Adaptive Risk-Limiting Comparison Audits

Benjamin Fuller, Abigail Harrison, and Alexander Russell

{benjamin.fuller,abigail.harrison,acr}@uconn.edu
Voting Technology Research Laboratory
University of Connecticut

February 19, 2024

Abstract

Risk-limiting audits (RLAs) are rigorous statistical procedures meant to detect invalid election results. RLAs examine paper ballots cast during the election to statistically assess the possibility of a disagreement between the winner determined by the ballots and the winner reported by tabulation. The design of an RLA must balance risk against efficiency: "risk" refers to a bound on the chance that the audit fails to detect such a disagreement when one occurs; "efficiency" refers to the total effort to conduct the audit.

The most efficient approaches—when measured in terms of the number of ballots that must be inspected—proceed by "ballot comparison." However, ballot comparison requires an (untrusted) declaration of the contents of each cast ballot, rather than a simple tabulation of vote totals. This "cast-vote record table" (CVR) is then spot-checked against ballots for consistency. In many practical settings, the cost of generating a suitable CVR dominates the cost of conducting the audit which has prevented widespread adoption of these sample-efficient techniques.

We introduce a new RLA procedure: an "adaptive ballot comparison" audit. In this audit, a global CVR is never produced; instead, a three-stage procedure is iterated: 1) a batch is selected, 2) a CVR is produced for that batch, and 3) a ballot within the batch is sampled, inspected by auditors, and compared with the CVR. We prove that such an audit can achieve risk commensurate with standard comparison audits while generating a fraction of the CVR. We present three main contributions: (1) a formal adversarial model for RLAs; (2) definition and analysis of an adaptive audit procedure with rigorous risk limits and an associated correctness analysis accounting for the incidental errors arising in typical audits; and (3) an analysis of efficiency. Finally, we observe that our results have security ramifications for conventional comparison RLAs: in particular, we note that ballot identifiers need not be assumed unique in order to preserve the standard statistical guarantees.

1 Introduction

We consider the task of conducting a risk-limiting audit of a conventional election based on paper ballots. This framework calls for the election to be organized in three stages:

Ballot casting: Voters mark paper ballots with their preferences, producing a *voter-verified paper trail* [7, 20].

Tabulation: Ballots are tabulated and aggregated by (untrusted) tabulators forming a tabulated outcome.

Storage: Ballots are stored in preparation for audits.

The tabulation and storage phases must ensure "ballot invariance": no ballots may be destroyed, introduced or modified. Many countries across the world and municipalities across the United States carry out elections modeled on this ideal.

Risk-limiting audits (RLAs) are techniques for testing the veracity of the tabulation step [15]. Assuming ballot invariance, RLAs explicitly bound the probability that a disagreement between the tabulated winner and the winner determined by the paper trail is undetected by the audit. RLAs must be transparent: it must be possible for an external observer to verify that the audit was conducted properly. While a variety of specific methods have been proposed, the basic landscape is dominated by two approaches (see the discussion in [3–5, 13, 15, 16, 18, 23, 24, 29, 33, 34, 36] and Section 1.2): a) "polling" randomly sampled ballots to directly estimate margins, and b) "comparing" randomly sampled ballots (or groups of ballots) against a *cast-vote record table*. We discuss this approach in detail below.

As mentioned above, the aim of the audit is to detect circumstances where the tabulated winner of the election is not, in fact, the winner as determined by the paper trail. The paper trail itself—typically consisting of paper ballots marked directly by voters—is assumed to have an unambiguous interpretation that serves as the ground truth for the audit.[1] The *risk* of the audit, denoted throughout by α, is (an upper bound on) the probability that the audit incorrectly concludes an election to be correct when the tabulated and ground truth outcomes disagree.

Polling. A ballot polling audit proceeds by drawing a collection of randomly sampled ballots; the votes cast on these sampled ballots are then used to statistically infer the winner of the election. For example, in a single two-candidate race, uniform sampling of ballots yields a direct estimate of the *diluted margin* μ of the race, equal to the number of votes cast for the winner minus those for the loser divided by the total number of ballots cast that contain the race. This estimate achieves risk α, correctly determining the winner with probability $1 - \alpha$, after sampling $\Theta(\log(\alpha)/\mu^2)$ ballots.

Comparison. Ballot comparison audits, in contrast to polling audits, require additional metadata about the election: a *cast-vote record table* (CVR) that declares the votes cast on each ballot in the election. This additional metadata—even though it is not assumed to be correct by the auditor—yields a dramatic reduction in the number of ballot examinations necessary for the same risk level: in particular, only $\Theta(\log(\alpha)/\mu)$ ballots need to be examined to achieve risk α, with μ as above.[2]

This would appear to establish ballot comparison as the dominant auditing paradigm as the number of ballots that must be examined scales more favorably in the margin. However, we are not aware of any mass-produced *voter-facing* tabulator that produces ballot-identifying CVRs suitable for a risk-limiting audit. (See the discussion in Section 1.1.2.) For elections with voting facing tabulation, CVRs must then be produced during a second round of processing by *transitive tabulators* that are specifically designed to produce CVRs. (The terminology here is meant to mimic the language of a "transitive ballot comparison audits" [15].) Unfortunately, this second round of processing—for reasons we discuss in detail below—tends to dominate the cost of the ballot comparison audit.[3]

For example, Rhode Island's RLA pilot estimated the setup cost for a ballot comparison audit to take roughly six times as long as conducting the audit [9, Table 2].[4] This was presumably the major factor in Rhode Island's adoption of ballot polling (rather than ballot comparison) for its RLA of the 2020 presidential election [14]. Connecticut's pilot found this ratio to be much higher, with CVR generation taking 99% of the audit execution [11, Section 6.2].[5] These pilots used different tabulators and different methods for identification—RI imprinted using a high speed scanner, while CT manually applied identifiers. While these figures are from pilots, they indicate that CVR generation is an important cost factor in the design and implementation of ballot comparison RLAs.

To conclude, ballot comparison audits offer significant advantages in ballot sample size. However, in many settings the generation of CVRs is an expensive, separate step that renders the approach non-competitive with ballot polling except in circumstances with small margins. We are not aware of any statewide election procedures in the United States that combine voter-facing tabulators with the efficiency benefits afforded by ballot comparison RLAs.

[1] In practice, audits may have to contend with disagreements among human interpretations of the paper trail and, in such cases, must provide a mechanism (majority vote, say) for yielding a final interpretation.

[2] The use of asymptotic notation here is meant to highlight how the efficiency of the audit—that is, the number of ballots that must be examined—scales with margin. Of course, practice demands explicit bounds which have been developed by a sizable literature; see [32] for a survey. We remark that the complexity can also be parameterized in terms of the *tabulated diluted margin*, equal to the margin defined above with the tabulated vote totals. See [35] for a detailed discussion.

[3] There are tabulators, such as the ES&S DS850 `https://www.essvote.com/products/ds850/`, designed for central tabulation that produce CVR tables suitable for comparison audits. These tabulators directly imprint identifiers on physical ballots in order to address the identification problem. Colorado, which uses mail-in voting and centrally processes ballots by county, uses such tabulators to support ballot comparison audits.

[4] This assumes a 10% margin and 10% risk limit with a 75% chance for the audit to complete.

[5] This analysis considers a 2% margin, 5% risk limit, and considers the expected number of ballots retrieved. The fraction of time dedicated to CVR generation increases as margin increases; one selects fewer ballots.

1.1 Our results: Adaptive Risk-Limiting Audits

Typical ballot storage organizes ballots into physical *batches*; in the context of ballot comparison audits, these provide a direct means for referencing and locating individual ballots. The election CVR required for the ballot comparison audit is then logically composed of a *batch CVR* associated with each batch.[6] To emphasize this distinction, we refer to the full election CVR as a *global CVR*. In cases where the total number of batches exceeds the number of ballots sampled during the audit, some batch CVRs will not be directly examined during the audit procedure. For example, Florida tabulates by precinct and has over 6000 precincts [22]. Even at a 1% margin, a comparison RLA would only select approximately 20% of these precincts for audit (see Table 1).

Development and analysis of adaptive risk-limiting audits. Considering the high cost of CVR generation, we propose an "on-the-fly" procedure for risk-limiting election audits by ballot comparison. The informal procedure is as follows. (The formal auditor is in Figure 3.)

(1) Ensure that the tabulation is consistent with batch sizes.

(2) Repeatedly (or, optionally, in parallel):

   (a) Sample a batch with probability proportional to its size. Request a CVR to be generated for the sampled batch. (The CVR contains a sequence of rows, each containing a ballot identifier and purported votes appearing on the corresponding ballot.)
   (b) Ensure that the produced batch CVR declares the same total size and votes for the winning and losing candidates as the tabulation of the batch, and declares a unique ballot identifier in each row.
   (c) Sample a row from the CVR and request a ballot with the identifier appearing in the row.
   (d) Compare the retrieved ballot with the votes declared in the CVR row and record their discrepancy.

(3) Compute risk using an appropriate statistical test.

We call this an **adaptive risk-limiting ballot comparison audit** because batch CVRs are created "on the fly" and only for batches for which ballot samples are actually drawn. The audit can additionally incorporate mechanisms to correct consistency failures that might arise in the checks of (1) and (2)b. The procedure can also benefit from carrying out CVR generation and sampling for different batches in parallel, known as *audit rounds*. As such, our techniques are never more costly than a conventional ballot comparison RLA.

Our main result is a rigorous analysis of the formal procedure which shows that *with the same number of ballot samples, adaptive comparison audits can achieve risk commensurate with standard comparison RLAs.*

Adaptive ballot comparison audits can provide significant efficiency improvements for RLAs of elections carried out using tabulators that do not provide ballot-identifying CVRs (that would directly support comparison RLAs). Twenty-three of the 50 United States fall into this category. We use Connecticut and Florida as running examples. They differ widely in size: Connecticut is 29th in population, Florida is 4th. In addition, Connecticut uses a transitive tabulator that produces CVRs [1]. Using precinct sizes from the 2020 general election as an example, for Connecticut, at a 1% margin and 5% risk limit, 78% of the CVR is generated; for larger margins, as little as 6% of the CVR is generated. For Florida, at a 1% margin and 5% risk limit, only 22% of the CVR is generated; for larger margins, as little as 1% of the CVR is generated. See Table 1 for full cost estimates and Appendix A for justification.

Adaptive RLAs moderate between the extremes of polling (which is efficient at large margins) and comparison (which is efficient at small margins). To explain, the overall time to conduct an adaptive RLA scales with (the inverse of) margin, while comparison has a large upfront cost to generate the full CVR and polling requires a sample size that grows quadratically with (the inverse of) margin.

In addition to our adaptive ballot comparison methods, we introduce an **adaptive group comparison audit** in Section 7 that is intended for settings where ballots are grouped into small groups (e.g., size 50) that are interpreted together if selected. In this setting, no order needs to be kept inside of a group and ballots do not need to be individually identified.

[6] For the purposes of this article, the word "batch" means a set of ballots that are physically co-located with the standard assumption that the size of each batch is known with confidence. We also require that each batch has an (untrusted) tabulated total, which arises naturally when batches are collections of ballots that were tabulated together (or unions of such collections).

	α = 5% Risk Limit					α = 1% Risk Limit				
		Adaptive Comparison					Adaptive Comparison			
		CT		FL			CT		FL	
Margin	Ballots	Batches	% CVR	Batches	% CVR	Ballots	Batches	% CVR	Batches	% CVR
1%	1532	590	78%	1321	22%	1886	633	84%	1579	26%
2%	548	331	44%	515	8%	725	401	53%	672	11%
3%	366	244	32%	350	6%	484	304	40%	458	8%
4%	274	192	26%	264	4%	363	242	32%	348	6%
5%	220	160	21%	213	3%	290	202	27%	279	5%
10%	110	86	11%	108	2%	145	109	14%	141	2%
15%	74	59	8%	73	1%	97	76	10%	95	2%
20%	55	44	6%	54	1%	73	57	8%	72	1%

Table 1: Fraction of CVR generated using the Adaptive RLA method for different states, margins, and risk limits. The number of ballot samples is computed with `rlacalc` [17]. The percentage of CVR generated by the audit is determined by simulation; see further discussion in Appendix A.

1.1.1 The analytic challenge

The rigorous analysis of an adaptive ballot comparison RLA must contend with new phenomena that do not arise in the standard setting: in particular, the batch CVRs relevant for the audit may be *adaptively* determined as a function of the entire history of the audit. Previous analyses also make direct use of the global CVR in order to define the basic probability-theoretic events of interest; of course, in our setting this global CVR is not even defined. These considerations lead to several modeling and analytic challenges, which we briefly summarize.

A formal model for RLAs. The obligation to rigorously handle such adaptivity motivates us to lay out a formal model for risk-limiting audits—borrowing from the successful framework of cryptographic games—that makes explicit the assumptions and guarantees offered by the audit. Adopting this model, we then prove the new procedure is risk-limiting.

Completeness and reflecting "typical" auditing errors. Such modeling must satisfactorily address the issue of "completeness," by which we mean the ability of the audit to survive the anticipated errors introduced during practical audit proceedings, such as occasional inconsistencies in human ballot interpretation and mismatches in tabulated batch sizes and CVR-declared sizes.

