in the election, which can then be corrected by the audit. For simplicity, however, we simply treat $\mathbf{B}$ as uniquely labeled in this setting.) Following this, the audit commences with the "judgment phase:" after considering some of the ballots, the audit either declares the audit to be Inconclusive or, for some candidate $\mathsf{A} \in C$, declares that "A is the winner."

The judgment phase; reconciling disagreements among the CVRs. If no pair of the submitted CVRs are contradictory, the audit concludes without considering any ballots. In this non-contradictory case, if any CVR declares victory the associated candidate is declared to be the winner of the audit; if, alternatively, no CVR declares a winner, the audit concludes with Inconclusive.

Otherwise, if two submitted CVRs are contradictory, the audit must settle the contradictory claims made by various submitted CVRs. In this case, some candidate A must be a declared winner according to one CVR and a declared loser according to another. We will see that there must then exist an identifier ι on which the two CVRs make contradictory assertions. Intuitively, one of these contradictory CVRs can then be removed from consideration by fetching the ballot with this identifier, if it exists, and relying on this ballot to settle the contradictory claims made by the two CVRs.

The natural procedure is complicated by three issues:

(1) two contradictory CVRs may declare different sets of identifiers rather than disagreeing on the interpretation of specific ballots;
(2) the audit should provide suitable protection against an imperfect or malicious environment that may not always respond with a given ballot, even when it exists; and
(3) the audit should withstand a small collection of errors in the creation of CVRs by advocates to reflect the practical difficulty of producing a perfect CVR, even by a well-intentioned advocate.

We remark that the second issue above forces the audit to treat "positive evidence" and "negative evidence" differently: for example, while the delivery of a ballot that disagrees with a particular CVR, either by possessing an identifier that is not declared in the CVR or having an interpretation that is inconsistent with those declared in the CVR, is considered evidence against the CVR, the failure of the environment to produce a ballot declared in a particular CVR is not. This cautious convention prevents a malicious environment from misleading the auditor. An environment that can refuse to retrieve ballots can always cause the audit to output Inconclusive.

Ultimately, we will prove two facts about the audit.

- If one of the submitted CVRs is consistent with the ballots and ballots are faithfully returned to the auditor dur-

	Baseline				Conservative				Bayesian			
μ	Mean	Stdev	Median	95%	Mean	Stdev	Median	95%	Mean	Stdev	Median	95%
.01	608	210	567	1028	595	181	576	938	583	175	545	920
.02	316	78	292	469	314	69	294	420	308	64	292	415
.03	213	44	202	283	212	38	219	263	210	39	202	271

Table 2: Number of ballots for Kaplan-Markov comparison audit between the baseline and the Bayesian auditor, across margins, and probability .5 of CVR and auditor determining marginal mark as for winner, $\alpha = .05, \gamma = 1.1$.

	Baseline				Conservative				Bayesian			
p_m	Mean	Stdev	Median	95%	Mean	Stdev	Median	95%	Mean	Stdev	Median	95%
1	471	115	438	704	514	164	494	824	468	116	438	704
.9	493	131	454	729	526	161	494	842	490	126	468	738
.8	520	149	470	817	533	158	494	824	510	135	485	773
.7	547	170	502	881	599	167	538	879	534	150	504	830
.6	575	188	539	930	579	178	544	944	559	155	533	856
.5	608	210	567	1028	595	181	576	938	583	175	545	920
.4	630	217	591	1071	617	181	576	975	616	189	576	1003
.3	662	231	616	1123	646	199	590	1020	641	199	596	1031
.2	685	227	644	1128	671	202	627	1071	671	211	619	1097
.1	707	227	627	1138	700	215	658	1108	692	209	630	1104
0	729	214	658	1184	729	214	658	1184	724	215	658	1184

Table 3: Number of ballots across p_m, $\alpha = .05, \gamma = 1.1, \mu = .01$.

ing the judgment phase, then—except with small probability that can be explicitly bounded as a function of the number of samples—the auditor will correctly conclude the audit with "W is the winner." Thus, if the advocate for W indeed acts in W's best interests, W can be confident that the audit will conclude favorably.

- If one of the submitted CVRs is consistent with the ballots and an alternate CVR declares that a losing candidate L won the election, the audit will not conclude that "L is the winner." Furthermore, this guarantee is robust against ballot suppression—it does not require that ballots are correctly returned when requested.