Adaptive statistical tests. Finally, this adaptive setting places new demands on the underlying statistical tests employed by the audit. Typical ballot comparison audits consider tests that consume *discrepancy vectors* which indicate how selected ballots differ from the corresponding CVR rows [15]. In contrast to standard RLA procedures, which can be given a simple analytic treatment in terms of independent and identically distributed random samples (from a fixed discrepancy vector), we require tests that provide guarantees for a broader class of dependent random variables that reflect our adaptive setting. We formulate a specific "induced sub-martingale" condition sufficient for our auditing framework. As shown in Section 5.1, many natural statistical tests satisfy the condition including the Kaplan–Markov test used in the "super simple" ballot comparison method [29, 31–33], the open-source RLA software Arlo,[7] and our open-source prototype of the adaptive auditor (Github repository and Jupyter notebook). RLA software design is complex [3] and our prototype is meant to inform future development.

1.1.2 Motivating the formal auditing model

Our model provides explicit, rigorous answers to natural questions that may be obscured by informal treatments. For example:

[7] `https://www.voting.works/risk-limiting-audits`.

- Must ballot identifiers be unique as they appear on physical ballots and/or as they appear in a CVR? More broadly, must ballot identifiers be determined by trusted auditors?
- What convention should be adopted for treating mismatches in CVR batch size and tabulated batch size?
- What effect can the—possibly adversarial—destruction of ballots have on audit risk and efficiency?[8]

And, finally, the question that originally motivated the model:

- What effect can adaptive, adversarial selection of CVRs have on audit risk?

The model itself introduces two parties, the *Auditor* and the *Adversary*. Formally, we consider an election to be defined by a set of physical ballots and a set of tabulation results (which, of course, need not match the ballots). The Auditor carries out a specific, fixed auditing procedure of interest; the Adversary, on the other hand, is responsible for all of the untrusted aspects of the audit, such as CVR generation and access to ballots. The notion of *risk*, for a particular auditor of interest, is now a probability upper bound that is guaranteed to hold for all possible behaviors of the adversary.

This corresponds to a guarantee of the risk of the audit even under situations where a powerful malicious party is attempting to deceive the auditor; of course, the same guarantees hold in the less adversarial circumstances that typically hold in practice. The model also provides a precise method for reasoning about *completeness*, which reflects the behavior of the audit when interacting with "honest adversaries with incidental errors" that exhibit the behavior one would expect from tabulators, CVRs, and human ballot handlers. (See Section 6.)

Remarks on practical relevance and conventional ballot comparison audits. Adaptive RLAs will improve efficiency in large-scale elections that (1) adopt tabulators that do not generate CVRs, or tabulators that generate CVRs without ballot identifying information, (2) maintain the natural ballot batching determined by tabulation, which is to say that ballots tabulated together appear in the same batch, (3) yield a number of batches that exceeds the anticipated number of sampled ballots, and (4) possess a mechanism to produce CVRs with a corresponding means for identifying individual ballots. Currently, 23 US states satisfy these conditions accounting for roughly half of the US population.

Remarks on ramifications for conventional comparison audits. Even in the context of a conventional ballot comparison RLA (in which the full CVR is generated, typically by the tabulator itself), there are two benefits to these techniques:

(1) Our proofs show it is safe to selectively release only the portion of the global CVR corresponding to batches containing selected ballots. This improves the privacy of the audit.

(2) Our model directly specializes to the setting of conventional (non-adaptive) comparison RLAs. Thus, the fact that uniqueness of ballot identifiers is not necessary for RLA risk guarantees applies to traditional comparison audits as well. To the best of our knowledge, this is the first time this question has been considered.

Remarks on tabulators, CVRs and ballot marking. Comparison audits require a reliable means for identifying specific physical ballots in order to compare against the CVR. There are two natural means for such ballot identification: (1) the physical location of a ballot and (2) identifying marks ("serial numbers") directly printed on ballots. Identifying a ballot by physical location has typically been implemented by referring to the *position* of the ballot in a named stack or batch. How this issue is addressed depends on the details of the tabulator. *Voter-facing tabulators* are those that support direct interaction with voters, providing sufficient physical security and privacy features in order for voters to cast their ballots at the tabulator. Typical voter-facing tabulators intentionally avoid maintaining ballot order to protect voter privacy; thus the batching of ballots generated directly from such a tabulator is unsatisfactory for comparison audits. A further difficulty with ballot position—even with tabulators that do preserve order—is that the ordering is transient, subject to corruption during handling, and prone to errors during ballot indexing; Colorado, which has successfully used ballot order for identification, has observed a small but significant error rate [21].

[8]The reader excited to know the answers can refer to Section 3.3.

Printing identifying marks directly on ballots addresses these concerns. However, printing identifiers on ballots *prior* to voters casting their votes is a privacy concern. A natural alternative is to indelibly "imprint" ballots with identifiers during tabulation. Unfortunately, this complicates tabulator design: it involves additional hardware which must provide firm guarantees that marking cannot interfere with cast vote interpretation and, of course, must not leak voter identity.[9] As stated above, these tabulators do not preserve order, so even with identifier imprinting, finding a matching ballot would be complex and time-consuming. This may explain why no mainstream voter-facing tabulators provide this functionality. These considerations suggest that the efficiency of near-term ballot comparison audits with voter-facing tabulators will indeed depend heavily on CVR generation, which is the principle metric we optimize.

Options for (post-tabulation) CVR generation currently fall into two categories (1) high-speed, centralized tabulators that provide imprinting and (2) tabulators specifically designed for transitive use that produce CVRs corresponding to ballot identifiers applied in a separate ballot identification pass.

Finally, while the election security landscape is complicated, there are reasons to prefer voter-facing tabulation. Elections are secure and trustworthy when voter registration, authentication, ballot delivery, vote casting, tabulation, and auditing are tightly coupled. In this context, voter-facing tabulators provide a strong coupling of vote casting and tabulation.

1.2 Related work

Risk-limiting audits, as the term is now understood, were first articulated in 2008 by Stark [28]. Following this, a body of work laid down the foundations, including key assumptions and guarantees [3, 10, 12, 15]. As indicated earlier, a variety of specific methods have been explored, often with an eye to optimize certain practical settings [6, 15, 16, 28, 30, 33]. A significant literature has also developed around various generalizations and refinements, including (1) supporting various social choice functions [4, 34], (2) managing multiple races across jurisdictions [13, 24, 29, 31, 33], (3) explicit p-value estimates [2, 6, 13, 16, 23, 28, 32, 35, 36] and (4) implementation issues [3, 10, 12].

Structure of the paper. After reviewing preliminaries in Section 2, we present the following: (1) an adaptive auditor (Section 3) that defines the details of the adaptive audit procedure; (2) a comprehensive model of election auditing (Section 4) expressive enough to reflect adaptive and traditional comparison RLAs, (3) a proof that the adaptive RLA procedure is risk-limiting for many existing statistical tests (Section 5), (4) a completeness analysis establishing that the audits have desirable properties in the presence of errors encountered in practical audits (Section 6), and (5) an adaptive group comparison audit (Section 7).

2 Preliminaries

The two-candidate single-race setting. We consider an audit of a single first-past-the-post race with two candidates denoted W and L. By our naming convention, the candidate W is reported to have received more votes. The general case—with multiple candidates and races—can be essentially reduced to this simpler case by conducting audits for each winner-loser pair simultaneously. The p-values for these can be appropriately combined both for candidate pairs in the same race and across races. Additional approximations can simplify the accounting; see [33], which proposes several techniques.

Notation. We provide a quick overview of notation in Table 2; this is reviewed as we introduce the adversarial model. Throughout, we use boldface to refer to "physical" objects, such as individual ballots (typically denoted $\mathbf{b}$) or groups of ballots (typically $\mathbf{B}_\beta$). Variables determined by these physical objects are typically denoted with a super- or subscript ($X^\mathbf{b}$) with the understanding that they can be determined from the physical object.

We define $\mathbb{N} = \{0, 1, \ldots\}$ to be the natural numbers (including zero). For a natural number k, we define $[k] = \{1, \ldots, k\}$ (and $[0] = \emptyset$). We let $\Sigma = \{-2, -1, 0, 1, 2\}$, a set that will play a special role in our setting. In general, for a finite set X, we define X^* to be the set of all finite-length sequences over X; that is, $X^* = \{(x_1, \ldots, x_k) \mid k \geq 0, x_i \in X\}$.

[9] The DVSOrder vulnerability is a notable example of an implementation that violated this.

	Notation	Description
concepts	S	size
	W	tabulated winner
	L	tabulated loser
	μ	diluted margin
	α	risk limit
	$\mathbf{b}, \mathbf{B}_\beta$	physical ballot, batch of ballots
	D	discrepancy
modifiers	act	on ballots
	tab	in tabulation results
	cvr	in CVR

Table 2: Notation, reviewed in detail in Section 2.

Note that this includes a sequence of length 0 which we denote λ. Finally, we define $X^{\mathbb{N}}$ to be the set of all sequences $\{(x_0, x_1, \ldots) \mid x_i \in X\}$.

2.1 Election Definitions

We now set down the elementary definitions of elections, manifests, and CVRs. Our setting demands some generalizations and variants of concepts that are standard in the literature. In particular, we consider tabulations with batch data and a batch-specific notion of CVR. See Definition 5 and the preceding discussion.

Definition 1 (Ballot family; ballot conventions)**.** *A* ballot family *is a collection of physical* ballots *partitioned into disjoint sets denoted* $\mathbf{B}_1, \ldots, \mathbf{B}_k$*. As a matter of notation, the ballot family is denoted* $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_k)$ *and the sets are referred to as "batches." For the sake of brevity, we use* $\mathbf{b} \in \mathbf{B}$ *as shorthand for* $\mathbf{b} \in \bigcup \mathbf{B}_\beta$ *and use* $|\mathbf{B}|$ *as shorthand for* $\sum |\mathbf{B}_\beta|$*. Throughout, we reserve the variable k to refer to the number of batches.*

Physical ballots have three properties:

(1) There is an immutable interpretation of the votes contained on the ballot. Each $\mathbf{b} \in \mathbf{B}$ *determines a pair* $(\mathsf{W}_\mathbf{b}, \mathsf{L}_\mathbf{b})$*, where each* $\mathsf{W}_\mathbf{b}, \mathsf{L}_\mathbf{b} \in \{0, 1\}$*.*

(2) For any $\mathbf{b} \in \mathbf{B}$*, one can determine the batch to which the ballot belongs. This defines an index* $\beta_\mathbf{b} \in [k]$ *such that* $\mathbf{b} \in \mathbf{B}_{\beta_\mathbf{b}}$*.*

(3) Each ballot $\mathbf{b} \in \mathbf{B}$ *is* labeled *with an indelible* identifier $\mathrm{id}_\mathbf{b} \in \{0, 1\}^*$*. Ballot identifiers are not necessarily unique; if the labels are unique, we say that the family is* uniquely labeled*.*

Some RLAs use the "location" of the ballot as the identifier (e.g., $\mathrm{id}_\mathbf{b}$ = *413th ballot in batch 6*); our framework works perfectly well in this setting. To reflect practical settings where certain ballots are actually unlabeled, these can be assigned a distinguished "unlabeled" identifier in $\{0, 1\}^*$.

Definition 2 (Tabulation; election)**.** *Let* $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_k)$ *be a ballot family. A* tabulation $T = (T_1, \ldots, T_k)$ *for* $\mathbf{B}$ *is a sequence where each* T_β *is a triple* $T_\beta = (\mathsf{S}_\beta^{\mathrm{tab}}; \mathsf{W}_\beta^{\mathrm{tab}}, \mathsf{L}_\beta^{\mathrm{tab}})$ *of natural numbers.* $\mathsf{S}_\beta^{\mathrm{tab}}$ *is the number of ballots declared by the tabulation in batch* β*,* $\mathsf{W}_\beta^{\mathrm{tab}}$ *is the number of votes for the declared winner, and* $\mathsf{L}_\beta^{\mathrm{tab}}$ *is the number of votes for the declared loser. For a tabulation T, the* tabulated totals *are*

$$\mathsf{W}^{\mathrm{tab}} = \sum_\beta \mathsf{W}_\beta^{\mathrm{tab}} \qquad \text{and} \qquad \mathsf{L}^{\mathrm{tab}} = \sum_\beta \mathsf{L}_\beta^{\mathrm{tab}}$$

with the convention that $\mathsf{W}^{\mathrm{tab}} > \mathsf{L}^{\mathrm{tab}}$*.*

An election E *is a pair* $E = (\mathbf{B}, T)$ *where* $\mathbf{B}$ *is a ballot family and* $T = (T_1, \ldots, T_k)$ *is a* tabulation *for* $\mathbf{B}$.

We do not treat elections that declare a tie between W and L, with the assumption that this would result in a runoff or a full hand-count audit.

Notational warning. The candidate W is the *declared winner* of the election (according to the tabulation). The tabulation may not, of course, accurately reflect the votes recorded on the ballots. The primary circumstance of interest arises when W is not the true winner of the election.

Definition 3 (Actual vote totals; ballot manifests)**.** *Let $E = (\mathbf{B}, T)$ be an election. Let*

$$((\mathsf{S}_1^{\mathrm{act}}; \mathsf{W}_1^{\mathrm{act}}, \mathsf{L}_1^{\mathrm{act}}), \ldots, (\mathsf{S}_k^{\mathrm{act}}; \mathsf{W}_k^{\mathrm{act}}, \mathsf{L}_k^{\mathrm{act}}))$$

denote the actual totals, where $\mathsf{S}_\beta^{\mathrm{act}} = |\mathbf{B}_\beta|$ is the actual size of batch β and

$$\mathsf{W}_\beta^{\mathrm{act}} = \sum_{\mathbf{b} \in \mathbf{B}_\beta} \mathsf{W}_{\mathbf{b}} \qquad \textit{and} \qquad \mathsf{L}_\beta^{\mathrm{act}} = \sum_{\mathbf{b} \in \mathbf{B}_\beta} \mathsf{L}_{\mathbf{b}}$$

are the total number of actual votes received by candidate W and candidate L in batch β. The actual totals are

$$\mathsf{W}^{\mathrm{act}} = \sum_\beta \mathsf{W}_\beta^{\mathrm{act}} \qquad \textit{and} \qquad \mathsf{L}^{\mathrm{act}} = \sum_\beta \mathsf{L}_\beta^{\mathrm{act}} .$$

The ballot manifest *of E is the tuple $\mathsf{S}_E^{\mathrm{act}} = (\mathsf{S}_1^{\mathrm{act}}, \ldots, \mathsf{S}_k^{\mathrm{act}})$.*

Definition 4 (Diluted margin; valid and invalid elections)**.** *The* tabulated diluted margin *of an election E is the quantity*

$$\mu^{\mathrm{tab}} = \frac{\mathsf{W}^{\mathrm{tab}} - \mathsf{L}^{\mathrm{tab}}}{|\mathbf{B}|} .$$

An election E is invalid *if the tabulated winner is incorrect: $\mathsf{L}^{\mathrm{act}} \geq \mathsf{W}^{\mathrm{act}}$; otherwise, we say that E is* valid.

The tabulated diluted margin is determined by both the number of physical ballots (as determined by the ballot manifest) and the tabulation; to emphasize this, we use the notation μ^{tab}. This is in contrast to the actual diluted margin *$\mu^{\mathrm{act}} = |\mathsf{W}^{\mathrm{act}} - \mathsf{L}^{\mathrm{act}}|/|\mathbf{B}|$ which is determined only by the physical ballots.*

Cast-vote records (CVRs). A cast-vote record table (CVR) is an (untrusted) declaration of both the ballots appearing in a particular physical batch and the votes appearing on the ballots. Each row of the CVR contains a ballot identifier and two entries in $\{0, 1\}$ indicating whether the purported ballot contains a vote for W or L.

In our setting, it is critical that tabulations provide batch-level subtotals which can be compared against the totals declared by adaptively generated CVR tables. Traditional RLAs require only a "global" CVR and the global consistency check that it induces the same winners and losers as the tabulation.

Ident.	W	L
id_1	1	0
$\bot_1$	1	0
id_3	0	1
$\vdots$	$\vdots$	$\vdots$

Figure 1: A CVR.

Definition 5 (Cast-Vote Record Table (CVR))**.** *Let $\mathbf{B}$ be a ballot family. A* Cast-Vote Record Table (CVR) *for batch β is a sequence of triples*

$$\mathrm{cvr} = ((\iota_1, \mathsf{W}_1, \mathsf{L}_1), \ldots, (\iota_s, \mathsf{W}_s, \mathsf{L}_s))$$

where each ι_r is a bitstring in $\{0, 1\}^*$ and each $\mathsf{W}_r, \mathsf{L}_r$ is an element of $\{0, 1\}$. We use the following language:

(1) The elements ι_r are identifiers.

(2) The number s is the size *of the CVR.*

(3) The rth row is a triple $\mathrm{cvr}_r = (\iota_r, \mathsf{W}_r, \mathsf{L}_r)$.

(4) *The values*

$$\mathsf{S}^{\mathrm{cvr}}_\beta = s,\ \mathsf{W}^{\mathrm{cvr}}_\beta = \sum_{1\le r\le s} \mathsf{W}^{\mathrm{cvr}}_r, \qquad \textit{and} \qquad \mathsf{L}^{\mathrm{cvr}}_\beta = \sum_{1\le r\le s} \mathsf{L}^{\mathrm{cvr}}_r .$$

These denote the number of ballots declared by the CVR and the number of votes declared for the two candidates in the CVR.

(5) *If the identifiers appearing in the CVR are unique, we say the CVR is* uniquely labeled. *If a CVR is uniquely labeled we use r_ι to refer to the (unique) row with identifier ι. Looking ahead, in Figure 4 we use the identifiers $\bot_i$ to transform a CVR to one with unique labels; such labels would not appear on CVRs generated by tabulators.*

Finally, a sequence $(\mathrm{cvr}_1, \dots \mathrm{cvr}_k)$*, where each* cvr_β *is a CVR for batch* β*, is a* global CVR.

Discrepancy. Discrepancy measures the disagreement between claimed vote tallies, either from a tabulation or CVR, and vote tallies determined by actual ballots.

Definition 6 (Batch and election discrepancy)**.** *Let $E = (\mathbf{B}, T)$ be an election. The* discrepancy *of a batch* $\mathbf{B}_\beta$ *is*

$$\mathsf{D}_\beta = (\mathsf{W}^{\mathrm{tab}}_\beta - \mathsf{L}^{\mathrm{tab}}_\beta) - \sum_{\mathbf{b}\in\mathbf{B}_\beta} \left(\mathsf{W}^{\mathrm{act}}_{\mathbf{b}} - \mathsf{L}^{\mathrm{act}}_{\mathbf{b}}\right) .$$

The overall discrepancy of an election is

$$\mathsf{D} = \sum_\beta \mathsf{D}_\beta = (\mathsf{W}^{\mathrm{tab}} - \mathsf{L}^{\mathrm{tab}}) - (\mathsf{W}^{\mathrm{act}} - \mathsf{L}^{\mathrm{act}}) .$$

For invalid elections $\mathsf{L}^{\mathrm{act}} \ge \mathsf{W}^{\mathrm{act}}$ and thus $\mu^{\mathrm{act}} = -(\mathsf{W}^{\mathrm{act}} - \mathsf{L}^{\mathrm{act}})/|\mathbf{B}|$. In this case

$$\frac{\mathsf{D}}{|\mathbf{B}|} = \frac{(\mathsf{W}^{\mathrm{tab}} - \mathsf{L}^{\mathrm{tab}}) - (\mathsf{W}^{\mathrm{act}} - \mathsf{L}^{\mathrm{act}})}{|\mathbf{B}|} = \mu^{\mathrm{tab}} + \mu^{\mathrm{act}}. \tag{1}$$

The discrepancy of a CVR is undefined until it is generated, which is why the above "global" definitions focus on the tabulation.

Definition 7 (CVR Discrepancy)**.** *Let* $\mathbf{B}$ *be a ballot family and let* $\mathrm{cvr} = ((\iota_1, \mathsf{W}_1, \mathsf{L}_1), \dots, (\iota_s, \mathsf{W}_s, \mathsf{L}_s))$ *be a CVR for batch β. For a row $r \in [s]$, define the* discrepancy $\mathsf{D}^{\mathrm{cvr}}_r$ *of the row r to be the value*

$$\mathsf{W}_r - \mathsf{L}_r + \min\left(\{1\} \cup \left\{-(\mathsf{W}_{\mathbf{b}} - \mathsf{L}_{\mathbf{b}}) \mid \mathrm{id}_{\mathbf{b}} = \iota_r, \mathbf{b} \in \mathbf{B}_\beta\right\}\right) . \tag{2}$$

The minimum is taken over all ballots for which $\mathrm{id}_{\mathbf{b}} = \iota_r$ *with the default value of* 1 *(intuitively corresponding to a "concealed vote" for the declared loser) when no ballot corresponds to the identifier.*

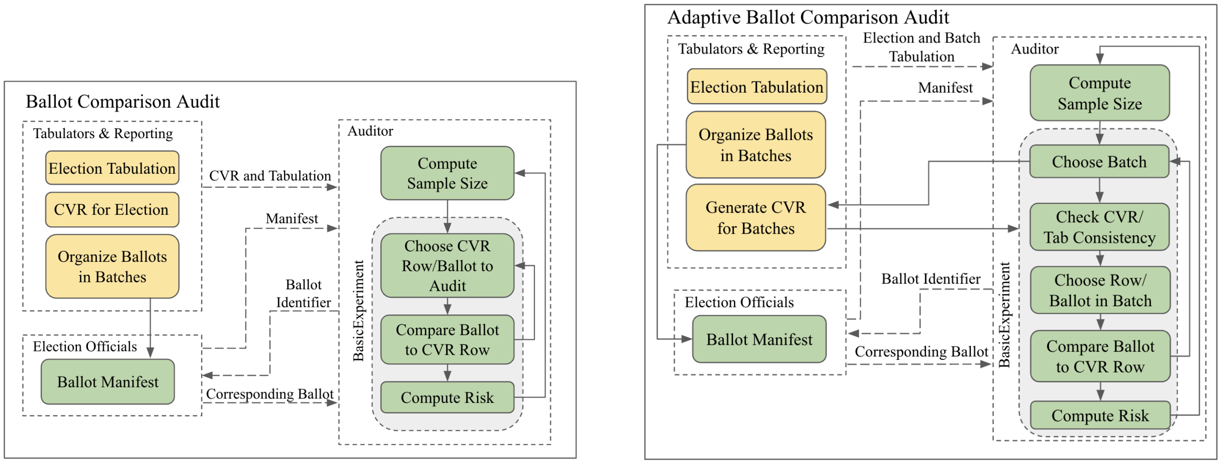
When discrepancy takes a positive value d we refer to it as a *d-vote overstatement*; likewise, when it takes a negative value $-d$ we refer to it as a *d-vote understatement*. In the context of a tabulation, then, a d-vote overstatement indicates that the reported difference, $\mathsf{W}^{\mathrm{tab}} - \mathsf{L}^{\mathrm{tab}}$, is d votes too large. Equation (2) assigns a notion of discrepancy to a particular row of a CVR, which always takes a value in the set $\Sigma = \{-2, -1, 0, 1, 2\}$. In the case when an identifier ι corresponds to a unique ballot $\mathbf{b}$, the discrepancy is the natural difference

$$\mathsf{W}_{r_\iota} - \mathsf{L}_{r_\iota} - (\mathsf{W}_{\mathbf{b}} - \mathsf{L}_{\mathbf{b}}) .$$

3 The Adaptive Auditor

A traditional ballot comparison audit proceeds as follows (illustrated in Figure 2a):

(1) An election is carried out, electronic tabulators generate an *untrusted* tabulation.



(a) Architecture of Ballot Comparison Audit with Rounds

(b) Architecture of Adaptive Ballot Comparison Audit with Rounds

Figure 2: Comparison of traditional and adaptive ballot comparison architectures. Yellow components are performed by untrusted components. Green components must be trustworthy. The dotted arrows represent information trade, while the solid arrows are procedure steps. The grey procedure is `BasicExperiment` can be done in parallel in both traditional and adaptive RLAs. Note that in a traditional audit, the CVR is generated as part of the audit process; in the adaptive setting, the CVR is generated only as the auditor chooses batches. The step of checking CVR and tabulation consistency is also absent from traditional comparison audits as an audit of the CVR is an audit of the tabulation as long as they show the same set of winning/losing candidates.

(2) Election officials store the physical ballots as a ballot family and produce a *trusted* ballot manifest that correctly indicates the number of physical ballots in the batch.

(3) An *untrusted* CVR is generated.

(4) The audit repeatedly selects a CVR row and ensures that the corresponding physical ballot matches the declaration of votes on the CVR row.

The audit either generates a risk-controlled declaration that the tabulated outcomes are consistent with the ballots or an inconclusive result.

The adaptive alternative. As described in the introduction we consider the *adaptive* version of the above (shown in Figure 2b) where CVRs are only generated when needed. This yields the following family of auditing procedures.

(1) An election is carried out and ballot family created as in steps (1)-(2) above. The tabulation declares a (sub-) tabulation for each batch in the ballot family.

(2) The audit consists of multiple instances of the following **basic experiment**, which may be carried out in parallel:

   (a) A batch is sampled with probability proportional to the number of ballots.

   (b) An (untrusted) CVR is generated for the batch.

   (c) The CVR is compared against the declared subtotals.

   (d) An entry in the CVR is drawn uniformly and compared with the corresponding ballot.

As above, the conclusion is either "consistent" or "inconclusive."

Multiple iterations of the basic experiment can be performed in parallel as in a traditional ballot comparison audit to allow audit workers to create their portion of the CVR simultaneously. These are known as audit *rounds* which yield a trade-off between the total number of examined ballots and the probability of carrying out an additional round of auditing. The impact of conducting multiple rounds can be quite high, so parameters are typically chosen to ensure a single-round audit with high probability. All of this existing machinery applies identically in our setting.