Taken together, we see that if a candidate's advocate submits a consistent CVR, the candidate can be assured a victory if no ballots are lost or suppressed, and can be assured that no losing candidate is declared the victor even in the face of ballot suppression.

6.1 Modeling Competitive Audits

Formally, we model a competitive audit with two parties, the *auditor* (denoted `Judge`) and the *environment* (denoted `Env`). The auditor is responsible for analyzing submitted conservative CVRs, requesting ballots to carry out the audit, and arriving at a final conclusion. The environment is responsible for servicing requests by the auditor for individual ballots: in particular, the environment is in possession of the ballots and, for each request for a ballot by the auditor, decides which ballot (if any) will be returned to the auditor and, moreover, how that ballot is to be interpreted (that is, which possible interpretation is to be adopted).

As mentioned above, this two party modeling of the audit reflects the fact that even a well-designed auditor may have to contend with such complexities as loss of ballots, potential malicious suppression of ballots, and the final choice of an "auditing board" in the interpretation of a ballot. These decisions are in the hands of the environment.

Definition 19 (Competitive auditor; competitive environments; honest environments)**.** *A competitive auditor* `Judge` *is a randomized procedure for carrying out a competitive audit. In the context of an election $E = (C, \mathbf{B}, \mathsf{S})$ with conservative ballots $\mathbf{B}$, the auditor is invoked with:*

(1) the set of candidates C,
(2) the size S of the election, and
(3) a list of conservative CVRs $\mathsf{cvr}_1^+, \ldots, \mathsf{cvr}_k^+$.

The audit proceeds in rounds, each round being an opportunity for the auditor to request delivery of a ballot (specified by an identifier) along with an interpretation of the ballot. After a sequence of requests, the auditor concludes the audit, resulting in either the statement Inconclusive*, or "A is the winner" for some candidate $\mathsf{A} \in C$.*

Ballot requests are handled by a second randomized procedure called the environment*. The environment is invoked with $(C, \mathbf{B}, \mathsf{cvr}_1^+, \ldots, \mathsf{cvr}_k^+)$. To each (identifier) request ι made by the auditor, the environment either responds with a ballot $\mathbf{b} \in \mathbf{B}$ and a single interpretation $I \in T_\mathbf{b}$ or a "no ballot" symbol $\perp$. While the environment may choose not to return a ballot when a matching ballot exists in $\mathbf{B}$, any ballot it does return is assumed to match the requested identifier: $\iota_\mathbf{b} = \iota$.*[3]

[3]These conventions are a convenience: the auditor can check the identifier of a returned ballot to ensure that it matches the request and, additionally, ensure that the returned interpretation is among those associated with the ballot; if either of these verifications fail, the auditor treats this ballot response as a $\perp$.

p_{cvr}	$p_{MAudit} = p_{cvr} + .4$			$p_{MAudit} = p_{cvr} + .2$			$p_{MAudit} = p_{cvr}$			$p_{MAudit} = p_{cvr} - .2$			$p_{MAudit} = p_{cvr} - .4$		
	Base	Cons	Bayes	Base	Cons	Bayes	Base	Cons	Bayes	Base	Cons	Bayes	Base	Cons	Bayes
1							704	824	704	1053	855	816	1307	924	750
.9							729	842	738	1094	893	840	1347	938	763
.8				780	810	981	817	824	773	1135	938	839	1347	975	740
.7				856	816	999	881	879	830	1139	975	836	1373	1020	768
.6	780	810	1129	860	842	1011	930	944	856	1165	975	850	1377	1071	754
.5	790	817	1164	871	900	1010	1028	938	920	1160	1052	823	1347	1153	748
.4	816	842	1203	905	924	1062	1071	975	1003	1100	1093	791	1271	1184	742
.3	816	893	1216	886	960	1079	1123	1020	1031	1109	1108	808			
.2	816	900	1290	861	975	1068	1128	1071	1097	1058	1184	815			
.1	780	938	1301	820	1020	1097	1138	1108	1104						
0	739	1020	1347	816	1102	1058	1184	1184	1184						

Table 4: 95% of number of ballots for Kaplan-Markov comparison audit, $\alpha = .05, \gamma = 1.1, \mu = .01$.