This section focuses on the audit procedure. However, a few preliminary remarks about modeling are in order. The risk guarantee associated with a standard comparison audit must hold for all possible CVRs that could be submitted for the election, even those that might be specifically designed to frustrate the audit or obscure an invalid election. This motivates our treatment of the environment in which an auditor operates as *adversarial*, including the CVRs that are produced. We additionally assume an arbitrary labeling of ballots.

The informal treatment above already highlights an important difference between conventional comparison audits and adaptive audits: the CVR generated and used for comparison by the auditor in steps (2)b–(2)d may depend on the prior history of the audit. The need to bound risk must hold when the CVRs proposed at intermediate steps of the audit might depend adversarially on prior CVRs, row selections, and comparison results. This ability of an adversary intent on concealing an invalid election appears to be very powerful: for example, if an adversary has been "caught" in a comparison iteration they may choose to declare subsequent CVRs with a low discrepancy in order to convince the statistical test that "everything is OK." The above procedure appears to be the first RLA involving an adaptive adversary that engages with the auditor.

We begin by introducing a "strict" auditor that enforces size checks, insisting that the CVR is consistent with tabulation. This auditor is not necessarily useful in practice, but is a convenient analytic tool. We then generalize this auditor by defining the notion of a *CVR transform function* that is applied before the auditor checks consistency. This extra flexibility makes it easy to construct and reason about more permissive auditors that are useful in practice. As we show in Lemma 1, if the original strict auditor (with the identity CVR transform) is risk-limiting then the resulting auditor is risk-limiting for *every* CVR transform. This allows us to introduce a transform that always produces "consistent" CVRs.

In the next three subsections, we discuss single-tailed statistical tests, the auditor, and the intuition for included checks. We then present the formal game including the definition of risk limit in Section 4, show that the auditor is risk-limiting for an appropriate statistical test in Section 5, and discuss completeness in Section 6.

3.1 Adaptive single-tailed statistical tests

A standard approach for designing RLAs is to consider the discrepancy $\mathsf{D}_r^{\mathrm{cvr}} = (\mathsf{W}_r - \mathsf{L}_r) - (\mathsf{W}_{\mathbf{b}} - \mathsf{L}_{\mathbf{b}})$ of a uniformly selected row r of a global CVR in comparison with a ballot $\mathbf{b}$ corresponding to this entry (as in Definition 7). In light of Equation 1, if the election is invalid one has that

$$\mathop{\mathbb{E}}_{r}[\mathsf{D}_r^{\mathrm{cvr}}] \geq \mu^{\mathrm{tab}} + \mu^{\mathrm{act}} \geq \mu^{\mathrm{tab}} \,.$$

Independently repeating this experiment results in a sequence of discrepancy observations $\mathsf{D}_1, \mathsf{D}_2, \ldots$ taking values in $\{-2, -1, 0, 1, 2\}$. With these random variables, one can formulate an RLA as a conventional statistical hypothesis test by adopting the null hypothesis that the election is invalid; then one is interested in bounding the probability that the null hypothesis is rejected when it is true. An RLA is determined by a single-tailed statistical test for these i.i.d. random variables with the hypothesis that "$\mathbb{E}[\mathsf{D}_i] \geq \mu^{\mathrm{tab}}$." The test decides whether to reject this hypothesis based on examination of a finite-length prefix $\mathsf{D}_1, \ldots, \mathsf{D}_\tau$ of the variables given by a "stopping time." Informally, such a test has *risk* (Type I error) α if $\alpha \geq \Pr[\text{hypothesis rejected}]$ when indeed $\mathbb{E}[\mathsf{D}_i] \geq \mu^{\mathrm{tab}}$. See [33, Equation 5] for further discussion.

The adaptive setting and the domination inequalities. In our setting with an adaptive adversary, we will require statistical tests with stronger properties. Specifically, as above we consider an infinite family of random variables $X_1, X_2, \ldots$ taking values in Σ with the weaker *domination* conditions recorded below.

Definition 8 (δ-dominating distributions and random variables)**.** *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 also use this terminology to apply to the distribution $\mathcal{D}$ corresponding to the random variables, writing $\delta \trianglelefteq \mathcal{D}$.

The variables are no longer required to be independent or have the same distribution; however, they still possess the property that under any conditioning on the past, each random variable has expectation bounded below by δ.

Definition 9 (Stopping time)**.** *Let $\Sigma = \{-2, -1, 0, 1, 2\}$. 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 $\mathsf{Stop}() = 1$.

With these preliminaries noted, we can define the family of statistical tests that we show can support adaptive audits.

Definition 10 (Adaptive Audit Test)**.** *An* adaptive audit test*, denoted $T = (\mathsf{Stop}, \mathsf{R})$, 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. Then we define the risk *of the test to be*

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

where this supremum is taken over all $\delta \in (0, 2]$ and over all probability distributions $\mathcal{D}$ for which $\delta \trianglelefteq \mathcal{D}$.

In Section 5 we observe that several families of statistical tests in common use—including the popular Kaplan-Markov test—are, in fact, adaptive audit tests.

3.2 The Adaptive Audit Procedure

We now present the adaptive auditor (Figure 3). The design of the audit procedure is motivated by three guiding principles:

(1) Ensure tabulation consistency with the ballot manifest. (This means the size must match, $\mathsf{W}^{\text{tab}} \leq \mathsf{S}^{\text{act}}$, and $\mathsf{L}^{\text{tab}} \leq \mathsf{S}^{\text{act}}$). Such checks ensure that the overall discrepancy is at least the margin for invalid elections. This principle motivates Steps (2) and (3).

(2) Ensure that duplicate labels appearing on distinct ballots cannot increase risk. This follows from (i.) forcing CVR tables to contain no duplicates, (ii.) adopting uniform selection of CVR rows for ballot selection and, (iii.) noting that among the collection of ballots that may be assigned a common identifier, there is a "pessimal" ballot that induces the minimum discrepancy. See `CheckConsistent` and Step (7) of `BasicExperiment`.

(3) Ensure that any produced CVR for a batch has the same number of votes for the winner and loser as the declared tabulation for that batch. This yields a lower bound on the discrepancy—determined only by the tabulation and the ballots—between any such CVR and the ballots. See the additional checks in `CheckConsistent`.

This auditor and the related treatment of ballot identifier uniqueness also have direct ramifications for traditional comparison audits; see the discussion in Section 4.1 below.

Figure 3 distinguishes two important algorithmic elements of the auditor by giving them separate "modular" treatment: the statistical test and the CVR transform.

(1) The statistical test. The auditor requires an adaptive audit test $(\mathsf{Stop}, \mathsf{R})$ as defined in Definition 10.

(2) The CVR transform. The auditor requires a CVR transform $\mathcal{T}$, which is a rule for rewriting a CVR before comparison.

Thus a full description of the auditor is written $C[\mathcal{T}; (\mathsf{Stop}, \mathsf{R})]$. In situations where the transform or the test are not directly relevant or can be inferred from context, we simply write C.

Auditor $\mathcal{C}[\mathcal{T}; (\mathsf{Stop}, \mathsf{R})]$ for an election E

(1) Receive ballot manifest and tabulation:

$$\mathsf{S}_E^{\text{act}} = (\mathsf{S}_1^{\text{act}}, \ldots, \mathsf{S}_k^{\text{act}}); \qquad T = (\mathsf{S}_1^{\text{tab}}; \mathsf{W}_1^{\text{tab}}, \mathsf{L}_1^{\text{tab}}), \ldots, (\mathsf{S}_k^{\text{tab}}; \mathsf{W}_k^{\text{tab}}, \mathsf{L}_k^{\text{tab}}))$$

(2) For $\beta = 1$ to k:
   (a) $\mathsf{S}_\beta^{\text{tab}} := \mathsf{S}_\beta^{\text{act}}$;
   (b) $\mathsf{W}_\beta^{\text{tab}} := \min(\mathsf{W}_\beta^{\text{tab}}, \mathsf{S}_\beta^{\text{act}})$;
   (c) $\mathsf{L}_\beta^{\text{tab}} := \min(\mathsf{L}_\beta^{\text{tab}}, \mathsf{S}_\beta^{\text{act}})$.

(3) Let $S^{\text{act}}, S^{\text{tab}} := \sum_{\beta=1}^{k} \mathsf{S}_\beta^{\text{tab}} = \sum_{\beta=1}^{k} \mathsf{S}_\beta^{\text{act}}$ and

$$\mu := \frac{\sum_{\beta=1}^{k} (\mathsf{W}_\beta^{\text{tab}} - \mathsf{L}_\beta^{\text{tab}})}{\mathsf{S}^{\text{act}}}.$$

(4) If $\mu \le 0$ return `Inconclusive`.
(5) Initialize $\mathsf{iter} = 0$.
(6) Repeat
   (a) Increment $\mathsf{iter} := \mathsf{iter} + 1$.
   (b) Perform $\mathsf{D}_{\mathsf{iter}} :=$ `BasicExperiment`
   until $\mathsf{Stop}_\mu(\mathsf{D}_1, \ldots, \mathsf{D}_{\mathsf{iter}}) = 1$
(7) If $\mathsf{R}_\mu(\mathsf{D}_1, \ldots, \mathsf{D}_{\mathsf{iter}}) = 1$ return `Consistent`; otherwise return `Inconclusive`.

`BasicExperiment`:

(1) Select batch β with probability $\mathsf{S}_\beta^{\text{tab}} / \mathsf{S}^{\text{tab}}$.
(2) Request CVR for batch β. Denote the response cvr_β.
(3) Apply $\mathcal{T}$: $\mathrm{cvr}_\beta := \mathcal{T}(\mathsf{S}_E^{\text{act}}, T, \mathrm{cvr}_\beta)$.
(4) `RowSelect`: Select a row $r \in [\mathsf{S}_\beta^{\text{tab}}]$ uniformly.
(5) If `CheckConsistent`$(\mathsf{S}_E^{\text{act}}, T, \mathrm{cvr}_\beta) =$ `Error`, return 2.
(6) Let ι be the ballot identifier in row r; request delivery of ballot ι from batch β.
(7) If a ballot **b** is delivered from batch β with identifier ι, let $\mathsf{W}^{\text{act}}, \mathsf{L}^{\text{act}}$ denote the $\{0, 1\}$ values on **b** for the declared winner and loser respectively. Otherwise, set $\mathsf{W}^{\text{act}} := 0, \mathsf{L}^{\text{act}} := 1$.
(8) Return $(\mathsf{W}_r^{\text{cvr}} - \mathsf{L}_r^{\text{cvr}}) - (\mathsf{W}^{\text{act}} - \mathsf{L}^{\text{act}})$.

`CheckConsistent`$(\mathsf{S}_E^{\text{act}}, T, \mathrm{cvr}_\beta)$:

(1) If cvr_β is not uniquely-labeled (Def. 5) return `Error`.
(2) If $\mathsf{S}_\beta^{\text{cvr}} \ne \mathsf{S}_\beta^{\text{act}}$ or $\mathsf{S}_\beta^{\text{act}} \ne \mathsf{S}_\beta^{\text{tab}}$, return `Error`.
(3) If $\mathsf{W}^{\text{cvr}} \ne \mathsf{W}^{\text{tab}}$ or $\mathsf{L}^{\text{cvr}} \ne \mathsf{L}^{\text{tab}}$, return `Error`.
(4) Return `OK`.

Figure 3: The auditor $\mathcal{C}[\mathcal{T}; (\mathsf{Stop}, \mathsf{R})]$. Here $\mathcal{T}$ is a CVR transform and $(\mathsf{Stop}, \mathsf{R})$ is an adaptive audit test.

Remarks on the auditor's handling of the CVR. As a convenience, our treatment permits the Auditor to carry out bookkeeping using the CVR, such as adding new rows or relabeling certain rows with new identifiers that are known not to match a physical ballot. For this purpose, we treat $\perp_1, \perp_2, \ldots$ as a sequence of special purpose identifiers known not to match any ballot. These modifications are for internal bookkeeping of the auditor only; the original CVR is still considered an immutable artifact of the audit.

The CVR separately records, for a given row r, whether it is associated with a vote for W or a vote for L; this

$\underline{\mathcal{T}_{\mathrm{Id}}(\mathsf{S}_E^{\mathrm{act}}, T, \mathrm{cvr}_\beta)}$:

(1) Return cvr_β.

$\underline{\mathcal{T}_{\mathrm{Force}}(\mathsf{S}_E^{\mathrm{act}}, T, \mathrm{cvr}_\beta)}$:

(1) While there exist two rows i and j where $i < j$ and both have identifier ι, replace the identifier in row j with an unused identifier in $\{\perp_t\}$.

(2) If $\mathsf{S}_\beta^{\mathrm{cvr}} \neq \mathsf{S}_\beta^{\mathrm{act}}$, then

  (a) While $\mathsf{S}_\beta^{\mathrm{cvr}} < \mathsf{S}_\beta^{\mathrm{act}}$ add a new row to cvr_β with an unused identifier in $\{\perp_t\}$ and zeroes for all votes.

  (b) While $\mathsf{S}_\beta^{\mathrm{cvr}} > \mathsf{S}_\beta^{\mathrm{act}}$ remove the last row of cvr_β.

(3) If $\mathsf{W}_\beta^{\mathrm{cvr}} \neq \mathsf{W}_\beta^{\mathrm{tab}}$. Set $i := \mathsf{S}_\beta^{\mathrm{cvr}}$.

  (a) While $\mathsf{W}_\beta^{\mathrm{cvr}} < \mathsf{W}_\beta^{\mathrm{tab}}$ set $\mathsf{W}_i^{\mathrm{cvr}} = 1$; set $i := i - 1$.

  (b) While $\mathsf{W}_\beta^{\mathrm{cvr}} > \mathsf{W}_\beta^{\mathrm{tab}}$ set $\mathsf{W}_i^{\mathrm{cvr}} = 0$, set $i := i - 1$.

(4) If $\mathsf{L}_\beta^{\mathrm{cvr}} \neq \mathsf{L}_\beta^{\mathrm{tab}}$. Set $i := \mathsf{S}_\beta^{\mathrm{cvr}}$.

  (a) While $\mathsf{L}_\beta^{\mathrm{cvr}} > \mathsf{L}_\beta^{\mathrm{tab}}$ set $\mathsf{L}_i^{\mathrm{cvr}} = 1$, set $i := i - 1$.

  (b) While $\mathsf{L}_\beta^{\mathrm{cvr}} < \mathsf{L}_\beta^{\mathrm{tab}}$ set $\mathsf{L}_i^{\mathrm{cvr}} = 0$, set $i := i - 1$.

(5) Return cvr_β.

Figure 4: CVR transform functions.

convention permits, in principle, rows of the CVR to contain votes for *both* candidates, known as an overvote, (a row with $1\ 1$ in the CVR table). This does not interfere with the risk limit of the auditor (even when used for an election that forbids overvotes) and is convenient for the Force transform. We point out in Appendix B that this is unnecessary, presenting a more complicated auditor that does not allow overvotes and a more complicated CVR transform function that never creates overvotes.

3.2.1 The CVR transform

The auditor also takes as input a CVR rewriting procedure, denoted $\mathcal{T}$, that will be used to "correct" the CVR before deciding if it is consistent with the tabulation. Our proof that the auditor is risk-limiting adopts the "identity" $\mathcal{T}$ that does not rewrite the CVR. In Lemma 1, we then show that if $C[\mathcal{T}_{\mathrm{Id}}; (\mathsf{Stop}, \mathsf{R})]$ is risk-limiting for the identity transform then it is risk-limiting for *any* procedure $\mathcal{T}'$. The goal of $\mathcal{T}_{\mathrm{Force}}$ is to make the CVR consistent with tabulation with minimal edits. We use $\mathcal{T}_{\mathrm{Force}}$ in all of our completeness analyses.

3.3 Discussion; an intuitive survey of the adaptive auditor

We prove the soundness of the auditor in Section 5; this informal discussion is for the sake of intuition.

The `CheckConsistent` procedure returns an error (resulting in a discrepancy of 2) in many settings that could occur naturally in practice, such as a mismatch between the number of ballots counted on the tabulator and the number of ballots on the CVR. Here we discuss the role played by the various properties checked by `CheckConsistent`. We remark again that a much more permissive auditor is obtained by the Force transform, discussed later.

Uniquely labeled CVRs. In our model and in many practical settings the auditor cannot ensure that ballots are uniquely labeled. This explains the convention that defines discrepancy for a row r as the minimum discrepancy across all ballots with the row identifier ι_r. The auditor does, however, ensure the uniqueness of identifiers appearing in the CVR. A concrete attack exists in the absence of this check. One simply labels all ballots with the same identifier and crafts a CVR to be consistent with tabulation. Then when a ballot is requested one simply returns a ballot with the votes listed in the CVR row. This attack succeeds as long as all vote patterns exist on at least one ballot. This is why a crucial step in $\mathcal{T}_{\text{Force}}$ in Figure 4 is to rewrite duplicate identifiers on a CVR.

Treatment of missing ballots. Missing ballots are treated as though cast for the loser. If not, the adversary can always choose to not return those ballots that show votes for the loser, effectively reducing the observed discrepancy. This treatment is similar to the "phantoms to zombies" approach [2].

Enforcing equality of batch sizes. The size checks ensure that the `RowSelect` operation selects both a uniform row in the CVR and (for an honest adversary) a uniform ballot in the batch.

Enforcing equality of CVR and tabulation subtotals. As discussed below, the tabulation effectively determines a lower bound on total discrepancy for the batch regardless of adversarial choice of the CVR. Without the check $\mathsf{W}^{\text{cvr}} = \mathsf{W}^{\text{tab}}$ and $\mathsf{L}^{\text{cvr}} = \mathsf{L}^{\text{tab}}$, the CVR could always be consistent with the ballots without actually auditing the tabulation.

4 An Adversarial Auditing model

As discussed in Section 3.1, the conventional formal approach to RLAs adopts the language of Neyman–Pearson statistical hypothesis testing. This picture emphasizes the role played by the culminating statistical test. Our more complex setting—involving adaptive selection of CVRs that may depend on the entire history of the audit—motivates us to extend the formal treatment of the audit to the entire procedure. We adopt the "security game" framework from the theory of cryptography, which has the expressive power to reflect such interactions between parties. The cryptographic model has the advantage that it explicitly identifies an *adversary*, a party that is charged with frustrating or subverting the audit, and precisely defines which aspects of the audit are under adversarial control.

In our framework, the adversary is responsible for producing CVRs and providing ballots to the auditor when requested; ballot labels are also effectively under adversarial control, as the final conclusions are guaranteed for all such labelings. The resulting game is a "physical cryptography game" along the lines of Fisch, Freund, and Naor [8]. In general, our definition gives the adversary control over parts of the process whenever possible. This explicitly identifies what aspects of the procedure must be honestly conducted for the statistical guarantees to hold. Finally, we remark that we adopt the classical nomenclature of "soundness" and "completeness" for cryptographic games that act as the analogues of Type I and Type II errors.

The Auditor–Adversary Game. The *Auditor–Adversary* game is played by two parties, the *Auditor* denoted by $\mathcal{C}$ and the *Adversary* denoted by $\mathcal{A}$. The game is played in the context of an election (Definition 2) and involves the exchange of both physical objects (ballots) and information (CVRs). Recall that we use boldface to refer to physical objects which may be exchanged between the formal parties in the game.

Figure 5 describes in detail the adaptive RLA game between the auditor and adversary. Before discussing the desired risk and completeness properties, we discuss our ballot identification convention.

Ballot identification. Our definition of a ballot family (Definition 1) includes identifiers on ballots. Recall that ballot identifiers are not assumed to be unique, which reflects an important feature of practical RLAs: in general, it's not possible for auditors to efficiently check physical identifiers to ensure that there are no collisions.

Our results work perfectly well if the adversary is permitted to (re-)assign identifiers to a batch each time they are asked to generate a CVR for that batch (this may be the case if a tabulator imprints during the audit). There are two crucial assumptions required for security in this setting: (1) the adversary cannot change ballot identifiers unless another CVR is requested for the batch, and (2) the auditor—if ever given the chance to observe the ballot—can reliably determine $\mathrm{id}_{\mathbf{b}}$.

Auditor (C)–Adversary ($\mathcal{A}$) game for election $E = (\mathbf{B}, T)$

(1) **Setup**.

  (a) **Ballot and tabulation delivery (to $\mathcal{A}$).** The physical ballots $\mathbf{B}$ and the tabulation T are given to the adversary $\mathcal{A}$.

  (b) **Ballot manifest and tabulation delivery (to C).** The ballot manifest $\mathsf{S}_E = (\mathsf{S}_1^{\text{act}}, \ldots, \mathsf{S}_k^{\text{act}})$ and the tabulation T are given to the auditor C.

(2) **Audit**. C repeatedly makes one of the following two requests of $\mathcal{A}$, or chooses to conclude the audit:

- **A CVR request**. C requests a CVR for batch β. $\mathcal{A}$ responds with a CVR denoted $\mathtt{CVR}_\beta$.
- **A ballot request**. C requests a ballot from the adversary with a specific identifier $\iota \in \{0,1\}^*$ from some batch β.
  (a) $\mathcal{A}$ either sends a physical ballot $\mathbf{b}$ in batch β, i.e. $\mathbf{b} \in \mathbf{B}_\beta$, to C or responds with **No ballot.**

(3) **Conclusion**. C returns one of the two values:

Consistent meaning "Audit consistent with tabulation," or
Inconclusive meaning "Audit inconclusive."

Figure 5: The $\mathsf{RLA}_{C,\mathcal{A}}(E)$ auditing game.

An adversary can effectively "destroy" a ballot by choosing not to reveal it when requested.

Definition 11 (Risk; soundness)**.** *Let C be an Auditor. For election E and adversary $\mathcal{A}$ let $\mathsf{RLA}_{C,\mathcal{A}}(E)$ denote the random variable equal to the conclusion of the audit as described in Figure 5. An auditor C has* α-risk *(or* α-soundness*) if, for all invalid elections E and all adversaries $\mathcal{A}$,*

$$\Pr[\mathsf{RLA}_{C,\mathcal{A}}(E) = \textsc{Consistent}] \leq \alpha\,.$$

(The probability here is taken over random choices of the auditor and the adversary.)

Definition 11 is a property of a C (the auditor) only. That is, it holds for all invalid elections and behaviors of the adversary. As we discuss in Section 6 completeness or Type-II errors will only be guaranteed for certain $\mathcal{A}$.

4.1 Modeling conventional RLAs

This modeling can apply directly to conventional ballot-comparison audits. In particular, by restricting the class of adversaries to those that draw all batch CVRs from a fixed global CVR, one obtains a model that corresponds to a conventional comparison audit. In particular, as this is a smaller class of adversaries, all of the conclusions of the paper apply to this setting (including the conclusions for the specific auditor we consider). This auditor can provide privacy improvements over traditional auditors, as it only needs to release portions of the global CVR table. As an alternate modeling approach, one can formulate an auditor that initially requests the entire CVR; with this convention, one can return to universally quantifying over all adversaries. The risk limits for this auditor follow directly from our proofs. Finally, we mention that these techniques demonstrate that traditional RLAs do not require the uniqueness of physical ballot identifiers.

The model can also be adapted to reason about polling audits, where auditors never issue CVR requests and tacitly assume a "position based" labeling. For simplicity, this variant calls for the adversary to label all ballots at the outset. These labels are never communicated to the auditor, who simply assumes that ballots are given labels of the form (b, n), where b is a batch number and n is a "sequence number" between 1 and the size of the batch. (Note that the auditor can deduce this label set from the ballot manifest.) Intuitively, this corresponds to the natural setting where ballots in each batch are placed in order and selection is determined by identifying a particular index in a particular batch. We remark that there are ballot polling techniques that are not directly reflected by this modeling: for example, techniques that treat "asking for a random ballot" as an atomic operation. (For example *k-cut* which cuts a stack of ballots an appropriate number of times [26].) Of course, with further alterations to the model, this could also be treated as a (necessarily) trusted operation.

5 $C[\mathcal{T}_{\mathrm{Id}}; (\mathsf{Stop}, \mathsf{R})]$ is Risk-Limiting

The key for establishing that C is risk-limiting is to demand that the generated CVR is nearly consistent with the previously generated tabulation. We observe that with this assurance, the tabulated results effectively generate a forcing "commitment" on the discrepancy of any CVR that the adversary may generate. Batch tabulations now play an essential role in the analysis by enforcing this commitment. In a conventional ballot comparison audit, the details of the tabulation itself can be ignored so long as the tabulation and CVR declare the same winner: The operational details of the audit are determined entirely by the CVR.

Theorem 1. *Let* $(\mathsf{Stop}, \mathsf{R})$ *be an adaptive audit test with risk* α*; let* $\mathcal{T}$ *be an arbitrary procedure that transforms CVRs to CVRs. Let* C *the auditor in Figure 3. Then* $C[\mathcal{T}; (\mathsf{Stop}, \mathsf{R})]$ *has risk* α.

Proof. We begin with the next Lemma, showing that a $\mathcal{T}$ does not affect whether an auditor is risk-limiting.

Lemma 1. *Let* $\mathcal{T}$ *be a (possibly randomized) procedure that takes as input* $(\mathsf{S}_E^{\mathrm{act}}, T, \mathrm{cvr}_\beta)$ *and rewrites* cvr_β*. Let* $(\mathsf{Stop}, \mathsf{R})$ *be a statistical test and let* C *be an auditor as in Figure 3.*

If $C[\mathcal{T}_{\mathrm{Id}}; (\mathsf{Stop}, \mathsf{R})]$ *satisfies Definition 11 with* α*-risk then* $C[\mathcal{T}; (\mathsf{Stop}, \mathsf{R})]$ *satisfies Definition 11 with* α*-risk.*

The proof of Lemma 1 has a simple core: For every adversary, $\mathcal{A}$ that succeeds in the presence of $\mathcal{T}$ one can define another adversary $\mathcal{A}'$ that applies $\mathcal{T}$ before returning the CVR to the auditor.