An environment is honest *if it responds to any requested identifier with a matching ballot* **b** *and an interpretation* $I^* \in \mathtt{I}_{\mathbf{b}}$ *when such a ballot exists.*

Definition 20. *Let* $E = (C, \mathbf{B}, \mathsf{S})$ *be an election. For a competitive auditor,* Judge*, environment* Env*, and a sequence of conservative CVR tables* $\mathsf{cvr}_1^+, \ldots, \mathsf{cvr}_k^+$*, define*

$$\mathsf{CompeteA}[\mathtt{Judge}, \mathtt{Env}; E; \mathsf{cvr}_1^+, \ldots, \mathsf{cvr}_k^+]$$

to be the result of the audit (returned by Judge*) described in Figure 5. As both the auditor and the environment may be randomized procedures, this is a random variable taking values in the set* $C \cup \{\text{Inconclusive}\}$.

Definition 21 (Disagreement Set)**.** *Define the Omission Set as* $\mathtt{Omission}(\mathsf{cvr_A}, \mathsf{cvr_B}) = \mathtt{Idents}(\mathsf{cvr_A}) \setminus \mathtt{Idents}(\mathsf{cvr_B})$.

Define the Conflict Set as $\mathtt{Conflict}(\mathsf{cvr_A}, \mathsf{cvr_B}) = \{\iota \mid \iota \in \mathtt{Idents}(\mathsf{cvr_A}) \cap \mathtt{Idents}(\mathsf{cvr_B})$ *and* $P_{\mathsf{cvr_A},\iota} \cap P_{\mathsf{cvr_B},\iota} = \emptyset\}$

The disagreement set is the set

$$\mathtt{Disagree}(\mathsf{cvr_A}, \mathsf{cvr_B}) = \begin{matrix}\mathtt{Omission}(\mathsf{cvr_A}, \mathsf{cvr_B}) \cup \\ \mathtt{Conflict}(\mathsf{cvr_A}, \mathsf{cvr_B})\end{matrix}$$

The identifiers ι *in the disagreement set are those for which* $\mathsf{cvr_A}$ *makes a claim about a ballot in* **B** *that is inconsistent with* $\mathsf{cvr_B}$*, in the sense that either*

(1) $\mathsf{cvr_B}$ *does not recognize that the ballot exists at all* ($\mathtt{Omission}(\mathsf{cvr}_i, \mathsf{cvr}_j)$) *or*
(2) $\mathsf{cvr_A}$ *and* $\mathsf{cvr_B}$ *make contradictory claims about ballot interpretation* ($\mathtt{Conflict}(\mathsf{cvr_A}, \mathsf{cvr_B})$).

Intuitively, a CVR $\mathsf{cvr_B}$ will be disqualified if a ballot **b** is found during the judgment phase with either i) an identifier ι that was not reported on $\mathsf{cvr_B}$ or ii) an interpretation I^* that was not in the set of possible interpretations $P_{\mathsf{cvr_A},\iota}$.

Definition 22. *(Disqualification Function) Define function* $\mathtt{Disqual} : \Sigma^* \to \{0, 1\}$ *that takes in an identifier,* ι*, an interpretation,* I^**, and a CVR,* cvr.

$$\mathtt{Disqual}(\iota, I^*, \mathsf{cvr}) = \begin{cases} 1 & \iota \notin \mathtt{Idents}(\mathsf{cvr}) \vee I^* \notin P_{\mathsf{cvr},\iota}, \\ 0 & \textit{otherwise.} \end{cases}$$

The intuition is that Disqual looks up the identifier ι in cvr and outputs a disqualification vote if the provided interpretation I^* is not present on the row $\mathsf{cvr}(\iota)$.

We now analyze the competitive auditor $\mathtt{Judge}^t$ in Figure 6. The parameter t is a positive integer that calibrates the robustness of the audit to possible interpretation errors on the part of the advocates; see Theorems 5 and 6.

6.2 Analysis of the Competitive Auditor $\mathtt{Judge}^t$

In this section, we make two claims arguing about the completeness and soundness of Judge. We first consider these claims in the case that an "honest" party makes no errors and then consider the setting with errors in the next subsection.