Proof. We show the result by the contrapositive. Fix some statistical test $(\mathsf{Stop}, \mathsf{R})$. Suppose that for some election E there exists an adversary $\mathcal{A}$ such that

$$\Pr_{C_{\mathcal{T},(\mathsf{Stop},\mathsf{R})}}[\mathsf{RLA}_{C_{\mathcal{T},(\mathsf{Stop},\mathsf{R})},\mathcal{A}}(E) = \textsc{Consistent}] > \alpha\,.$$

Consider $C[\mathcal{T}_{\mathrm{Id}}; (\mathsf{Stop}, \mathsf{R})]$. Assume for a moment that the test Stop always outputs 0. (This is just to define a sequence of length ℓ, noting that the selection of batches/ballots is independent in each iteration though the resulting discrepancies need not be independent).

Fix some positive number ℓ and consider a sequence of selected batches $\beta_1, ..., \beta_\ell$ and selected locations within a batch $\iota_1, ..., \iota_\beta$ with $\iota_\beta = \perp$ as a special value indicating that no ballot is selected. Here we that note both of these sequences of random variables are independent of an adversary and only depend on the election E. Furthermore, note that these sequences are identically distributed in $C[\mathcal{T}_{\mathrm{Id}}; (\mathsf{Stop}, \mathsf{R})]$ and $C[\mathcal{T}; (\mathsf{Stop}, \mathsf{R})]$ except that some locations may be $\perp$ in either sequence but not in the other. Consider the following adversary $\mathcal{A}'$ for the auditing experiment with $C[\mathcal{T}_{\mathrm{Id}}; (\mathsf{Stop}, \mathsf{R})]$.

- $\mathcal{A}'$ initializes $\mathcal{A}$ with E.
- $\mathcal{A}'$ runs $\mathcal{A}$ and forwards all audit requests to $\mathcal{A}$. Upon receiving a response cvr_β from $\mathcal{A}$, compute $\mathrm{cvr}'_\beta = \mathcal{T}(\mathsf{S}_E^{\mathrm{act}}, T, \mathrm{cvr}_\beta)$ and return cvr'_β to $C[\mathcal{T}_{\mathrm{Id}}; (\mathsf{Stop}, \mathsf{R})]$.
- Upon receiving request for ballot ι, forward request to $\mathcal{A}$ and return ballot returned by $\mathcal{A}$.

$\mathcal{A}'$ exactly replicates the view that $\mathcal{A}$ would experience interacting with $C[\mathcal{T}; (\mathsf{Stop}, \mathsf{R})]$. The sequence of batches and locations selected in $C[\mathcal{T}_{\mathrm{Id}}; (\mathsf{Stop}, \mathsf{R})]$ when interacting with $\mathcal{A}'$ is identically distributed to $C[\mathcal{T}; (\mathsf{Stop}, \mathsf{R})]$ when interacting with $\mathcal{A}$.

We define $\vec{\mathsf{D}}_{C[\mathcal{T};(\mathsf{Stop},\mathsf{R})],\mathcal{A}}$ as the sequence of discrepancies produced by $\mathcal{A}$ when interacting with $C[\mathcal{T}; (\mathsf{Stop}, \mathsf{R})]$. Similarly, define $\vec{\mathsf{D}}_{C[\mathcal{T}_{\mathrm{Id}};(\mathsf{Stop},\mathsf{R})],\mathcal{A}'}$ as the sequence of discrepancies produced by $\mathcal{A}'$ interacting $C[\mathcal{T}_{\mathrm{Id}}; (\mathsf{Stop}, \mathsf{R})]$. We now remove the assumption that Stop always outputs 0. Then, the two sequences $\vec{\mathsf{D}}_{C[\mathcal{T};(\mathsf{Stop},\mathsf{R})],\mathcal{A}}$ and $\vec{\mathsf{D}}_{C[\mathcal{T}_{\mathrm{Id}};(\mathsf{Stop},\mathsf{R})],\mathcal{A}'}$ are identically distributed. Thus, it must be the case that

$$\Pr_C[\mathsf{RLA}_{\vec{\mathsf{D}}_{C[\mathcal{T}_{\mathrm{Id}};(\mathsf{Stop},\mathsf{R})],\mathcal{A}'}}(E) = \textsc{Consistent}] > \alpha\,.$$

This is a contradiction and proves Lemma 1. □

We then analyze `BasicExperiment` defined in Figure 3 where a batch is selected with probability proportional to its actual size and a uniform row is selected from the generated CVR. Before analyzing a single iteration of `BasicExperiment`, we consider the result of Steps (2) to (8) in `BasicExperiment` for some fixed β and adversary $\mathcal{A}$ (and the identity CVR transform). That is, we focus on the random variables r and $\mathsf{D}^{\mathcal{A}}_{\beta}$ defined by the following procedure and denoted as $\texttt{BasicExperiment}_{\beta}$.

Definition of the random variables r and $\mathsf{D}^{\mathcal{A}}_{\beta}$:

(1) $\mathcal{A}$ generates a CVR for β, denoted cvr.

(2) A row $r \in [\mathsf{S}^{\mathrm{act}}_{b}]$ is drawn independently and uniformly at random.

(3) $\mathsf{D}^{\mathcal{A}}_{\beta}$ is defined to be 2 if `CheckConsistent` outputs `Error`.

(4) If $\mathsf{D}^{\mathcal{A}}_{\beta}$ has not already been set to 2 in the step above, let ι be the identifier appearing in row r. The adversary is asked to return a ballot from batch β with identifier ι. If the adversary responds with such a ballot $\mathbf{b}$, $\mathsf{D}^{\mathcal{A}}_{\beta} = (\mathsf{W}_r - \mathsf{L}_r) - (\mathsf{W}_{\mathrm{b}} - \mathsf{L}_{\mathrm{b}})$; otherwise $\mathsf{D}^{\mathcal{A}}_{\beta} = (\mathsf{W}_r - \mathsf{L}_r) + 1$.

Claim 1. *Consider* $\texttt{BasicExperiment}_{\beta}$ *for an adversary $\mathcal{A}$ and a batch β. Then* $\mathbb{E}[\mathsf{D}^{\mathcal{A}}_{\beta}] \geq \mathsf{D}_{\beta}/\mathsf{S}_{\beta}$.

Proof. The random variable $\mathsf{D}^{\mathcal{A}}_{\beta}$ is determined by selection of cvr by $\mathcal{A}$, (independent) uniform selection of r by C, and final selection by $\mathcal{A}$ of a ballot to return. The proof only requires that cvr and r are independent; in particular, cvr may be chosen with arbitrary dependence on the history of the audit. We remark that the same guarantee holds if multiple instances of $\texttt{BasicExperiment}_{\beta}$ occur in parallel, as the independence assumption is guaranteed by C.

We will show that the inequality holds conditioned on any fixed CVR cvr produced by the adversary in the first step; hence it holds for any distribution of CVRs. Note that if `CheckConsistent` = `Error` for this CVR then $\mathsf{D}^{\mathcal{A}}_{\beta} = 2$ and the claim is clearly true. Otherwise, `CheckConsistent` = `OK`, the CVR $\mathrm{cvr} = ((\iota_1, \mathsf{W}_1, \mathsf{L}_1), \ldots, (\iota_s, \mathsf{W}_s, \mathsf{L}_s))$ is uniquely-labeled, $s = \mathsf{S}^{\mathrm{cvr}}_{\beta} = \mathsf{S}^{\mathrm{act}}_{\beta}$, $\mathsf{W}^{\mathrm{cvr}}_{\beta} = \mathsf{W}^{\mathrm{tab}}_{\beta}$, and $\mathsf{L}^{\mathrm{cvr}}_{\beta} = \mathsf{L}^{\mathrm{tab}}_{\beta}$.

For any particular row r of the cvr, let $\mathbf{B}(r) = \{\mathbf{b} \in \mathbf{B}_{\beta} \mid \iota_{\mathrm{b}} = \iota_r\}$ denote the set of ballots with identifier that matches ι_r. Consider the following function of ballots in batch β, denoted $\mathsf{OneB} : [\mathsf{S}_{\beta}] \to \mathbf{B}_{\beta}$ to rows in the CVR:

(1) For a row r for which $|\mathbf{B}(r)| \geq 1$ associate any ballot $\mathbf{b} \in \mathbf{B}(r)$ with r that minimizes the resulting discrepancy (and hence achieves D_r from Definition 7).

(2) Of the remaining, yet unassociated, ballots, assign them arbitrarily, but in a one-to-one fashion, to the rows of the CVR which have ballot identifiers that do not match a physical ballot.

As the CVR is uniquely-labeled there is no contention for the ballots assigned by the first rule. That is, OneB is a one-to-one function between rows and physical ballots. Furthermore, since $\mathsf{S}^{\mathrm{act}}_{\beta} = \mathsf{S}^{\mathrm{cvr}}_{\beta}$ the function OneB is also onto; thus OneB is bijective.

For this fixed β and fixed cvr provided by $\mathcal{A}$, let $\mathsf{D}^{\mathrm{cvr},\mathsf{OneB}}_{r}$ denote the random variable (determined by the random variable r) given by the discrepancy between the votes appearing in row r and $\mathsf{OneB}(r)$. That is,

$$\mathsf{D}^{\mathrm{cvr},\mathsf{OneB}}_{r} = (\mathsf{W}_r - \mathsf{L}_r) - (\mathsf{W}_{\mathrm{b}} - \mathsf{L}_{\mathrm{b}}).$$

We then note that, conditioned on observing a fixed cvr,

$$\mathsf{D}^{\mathrm{cvr},\mathsf{OneB}}_{r} \leq^{(1)} \mathsf{D}^{\mathrm{cvr}}_{r} \leq^{(2)} \mathsf{D}^{\mathcal{A}}_{\beta}$$

with certainty over the uniform choice of r.

The inequality $\leq^{(1)}$ follows immediately from the definition of $\mathsf{D}_r^{\text{cvr}}$: to see this, observe that if $\mathbf{B}(r) \geq 1$ then there is a matching ballot and $\mathsf{D}_r^{\text{cvr},\mathsf{OneB}} = \mathsf{D}_r^{\text{cvr}}$ as they are both determined by minimum discrepancy obtained over all matching ballots; if, on the other hand, there is no matching ballot then the inequality follows because $(\mathsf{W}_r - \mathsf{L}_r) - (\mathsf{W}_\mathbf{b} - \mathsf{L}_\mathbf{b}) \leq (\mathsf{W}_r - \mathsf{L}_r) + 1$ for any ballot $\mathbf{b}$.

As for the second inequality $\leq^{(2)}$, note that if the adversary returns a ballot that matches the identifier for row r, $\mathsf{D}_r^{\text{cvr}} \leq \mathsf{D}_\beta^{\mathcal{A}}$ as above, since D_r is defined to be the minimum value over all matching ballots. If the adversary does not return a matching ballot then $\mathsf{D}_r \leq (\mathsf{W}_r - \mathsf{L}_r) + 1 = \mathsf{D}_\beta^{\mathcal{A}}$, as desired.

We conclude that

$$\mathbb{E}\left[\mathsf{D}_\beta^{\mathcal{A}}\right] = \sum_{\text{cvr}} \Pr[\mathcal{A} \text{ generates cvr}]\, \mathbb{E}[\mathsf{D}_\beta^{\mathcal{A}} \mid \text{cvr}] \geq \sum_{\text{cvr}} \Pr[\mathcal{A} \text{ generates cvr}] \mathop{\mathbb{E}}_{r}\left[\mathsf{D}_r^{\text{cvr},\mathsf{OneB}}\right] . \tag{4}$$

For a fixed cvr, we may expand $\mathbb{E}\left[\mathsf{D}_r^{\text{cvr},\mathsf{OneB}}\right]$ as the sum

$$\frac{1}{\mathsf{S}_\beta} \sum_{r=1}^{\mathsf{S}_\beta} \left((\mathsf{W}_r^{\text{cvr}} - \mathsf{L}_r^{\text{cvr}}) - (\mathsf{W}_{\mathsf{OneB}(r)}^{\text{act}} - \mathsf{L}_{\mathsf{OneB}(r)}^{\text{act}}) \right) . \tag{5}$$

As OneB is bijective, every ballot appears exactly once in this sum, so we can rewrite the quantity in (5)

$$\frac{1}{\mathsf{S}_\beta} \left(\sum_R (\mathsf{W}_R^{\text{cvr}} - \mathsf{L}_R^{\text{cvr}}) - \sum_{\mathbf{b} \in \mathbf{B}_\beta} (\mathsf{W}_\mathbf{b}^{\text{act}} - \mathsf{L}_\mathbf{b}^{\text{act}}) \right) = \frac{\mathsf{D}_\beta}{\mathsf{S}_\beta} .$$

Returning to (4), we have

$$\begin{aligned}
\mathbb{E}\left[\mathsf{D}_\beta^{\mathcal{A}}\right] &\geq \sum_{\text{cvr}} \Pr[\mathcal{A} \text{ generates cvr}]\, \mathbb{E}\left[\mathsf{D}_r^{\text{cvr},\mathsf{OneB}}\right] \\
&= \sum_{\text{cvr}} \Pr[\mathcal{A} \text{ generates cvr}] \frac{\mathsf{D}_\beta}{\mathsf{S}_\beta} \\
&= \frac{\mathsf{D}_\beta}{\mathsf{S}_\beta} \sum_{\text{cvr}} \Pr[\mathcal{A} \text{ generates cvr}] = \frac{\mathsf{D}_\beta}{\mathsf{S}_\beta} ,
\end{aligned}$$

which completes the proof of Claim 1. □

We now turn to analyzing a single iteration of `BasicExperiment`. We define the result of this experiment to be a random variable $\mathsf{D}^{\mathcal{A}}$, defined by the following procedure:

(1) Select a batch β with probability $\mathsf{S}_\beta^{\text{act}}/\mathsf{S}^{\text{act}}$.

(2) Carry out the local experiment with batch β.

Claim 2. *For any adversary $\mathcal{A}$, the expectation of $\mathsf{D}^{\mathcal{A}}$ over a single iteration satisfies*

$$\mathbb{E}[\mathsf{D}^{\mathcal{A}}] = \sum_\beta \left(\frac{\mathsf{S}_\beta^{\text{act}}}{\mathsf{S}^{\text{act}}} \cdot \mathbb{E}[\mathsf{D}_\beta^{\mathcal{A}}] \right) \geq \sum_\beta \left(\frac{\mathsf{S}_\beta^{\text{act}}}{\mathsf{S}^{\text{act}}} \cdot \frac{\mathsf{D}_\beta}{\mathsf{S}_\beta^{\text{act}}} \right) = \frac{\mathsf{D}}{\mathsf{S}} .$$

Theorem 1 follows from Claim 2 by noting that for any invalid election the input $\mathsf{D}^{\mathcal{A}}$ to $(\mathsf{Stop}, \mathsf{R})$ is a $\mathsf{D}/\mathsf{S} \geq \mu^{\text{tab}}$ dominated random variable and by application of Lemma 1. □

5.1 Concrete statistical tests

We recall the Kaplan-Markov test.

Definition 12 (Kaplan-Markov [29,31–33])**.** *Let* $\alpha \in [0,1]$, $\gamma > 1$, $\ell_{\min}, \ell_{\max} \in \mathbb{Z}^{+}$. *Define the value*

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

The $(\alpha, \gamma, \ell_{\min}, \ell_{\max})$-Kaplan-Markov *audit statistical test is* $(\mathsf{Stop}, \mathsf{R})$ *where* $\mathsf{Stop}_\delta(\mathsf{D}_1, ..., \mathsf{D}_\ell) = 1$ *if and only if*

$$\ell \geq \ell_{\max} \vee \left(\mathsf{Risk}_{\delta}^{(\gamma)}(\mathsf{D}_1, ..., \mathsf{D}_\ell,) \leq \alpha \wedge \ell \geq \ell_{\min} \right) \quad \textit{and} \quad \mathsf{R}_\delta(\mathsf{D}_1, ..., \mathsf{D}_\ell, ; \gamma) = \left(\mathsf{Risk}_{\delta}^{(\gamma)}(\mathsf{D}_1, ..., \mathsf{D}_\ell, ; \gamma) \leq \alpha \right).$$

Note: One can define the test without $\ell_{\min}$ or $\ell_{\max}$. The parameter $\ell_{\max}$ is usually set to some small fraction of the overall number of ballots where hand counting becomes more efficient. The parameter $\ell_{\min}$ is usually set so that some number of sampled ballots can display 1-vote overstatements while meeting the risk limit. For a $\lambda * \delta$ fraction of 1-vote overstatements to be acceptable

$$\ell_{\min} = -\log \alpha / \left(\delta \left(\frac{1}{2\gamma} + \lambda \log \left(1 - \frac{1}{2\gamma} \right) \right) \right)$$

suffices [31].

Claim 3. *The Kaplan-Markov test is an adaptive audit test.*

Proof of Claim 3. Consider a sequence of bounded, non-negative and i.i.d. real-valued random variables $X_1, \ldots$, each with mean δ. The Kaplan–Markov inequality asserts that

$$\Pr\left[\max_{t=0}^{n} \prod_{i=1}^{t} (X_i/\delta) \geq 1/\alpha \right] \leq \alpha \qquad \text{for any } \alpha > 0. \tag{6}$$

Critically, we observe that the Kaplan-Markov inequality applies to random variables under the weaker δ-dominating condition. Specifically, assume that $X_1, X_2, \ldots$ are δ-dominating (but not necessarily i.i.d.). Then the sequence of random variables $Z_1 = X_1/\delta, Z_2 = (X_1/\delta)(X_2/\delta), \ldots$ form a nonnegative sub-martingale, which is to say that $\mathbb{E}[Z_t | Z_1, \ldots, Z_{t-1}] \geq Z_{t-1}$. According to the Doob (sub-)martingale inequality, $\mathbb{E}[\max_{i=1}^{n} Z_i] \leq \mathbb{E}[Z_n]$ and hence Markov's inequality can be applied to yield (6), as desired. (See, e.g., [37, §14.6] for a detailed account of the Doob inequality). Finally, the Kaplan-Markov test for δ-dominated random variables is obtained by applying (6) to the observed discrepancies under the transformation $D \mapsto 1 - D/(2\gamma)$. □

Other classical tail bounds directly yield adaptive audit tests by monotonicity or stochastic domination arguments. For example, the Azuma-Hoeffding inequality applies to this situation as it applies directly to submartingales. (See, e.g., [19] for a detailed account.) Inequalities that optimize one side of the tail bound (e.g., the upper Chernoff bound) can be applied to this situation via a stochastic dominance argument that exploits the fact that the test criteria are monotone.

6 Completeness

The second natural figure of merit for an audit is the probability that it correctly concludes that a valid election is "Consistent." Treating this issue is complicated by the fact that inconsistencies between the CVR and the physical ballots are frequently observed even during vigilant audits of valid elections. Thus, the underlying statistical tests must be parameterized in order to tolerate a certain frequency of errors. Ultimately, this leads to a trade-off between risk, sample size, and the probability that a valid election will be found inconclusive when the audit is subject to some presumed rate of inconsistencies. This third quantity we call "completeness"; this is non-standard terminology motivated by directly analogous definitions in cryptography.

The traditional analysis of completeness focuses on the number of overstatements and understatements, either according to the actual ballot population or observed empirically during the audit. The relationship to sample size and risk then depends largely on the details of the adopted statistical test (see [17, 27] and Section 5.1). However, our setting introduces new types of inconsistencies that may arise during an audit: in particular, mismatches between the tabulation and CVR yield a new source of non-zero observed discrepancy.

To provide a comprehensive treatment, we augment the traditional accounting of under- and overstatement errors with two further classes of errors. **Ballot Additions** can result from ballots that are scanned or tabulated more than once (which a tabulator cannot detect without an identifier). **Ballot Deletions** can result from ballots that were cast but never scanned or whose interpretations were not included in the reported results. We remark such errors can also arise in traditional settings. 15% of audited precincts in Connecticut in the 2020 presidential election reported a different ballot count from the tabulation [25]. To the best of our knowledge, this is the first formal detailed analysis of the effect of additions and deletions.

Handling size, tally, and uniquely-labeled failures via the CVR transform mechanism. Recall that the strict "default" auditor (that is, the procedure of Figure 3 using $\mathcal{T}_{\mathrm{Id}}$) rejects CVRs resulting from commonplace errors. For example, if the CVR has one fewer row than the size of the batch or if $\mathsf{W}^{\mathrm{cvr}} = \mathsf{W}^{\mathrm{tab}} + 1$. To eliminate such errors, $\mathcal{T}_{\mathrm{Force}}$ forcibly revises the CVR so as to declare sizes and vote totals consistent with the manifest and tabulation. While this transformation corrects the CVR in this sense, it may generate new overstatements or understatements. The CVR transform paradigm provides a unified way to treat such errors by converting them into understatement and overstatement errors, which have a well understood effect on standard statistical tests.

In light of the discussion above, this section provides precise control on the effect of size mismatches, vote tally disagreements, or duplicated identifiers on the resulting number of overstatements and understatements. With these equivalencies in hand, one can compute appropriate sample sizes for different statistical tests by established techniques [17, 27]. As remarked above, this approach can also be used to treat similar issues in traditional comparison audits.

We separately present and analyze two different settings. The first setting considers a consistent CVR and tabulation that disagree with the physical ballots. The second setting considers an arbitrary tabulation in context of an inconsistent CVR. We compose these in Section 6.1 to handle the general case.

Definition 13 (The canonical CVR)**.** *Let* $\mathbf{B}$ *be a uniquely labeled ballot family. A global CVR* $\mathrm{cvr}^* = (\mathrm{cvr}_1^*, \ldots, \mathrm{cvr}_k^*)$ *is* canonical *if it correctly reflects the ballots. That is, the ballots* $\mathbf{B}_\beta$ *can be placed in one-to-one correspondence with the rows of* cvr_β *in such a way that both the identifiers and votes match. For the ballot family* $\mathbf{B}$*,* $\mathrm{cvr}_{\mathbf{B}}^*$ *indicates a canonical CVR.*

Observe that any canonical CVR is uniquely labeled. The canonical CVR is only determined up to a permutation of the rows. Despite this, we say "the canonical CVR" of a ballot family.

Definition 14 (The honest adversary)**.** *Let* $E = (\mathbf{B}, T)$ *be an election with uniquely labeled ballots and let* cvr *be a uniquely labeled global CVR. The* honest adversary $\mathcal{H}(\mathbf{B}, \mathrm{cvr}, T)$ *is the adversary that responds to any CVR request with the appropriate* cvr_i *and responds to any request for an (existing) ballot identifier* ι *with the matching ballot* $\mathbf{b}$*. If no ballot exists matching the identifier, it returns* ***No ballot****.*

The honest adversary's behavior is only defined if all ballots have unique identifiers and the cvr is uniquely labeled.

Definition 15 (Pairwise CVR discrepancy)**.** *Let* $\mathrm{cvr}_1, \mathrm{cvr}_2$ *be two uniquely labeled CVRs (for the same batch of a ballot family). For an identifier* ι *that appears in both CVRs, define*

$$\mathsf{D}(\mathrm{cvr}_1, \mathrm{cvr}_2, \iota) = (\mathsf{W}_{r_\iota}^{\mathrm{cvr}_1} - \mathsf{L}_{r_\iota}^{\mathrm{cvr}_1}) - (\mathsf{W}_{r_\iota}^{\mathrm{cvr}_2} - \mathsf{L}_{r_\iota}^{\mathrm{cvr}_2})\,.$$

Definition 16 (CVR distortion)**.** *Let* $\mathbf{B}$ *be a uniquely labeled ballot family and* $\mathrm{cvr} = (\mathrm{cvr}_1, \ldots, \mathrm{cvr}_k)$ *be a global CVR for* $\mathbf{B}$*. Let* $(o_1, o_2, u_1, u_2, a, d)$ *be natural numbers such that* $\mathsf{W}^{\mathrm{cvr}}, \mathsf{L}^{\mathrm{cvr}}, \mathsf{S}^{\mathrm{cvr}} - \mathsf{W}^{\mathrm{cvr}}, \mathsf{S}^{\mathrm{cvr}} - \mathsf{L}^{\mathrm{cvr}}$ *are all at least* $o_1 + o_2 + u_1 + u_2 + a + d$*. Then* $\tilde{\mathrm{cvr}}$ *is a* $(o_1, o_2, u_1, u_2, a, d)$-distortion *of* cvr *if* $\tilde{\mathrm{cvr}} = \mathrm{cvr}$ *with the following exceptions:*

Overstatements/Understatements. *There are*

- o_1 *identifiers* ι *where* $\mathsf{D}(\tilde{\mathrm{cvr}}, \mathrm{cvr}, \iota) = 1$,
- o_2 *identifiers* ι *where* $\mathsf{D}(\tilde{\mathrm{cvr}}, \mathrm{cvr}, \iota) = 2$,
- u_1 *identifiers* ι *where* $\mathsf{D}(\tilde{\mathrm{cvr}}, \mathrm{cvr}, \iota) = -1$,
- u_2 *identifiers* ι *where* $\mathsf{D}(\tilde{\mathrm{cvr}}, \mathrm{cvr}, \iota) = -2$,

Deletions *There are d identifiers ι appearing in* cvr *that do not appear in* $\tilde{\mathrm{cvr}}$.

Additions *There are a identifiers ι appearing in* $\tilde{\mathrm{cvr}}$ *that do not appear in* cvr *or on any ballot.*

Definition 17 (Tabulation of CVR)**.** *Let* cvr *be a global CVR for a ballot family* $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_k)$. *The tabulation of* cvr *is*

$$\mathrm{Tab}(\mathrm{cvr}) = ((\mathsf{S}^{\mathrm{cvr}_1}, \mathsf{W}^{\mathrm{cvr}_1}, \mathsf{L}^{\mathrm{cvr}_1}), ..., (\mathsf{S}^{\mathrm{cvr}_k}, \mathsf{W}^{\mathrm{cvr}_k}, \mathsf{L}^{\mathrm{cvr}_k}))\ .$$

A tabulation T is consistent *with a global CVR* cvr *if* $T = \mathrm{Tab}(\mathrm{cvr})$.

Our first claim bounds the (probability distribution of) discrepancy when the tabulation and CVR are consistent but are inconsistent with the physical ballots.