Theorem 3 shows that if there is an election winner W and some $\mathsf{cvr_W}$ is consistent with the physical ballots, then W will be the output of Judge. This theorem assumes an honest environment. Recall that (i.) the consistency of a CVR means that for each ballot, the CVR submits a superset of the actual interpretations for that ballot and (ii.) a candidate is an election winner only if it wins regardless of the interpretation of ballots with multiple interpretations.

Theorem 4 shows that if a candidate is an election loser L and one of the submitted cvrs is canonical, then L will never be the output of Judge. This theorem does not require an honest environment. Recall that a CVR is canonical if for every ballot it declares exactly the set of interpretations on the physical ballots. Intuitively, since a loser loses no matter the interpretation declared by the environment, a canonical CVR is never disqualified by Judge and L can never be the output.

The main requirement for both theorems is that some advocate asserts the full set of interpretations for each physical ballot on the CVR; this may be either a superset or the exact set, depending on the circumstances. An advocate is disqualified when a ballot is delivered with a ground-truth interpretation that does not appear in the advocate's CVR. Note that any ballot for which no ground-truth interpretation appears on the CVR is sure to disqualify the advocate and, in fact, the

Competitive auditor (Judge)–Environment (Env) game

$$\mathsf{CompeteA}[\mathtt{Judge},\mathtt{Env};\mathbf{B},C,\mathsf{S}].$$

(1) **Advocate CVR generation**. For each candidate $\mathsf{A} \in C$, an advocate for A—with access to the full candidate set C and the physical ballots $\mathbf{B}$—generates a cast-vote record table, $\mathsf{cvr}^+_{\mathsf{A}}$.

(2) **Setup**.

(a) **Candidates, ballots, and CVR delivery (to Env).** $(C,\mathbf{B},\{\mathsf{cvr}^+_{\mathsf{A}} \mid \mathsf{A} \in C\})$ are given to the environment Env.

(b) **Candidate list, ballot manifest, and CVR delivery (to Judge).** $(C,\mathsf{S},\{\mathsf{cvr}^+_{\mathsf{A}} \mid \mathsf{A} \in C\})$ are given to the auditor Judge.

(3) **Audit**. Judge repeatedly makes a request to Env, or chooses to conclude the audit:

- **A ballot request**. Judge requests a ballot from Env with identifier $\iota \in \{0,1\}^*$.

  (a) Env either responds either with an interpretation I where $I \in T_{\mathbf{b}}$ for some ballot $\mathbf{b}$ for which $\iota = \iota_{\mathbf{b}}$ or responds with $\perp$ (meaning **No ballot**).

(4) **Conclusion**. Judge either returns:

A **is the winner** meaning "The audit is consistent with victory by candidate A," or

Inconclusive meaning "Audit inconclusive."

Figure 5: The competitive auditing game CompeteA with auditor Judge, environment Env, election E, and advocate CVRs $\{\mathsf{cvr}^+_{\mathsf{A}} \mid \mathsf{A} \in C\}$.

analysis only requires this weaker statement.

Theorem 3 (Competitive completeness)**.** *Let* $\mathtt{Judge}^t$ *be the competitive auditor with sample parameter* $t \in \mathbb{Z}^+$*. Let* $E = (C,\mathbf{B},\mathsf{S})$ *be an election with winner* $\mathsf{W} \in C$*. Let* cvr_* *be consistent with E and declare* W *to be the winner. Let* $\mathsf{cvr}_1,\ldots,\mathsf{cvr}_{k-1}$ *be any collection of CVRs. Then*

$$\mathsf{CompeteA}[\mathtt{Judge}^t,\mathtt{Env};E;\mathsf{cvr}_*,\mathsf{cvr}_1,\ldots,\mathsf{cvr}_{k-1}] = \mathsf{W},$$

for any honest environment Env. *Furthermore,* $\mathtt{Judge}^t$ *completes the audit with no more than* $tk(k-1)$ *ballot samples.*

We remark that the conclusion of the Theorem holds even for $t = 1$, in which case that auditor only draws $k(k-1)$ ballots. Larger values of t are relevant for managing interpretation errors; see the next subsection.

Theorem 4 (Competitive soundness)**.** *Let* $\mathtt{Judge}^t$ *be the competitive auditor with integer sample parameter* $t > 0$*. Let* $E = (C,\mathbf{B},\mathsf{S})$ *be an election with loser* $\mathsf{L} \in C$*. Let* cvr_* *be the canonical CVR for E. Let* $\mathsf{cvr}_1,\ldots,\mathsf{cvr}_{k-1}$ *be any collection of CVRs. Then*

$$\Pr[\mathsf{CompeteA}[\mathtt{Judge},\mathtt{Env};E;\mathsf{cvr}_*,\mathsf{cvr}_1,\ldots,\mathsf{cvr}_{k-1}] = \mathsf{L}] = 0,$$

for any environment Env. *Furthermore,* $\mathtt{Judge}^t$ *completes the audit with no more than* $tk(k-1)$ *ballot samples.*

Competitive auditor $\mathtt{Judge}^t[C,\mathsf{S};\mathsf{cvr}_1,\ldots,\mathsf{cvr}_k]$;

The $\mathtt{Judge}^t$ immediately outputs Inconclusive if any advocate announces duplicate labels. For each submitted CVR $\mathsf{cvr}_{\mathsf{A}}$, define $\mathtt{Idents}_{\mathsf{A}} := \mathtt{Idents}(\mathsf{cvr}_{\mathsf{A}}) = \{\iota \mid$ the identifier ι appears in $\mathsf{cvr}_{\mathsf{A}}\}$ and, for an identifier $\iota \in \mathtt{Idents}_{\mathsf{A}}$, let $\mathsf{P}_{\mathsf{A},\iota}$ denote the set of interpretations associated with ι in $\mathsf{cvr}_{\mathsf{A}}$.

(1) Remove from consideration any CVRs that do not have size S.

(2) For each ordered pair (A,B) for which the (remaining) CVRs $\mathsf{cvr}_{\mathsf{A}}$ and $\mathsf{cvr}_{\mathsf{B}}$ are contradictory:

- Let $\iota^{(1)},\ldots,\iota^{(t)}$ be independent and uniformly random elements of $\mathtt{Disagree}(\mathsf{cvr}^+_{\mathsf{A}},\mathsf{cvr}^+_{\mathsf{B}})$. For each $s \in \{1,\ldots,t\}$, request the ballot matching $\iota^{(s)}$. Let $I^{(1)},...,I^{(t)}$ denote the sequence of returned identifiers (including $\perp$).
- For each request s, we define the sequence $\mathtt{Disqual}_s = \mathtt{Disqual}(\iota^{(s)},I^{(s)},\mathsf{cvr}^+_{\mathsf{B}})$.
- Finally, if $\sum_{s=1}^{t} \mathtt{Disqual}_s > t/2$, disqualify $\mathsf{cvr}^+_{\mathsf{B}}$.

(3) Remove all disqualified CVRs from consideration.

(4) If there remains any pair of contradictory CVRs, or no remaining CVR declares victory, the audit concludes with Inconclusive.

(5) Otherwise, a winner W is declared by a remaining CVR that is not contradicted by any other remaining CVR and the audit concludes with "W is the winner."

Figure 6: The competitive auditor $\mathtt{Judge}^t$, with input C, S, and advocate CVRs $\{\mathsf{cvr}^+_{\mathsf{A}} \mid \mathsf{A} \in C\}$ where there are k submitted CVRs. The parameter t determines the number of ballot samples collected by the auditor: in particular, the auditor requests no more than $tk(k-1)$ ballots.

6.3 Handling Advocacy Errors

As producing cast-vote records is a process involving manipulation and interpretation of physical ballots, it may be unrealistic to demand that even a well-intentioned auditor can produce a perfectly consistent (or canonical) CVR. Occasional mistakes may occur during ballot interpretation, there may be some uncertainty about establishing a set I that necessarily contains all interpretations that an auditing board may assign to a ballot, and ballots may be overlooked. As we insist in our setting that CVRs have size consistent with the election, we note that an advocate can always pad out a CVR by adding invented identifiers that do not correspond to any ballot. All told, these considerations lead us to consider CVRs that are "nearly" consistent, defined formally below.

Theorems 5 and 6 extend these theorems to the case when "honest" parties make a small fraction of errors.

Definition 23. *Let $E = (C, \mathsf{S}, \mathbf{B})$ be an election. We say that a* cvr *(over C) is $(1-\epsilon)$-consistent if the size of the CVR is* S *and there is a subset C of* $\mathtt{Idents}(\mathsf{cvr})$ *of size at least $(1-\epsilon)\mathsf{S}$ on which* cvr *is consistent, which is to say that for each $\iota \in C$, there is a ballot $\mathbf{b} \in \mathbf{B}$ for which $\iota = \iota_{\mathbf{b}}$ and $\mathtt{I}_{\mathbf{b}} \subset \mathtt{I}_\iota$ (where $\mathtt{I}_\iota$ is the set of interpretations declared by* cvr *for ι). Define a $(1-\epsilon)$-canonical* cvr *analogously.*

Let $Bin[\gamma, t; t/2]$ denotes the tail of the binomial distribution: specifically, if $X_1, \ldots, X_t$ are independent Bernoulli random variables for which $\Pr[X_i = 1] = \gamma$, then $Bin[\gamma, t; t/2]$ is the probability that $\sum_i X_i \geq t/2$.

Theorem 5 (Competitive completeness with imperfect advocacy)**.** *Let $\mathtt{Judge}^t$ be the competitive auditor with integer sample parameter $t > 0$. Let $E = (C, \mathbf{B}, \mathsf{S})$ be an election with winner $\mathsf{W} \in C$. Let cvr_* be $(1-\epsilon)$-consistent with E and declare W to be the winner with margin $\mu_{\mathsf{cvr}_*} > 4\epsilon$. Let $\mathsf{cvr}_1, \ldots, \mathsf{cvr}_{k-1}$ be any collection of CVRs and let $\mathtt{Env}$ be an honest environment. Then*

$$\Pr[\mathsf{CompeteA}[\mathtt{Judge}^t, \mathtt{Env}; E; \mathsf{cvr}_*, \mathsf{cvr}_1, \ldots, \mathsf{cvr}_{k-1}] = \mathsf{W}] \geq 1 - 2(k-1)Bin[2\epsilon/\mu_{\mathsf{cvr}_*}, t; t/2].$$

For $\gamma < 1/2$ (guaranteed above by $4\epsilon < \mu_{\mathsf{cvr}_*}$), the quantity $Bin[\gamma, t; t/2] = \exp(-\omega(t))$. Under the mild assumption that $\epsilon < \mu_{\mathsf{cvr}_*}/6$, say, the number of samples required by such an audit to achieve any particular risk limit is independent of μ_{cvr_*}. An advocate may reduce μ_{cvr_*} in an effort to be consistent with the underlying ballots.

Theorem 6 (Competitive soundness)**.** *Let $\mathtt{Judge}^t$ be the competitive auditor with integer sample parameter $t > 0$. Let $E = (C, \mathbf{B}, \mathsf{S})$ be an election with loser $\mathsf{L} \in C$. Let cvr_* be a $(1-\epsilon)$-canonical CVR for E. Define*

$$\mu_{\mathsf{cvr}_*} = \frac{\max_{\mathsf{A} \in C \setminus \mathsf{L}}(\mathsf{cvr}_*^-(\mathsf{A}) - \mathsf{cvr}_*^+(\mathsf{L}))}{\mathsf{S}}.$$

Let $\mathsf{cvr}_1, \ldots, \mathsf{cvr}_{k-1}$ be any collection of CVRs. Then

$$\Pr[\mathsf{CompeteA}[\mathtt{Judge}^t, \mathtt{Env}; E; \mathsf{cvr}_*, \mathsf{cvr}_1, \ldots, \mathsf{cvr}_{k-1}] = \mathsf{L}] \leq 2(k-1)Bin[2\epsilon/\mu, t; t/2]$$

for any environment $\mathtt{Env}$. Furthermore, $\mathtt{Judge}^t$ completes the audit with no more than $tk(k-1)$ ballot samples.

6.4 Remarks on Bayesian competitive audits

While we do not present any details, we make a few remarks about competitive audits in the Bayesian case. This calls for each advocate to submit a Bayesian CVR. Now we note that identification of a collection of ballots D on which the two CVRs substantively disagree permits us to define two different probability distributions on this set of ballots: (i.) select a ballot at random from D, and (ii.) output an interpretation given by the CVR in question. Then the classical sequential probability ratio test provides a statistical test for distinguishing the two models [42]. This can serve as the statistical mechanism for selecting between contradictory CVRs.

Acknowledgements

These results were developed as part of a collaboration with the Office of the Connecticut Secretary of State and, additionally, were supported in part by a grant from that office. Discussions with anonymous reviewers improved the narrative and technical treatment. The authors also thank Ron Rivest for his helpful comments. B.F. is supported by NSF Grants #2232813 and #2141033 and the Office of Naval Research. A.R was supported by NSF Grant #1801487.

References

[1] Mark Lindeman and Philip B Stark. A gentle introduction to risk-limiting audits. *IEEE Security & Privacy*, 10(5):42–49, 2012.

[2] National Academies of Sciences Engineering and Medicine. *Securing the Vote: Protecting American Democracy*. National Academies Press, 2018.

[3] Filip Zagórski, Grant McClearn, Sarah Morin, Neal McBurnett, and Poorvi L Vora. Minerva–an efficient risk-limiting ballot polling audit. In *USENIX Security Symposium*, pages 3059–3076. USENIX Association, 2021.

[4] Andrea Bajcsy, Ya-Shian Li-Baboud, and Mary Brady. Systematic measurement of marginal mark types on voting ballots, 2015. https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8069.pdf.

[5] Benjamin Fuller, Abigail Harrison, and Alexander Russell. Adaptive risk-limiting comparison audits. In *IEEE Symposium on Security and Privacy*, pages 2002–2019, Los Alamitos, CA, USA, may 2023.

[6] Philip B Stark. Auditing a collection of races simultaneously. *arXiv preprint arXiv:0905.1422*, 2009.

[7] Philip B Stark. Super-simple simultaneous single-ballot risk-limiting audits. In *EVT/WOTE*, 2010.

[8] Michael J Higgins, Ronald L Rivest, and Philip B Stark. Sharper p–values for stratified election audits. *Statistics, Politics, and Policy*, 2(1), 2011.

[9] Kellie Ottoboni, Philip B Stark, Mark Lindeman, and Neal McBurnett. Risk-limiting audits by stratified union-intersection tests of elections (SUITE). In *International Joint Conference on Electronic Voting*, pages 174–188. Springer, 2018.

[10] Philip B Stark. Sets of half-average nulls generate risk-limiting audits: Shangrla. In *Financial Cryptography and Data Security*, pages 319–336. Springer, 2020.

[11] Ian Waudby-Smith, Philip B Stark, and Aaditya Ramdas. Rilacs: Risk limiting audits via confidence sequences. In *International Joint Conference on Electronic Voting*, pages 124–139. Springer, 2021.

[12] Michelle Blom, Jurlind Budurushi, Ronald L Rivest, Philip B Stark, Peter J Stuckey, Vanessa Teague, and Damjan Vukcevic. Assertion-based approaches to auditing complex elections, with application to party-list proportional elections. In *International Joint Conference on Electronic Voting*, pages 47–62. Springer, 2021.

[13] Colorado Secretary of State. 2020 general election risk-limiting audit discrepancy report, 2020. https://www.sos.state.co.us/pubs/elections/RLA/2020/general/DiscrepancyReport.pdf.

[14] A. Russell, L. Michel, B. Fuller, J. Wohl, and G. Johnson. Statistical analysis of post-election audit data for the November 8, 2022 state election. https://voter.engr.uconn.edu/wp-content/uploads/sites/3651/2023/02/2022-11-08-hand-count-statistics.pdf.

[15] Kai Wang, Eric Rescorla, Hovav Shacham, and Serge J Belongie. Openscan: A fully transparent optical scan voting system. *EVT/WOTE*, 10:1–13, 2010.

[16] Philip B. Stark. Conservative statistical post-election audits. *The Annals of Applied Statistics*, 2(2):550 – 581, 2008.

[17] Matthew Bernhard. Risk-limiting audits: A practical systematization of knowledge. In *International Joint Conference on Electronic Voting*, 2021.

[18] Lynn Garland, Neal McBurnett, Jennier Morrell, Marian K. Schneider, and Stephanie Singer. Principles and best practices for post-election tabulation audits, 2018.

[19] Joseph Lorenzo Hall, Luke W Miratrix, Philip B Stark, Melvin Briones, Elaine Ginnold, Freddie Oakley, Martin Peaden, Gail Pellerin, Tom Stanionis, and Tricia Webber. Implementing risk-limiting post-election audits in California. In *EVN/WOTE*, Montreal, Canada, 2009.

[20] Mark Lindeman, Philip B Stark, and Vincent S Yates. Bravo: Ballot-polling risk-limiting audits to verify outcomes. In *EVT/WOTE*, 2012.

[21] Stephen Checkoway, Anand Sarwate, and Hovav Shacham. Single-ballot risk-limiting audits using convex optimization. In *Proceedings of the 2010 International Conference on Electronic Voting Technology/Workshop on Trustworthy El ections*, EVT/WOTE'10, pages 1–13, USA, 2010. USENIX Association.

[22] Philip B. Stark. Cast: Canvass audits by sampling and testing. *IEEE Transactions on Information Forensics and Security*, 4(4):708–717, 2009.

[23] Michelle Blom, Peter J Stuckey, Vanessa Teague, and Damjan Vukcevic. A first approach to risk-limiting audits for single transferable vote elections. In *International Conference on Financial Cryptography and Data Security*, pages 366–380. Springer, 2022.

[24] Philip B Stark. Efficient post-election audits of multiple contests: 2009 california tests. In *CELS 2009 4Th annual conference on empirical legal studies paper*, 2009.

[25] Kellie Ottoboni, Matthew Bernhard, J Alex Halderman, Ronald L Rivest, and Philip B Stark. Bernoulli ballot polling: a manifest improvement for risk-limiting audits. In *International Conference on Financial Cryptography and Data Security*, pages 226–241, 2019.

[26] Philip B. Stark. Risk-limiting postelection audits: Conservative P-values from common probability inequalities. *IEEE Transactions on Information Forensics and Security*, 4(4):1005–1014, 2009.

[27] Philip B Stark. Alpha: audit that learns from previously hand-audited ballots. *The Annals of Applied Statistics*, 17(1):641–679, 2023.

[28] Jorge H Banuelos and Philip B Stark. Limiting risk by turning manifest phantoms into evil zombies. *arXiv preprint arXiv:1207.3413*, 2012.

[29] Filip Zagórski, Grant McClearn, Sarah Morin, Neal McBurnett, and Poorvi L Vora. The athena class of risk-limiting ballot polling audits. *arXiv preprint arXiv:2008.02315*, 2020.

[30] Jennifer Morrell. Knowing it's right, part two. risk-limiting audit implementation workbook., 2019.

[31] Jennie Bretschneider, Sean Flaherty, Susannah Goodman, Mark Halvorson, Roger Johnston, Mark Lindeman, Ronald L. Rivest, Pam Smith, and Phillip B. Stark. Risk-limiting post-election audits: Why and how, 2012.

[32] Amanda K Glazer, Jacob V Spertus, and Philip B Stark. Bayesian audits are average but risk-limiting audits are above average. In *E-Vote-ID 2020*, pages 84–94. Springer, 2020.

[33] Josh Benaloh, Kammi Foote, Philip B Stark, Vanessa Teague, and Dan S Wallach. VAULT-style risk-limiting audits and the Inyo county pilot. *IEEE Security & Privacy*, 19(4):8–18, 2021.

[34] Douglas W Jones, Sunoo Park, Ronald L Rivest, and Adam Sealfon. Scan, shuffle, rescan: Machine-assisted election audits with untrusted scanners. In *Financial Cryptography*, 2024.

[35] John H Reif. The complexity of two-player games of incomplete information. *Journal of computer and system sciences*, 29(2):274–301, 1984.

[36] M. Kiwi, C. Lund, D. Spielman, A. Russell, and R. Sundaram. Alternation in interaction. *Computational Complexity*, 9(3):202–246, 2000.

[37] László Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational complexity*, 1:3–40, 1991.

[38] Benjamin Fuller, Rashmi Pai, and Alexander Russell. The decisive power of indecision: Low-variance risk-limiting audits and election contestation via marginal mark recording. *arXiv preprint arXiv:2402.06515*, 2024.

[39] Rashmi Pai. Marginal mark RLA tools, 2024.

[40] Abigail Harrison. Adaptive RLA tools, 2023.

[41] Jacob V Spertus. Cobra: Comparison-optimal betting for risk-limiting audits. In *International Conference on Financial Cryptography and Data Security*, pages 95–109. Springer, 2023.

[42] Abraham Wald. Sequential tests of statistical hypotheses. In *Breakthroughs in statistics: Foundations and basic theory*, pages 256–298. Springer, 1992.