Claim 4. *Let $(o_1, u_1, o_2, u_2, a, d)$ be natural numbers, let $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_k)$ be a ballot family with canonical CVR* cvr^**, and let* $\tilde{\mathrm{cvr}} = (\tilde{\mathrm{cvr}}_1, ..., \tilde{\mathrm{cvr}}_k)$ *be a $(o_1, u_1, o_2, u_2, a, d)$-distortion of* cvr^*. *For a single iteration of $C[\mathcal{T}_{\mathrm{Force}}]$ interacting with* $\mathcal{H}(\mathbf{B}, \tilde{\mathrm{cvr}}, \mathrm{Tab}(\tilde{\mathrm{cvr}}))$,

$$\begin{aligned}
\frac{o_2 - 2a - d}{\mathsf{S}^{\mathrm{act}}} &\le \Pr[\mathsf{D}^{\mathcal{H}} = 2] \le \frac{o_2 + a + 2d}{\mathsf{S}^{\mathrm{act}}},\\
\frac{o_1 - 3a - 2d}{\mathsf{S}^{\mathrm{act}}} &\le \Pr[\mathsf{D}^{\mathcal{H}} = 1] \le \frac{o_1 + 2a + 3d}{\mathsf{S}^{\mathrm{act}}},\\
\frac{u_1 - 3a - 2d}{\mathsf{S}^{\mathrm{act}}} &\le \Pr[\mathsf{D}^{\mathcal{H}} = -1] \le \frac{u_1 + 2a + 2d}{\mathsf{S}^{\mathrm{act}}},\\
\frac{u_2 - 2a - d}{\mathsf{S}^{\mathrm{act}}} &\le \Pr[\mathsf{D}^{\mathcal{H}} = -2] \le \frac{u_2 + a + d}{\mathsf{S}^{\mathrm{act}}}.
\end{aligned}$$

Furthermore, for $e = o_1 + o_2 + u_1 + u_2$ we have

$$\frac{1 - e - (3a + 3d)}{\mathsf{S}^{\mathrm{act}}} \le \Pr[\mathsf{D}^{\mathcal{H}} = 0] \le \frac{1 - e + (3a + 3d)}{\mathsf{S}^{\mathrm{act}}}.$$

Proof. Consider some fixed batch β. In the absence of additions and delections, overstatement and understatement errors are immediate. We now consider two cases where the size of the batch is too large and when it is too small.

Let $\mathsf{S}^{\mathrm{cvr}}_\beta > \mathsf{S}^{\mathrm{act}}_\beta$. Then $\mathsf{S}^{\mathrm{cvr}}_\beta - \mathsf{S}^{\mathrm{act}}_\beta$ rows will be deleted from the cvr. These deleted rows could correspond to any possible discrepancy value. Note other rows will be adjusted to deal with the discrepancy of the deleted rows. At most one vote for a winner can be added to a single row and at most one vote for a loser can be added to a single row. If these are added the same row they do not change the discrepancy. Otherwise, they increase the discrepancy of one row and decrease the discrepancy of another row. Thus, to compensate for the removal of a row 2 instances of a discrepancy of $-1, 0, 1$ can be removed and 2 added. Compensation can remove two instances of $2, -2$ discrepancy and create at most 1 row of discrepancy $2, -2$ since a discrepancy of $2, -2$ can never be achieved by subtracting or increasing discrepancy respectively. This yields the bounds for a in Claim 4.

Now consider the case when $\mathsf{S}^{\mathrm{cvr}}_\beta < \mathsf{S}^{\mathrm{act}}_\beta$; then $\mathsf{S}^{\mathrm{act}}_\beta - \mathsf{S}^{\mathrm{cvr}}_\beta$ rows will be added to the CVR with identifier $\perp_i$. Note that the votes on this row can be any value but there will be no matching ballot leading to a discrepancy value of 0, 1 or 2. To keep the CVR consistent with the CVR at most 2 records can have their totals adjusted as with additions. As before, only a single row can be created with a discrepancy of $2, -2$ per deletion. ◻

Recall that $\mathcal{T}_{\text{Force}}$ forces the CVR to be consistent with the tabulation; thus the transformed CVR has the same discrepancy as the tabulation with the actual ballots. Ideally, the observed random variable D, arising from $\tilde{\text{cvr}}$ under $\mathcal{T}_{\text{Force}}$, would be identical to that arising from the original CVR $\tilde{\text{cvr}}$. In the case of only overstatement and understatement errors this is achieved.

However, this is not achieved in the case of additions and deletions. Recall that the tabulation and $\tilde{\text{cvr}}$ are consistent. The corrections that happen in $\mathcal{T}_{\text{Force}}$ are size corrections due to additions and deletions. Ideally, $\mathcal{T}_{\text{Force}}$ would respond to a deletion by "adding back" the deleted row but it has no information about the votes or identifier on the deleted ballot. Furthermore, any row that is added back may require other rows of the $\tilde{\text{cvr}}$ to be adjusted for consistency with the tabulation.

Similarly, $\mathcal{T}_{\text{Force}}$ would ideally respond to addition by deleting the added row but in general it cannot identify the added row. The row it chooses to delete can then yield changes to the discrepancy distribution as indicated above. Thus, the response to additions can increase or decrease the mean of D depending on where they are located. The response to deletions can never cause a negative discrepancy value because the added row's identifier does not appear on any ballot.

We now consider the case where errors are introduced between the tabulation and the CVR. In this setting we assume that the tabulation has arbitrary disagreements with the canonical CVR so that the effect of $\mathcal{T}_{\text{Force}}$ is to ensure that the CVR for β has the same discrepancy as the tabulation. This means that the expectation of observed discrepancy will have the same mean but $\mathcal{T}_{\text{Force}}$ can increase the probability that the observed discrepancy is nonzero, increasing the variance. That is, errors reduce the chance that the observed discrepancy will be 0. In both Claims 4 and 5 the actual distribution of discrepancy depends on the distribution of errors between batches.

Claim 5. *Let $(o'_1, u'_1, o'_2, u'_2, a', d')$ be natural numbers and let $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_k)$ be a ballot family. Let T be a tabulation for $\mathbf{B}$ and let cvr_T be a uniquely labeled global CVR that is consistent with T (so that $T = \text{Tab}(\text{cvr}_T)$). Define $d_{-2}, d_{-1}, d_0, d_1, d_2$ so that for a single iteration of $C[\mathcal{T}_{\text{Force}}]$ interacting with $\mathcal{H}(\mathbf{B}, \text{cvr}_T, T)$),*

$$\forall i, d_i = \Pr[\mathsf{D}^{\mathcal{H}} = i] \text{ and } d_e := \sum_i i \cdot d_i.$$

Let $\tilde{\text{cvr}} = (\tilde{\text{cvr}}_1, ..., \tilde{\text{cvr}}_k)$ be a $(o'_1, u'_1, o'_2, u'_2, a', d')$-distortion of cvr_T. For a single iteration of $C[\mathcal{T}_{\text{Force}}]$ interacting with $\mathcal{H}((\mathbf{B}_1, \ldots, \mathbf{B}_k), \tilde{\text{cvr}}, T)$ one has that

$$\begin{aligned}
\Pr[\mathsf{D}^{\mathcal{H}} = 2] &\in d_2 \pm \frac{o'_2 + o'_1 + 2u'_2 + u'_1 + 2a' + 3d'}{\mathsf{S}^{\text{act}}}, \\
\Pr[\mathsf{D}^{\mathcal{H}} = 1] &\in d_1 \pm \frac{2o'_2 + 2o'_1 + 2u'_2 + 2u'_1 + 2a' + 3d'}{\mathsf{S}^{\text{act}}}, \\
\Pr[\mathsf{D}^{\mathcal{H}} = 0] &\in d_0 \pm \frac{2o'_2 + 2o'_1 + 2u'_2 + 2u'_1 + 3a' + 3d'}{\mathsf{S}^{\text{act}}}, \\
\Pr[\mathsf{D}^{\mathcal{H}} = -1] &\in d_{-1} \pm \frac{2o'_2 + 2o'_1 + 2u'_2 + 2u'_1 + 2a' + 3d'}{\mathsf{S}^{\text{act}}}, \\
\Pr[\mathsf{D}^{\mathcal{H}} = -2] &\in d_{-2} \pm \frac{2o'_2 + o'_1 + u'_2 + u'_1 + 2a' + 3d'}{\mathsf{S}^{\text{act}}},
\end{aligned}$$

and $\mathbb{E}[\mathsf{D}^{\mathcal{H}}] = d_e$.

Proof. Consider some fixed batch β. For a batch with an addition, some row will be deleted which can have an arbitrary discrepancy value. As in the proof of Claim 4 in the worst case to compensate for the vote totals on the deleted row, one row will have $\mathsf{W}_r - \mathsf{L}_r$ increased and another row will have $\mathsf{W}_r - \mathsf{L}_r$ decreased.

We now consider deletions. A row may be added which begins with discrepancy 0. The deleted row had an arbitrary discrepancy. When new rows are added to compensate for the deleted rows the discrepancy of the $\tilde{\text{cvr}}$ must be adjusted to match the tabulation. For each deletion, the newly added row can have any vote pattern. As before, the created row could have a vote pattern different from the ballot that was deleted. This leads to other ballots having their vote totals adjusted to ensure the total discrepancy between $\tilde{\text{cvr}}_\beta$ and the tabulation is 0. At most two ballots have to be adjusted to compensate for this created row. These adjustments can create any discrepancy.

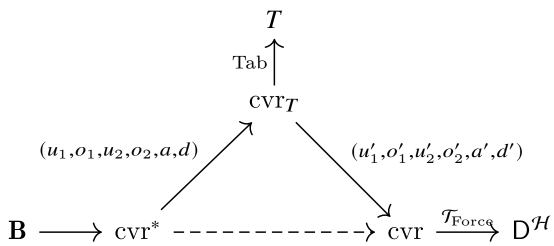


Figure 6: Claim 4 bounds the probability of each discrepancy value for the case when errors are introduced from canonical CVR and tabulation. Claim 5 bounds the probability of each discrepancy value for the case when (additional) errors are introduced from tabulation to the produced batch CVRs.

Now consider an o_2 error. This means there is some row $(\iota, 0, 1)$ moved to $(\iota, 1, 0)$ in the CVR. As such in the worst case the checks in Step (3) and (4) will not pass in Figure 3 (this would not be the case if u_1 or u_2 errors occur in the same batch). Namely, $\mathsf{W}^{\text{cvr}} > \mathsf{W}^{\text{tab}}$ and $\mathsf{L}^{\text{cvr}} < \mathsf{L}^{\text{tab}}$. To compensate for this the procedure in Figure 4 will change some W vote from 1 to 0 and some L vote from 0 to 1. If both of these changes happen on the vote with the o_2 error then no problem occurs. If it happens on two separate this decreases the discrepancy of two rows. Analysis for the other cases proceeds in a similar fashion. □

Since the CVR is forced to have the same discrepancy as tabulation, after applying $\mathcal{T}_{\text{Force}}$ the produced CVR has the same discrepancy as the tabulation. But $\mathcal{T}_{\text{Force}}$ could increase the probability that discrepancy is nonzero. There are statistical tests that only depend on the expected value of $\mathsf{D}^{\mathcal{A}}$. However, Risk, and thus Stop, of Kaplan-Markov (and many other statistical tests) depends on the entire distribution of D (not just its expectation), so these errors do affect stopping time.

6.1 Composing the two error models

Figure 6 describes a comprehensive error model where errors are first added from the canonical CVR and the tabulation and then further errors are added to the CVRs provided to the honest adversary. The bounds obtained by composing Claims 4 and 5 are below:

$$\begin{aligned}
\Pr[\mathsf{D}^{\mathcal{H}} = 2] &\in o_2 \pm \frac{2a + 2d + o'_2 + o'_1 + 2u'_2 + u'_1 + 2a' + 3d'}{\mathsf{S}^{\text{act}}},\\
\Pr[\mathsf{D}^{\mathcal{H}} = 1] &\in o_1 \pm \frac{3a + 3d + 2o'_2 + 2o'_1 + 2u'_2 + 2u'_1 + 2a' + 3d'}{\mathsf{S}^{\text{act}}},\\
\Pr[\mathsf{D}^{\mathcal{H}} = -1] &\in u_1 \pm \frac{3a + 2d + 2o'_2 + 2o'_1 + 2u'_2 + 2u'_1 + 2a' + 3d'}{\mathsf{S}^{\text{act}}},\\
\Pr[\mathsf{D}^{\mathcal{H}} = -2] &\in u_2 \pm \frac{2a + d + 2o'_2 + o'_1 + u'_2 + u'_1 + 2a' + 3d'}{\mathsf{S}^{\text{act}}},\\
\Pr[\mathsf{D}^{\mathcal{H}} = 0] &\geq 1 - \frac{(o_2 + o_1 + u_1 + u_2) + 2(o'_2 + o'_1 + u'_2 + u'_1)}{\mathsf{S}^{\text{act}}} - \frac{3(a + d + a' + d')}{\mathsf{S}^{\text{act}}}.
\end{aligned}$$

Ideally one would show error bounds for an arbitrary combination of ballots, tabulation, and global CVR. Our bounds assume errors are added to global CVRs in two stages first to tabulation, and then to CVRs returned in the audit. We found global CVRs for each stage to be the most natural way to track differences. This leads to final bounds that assume a particular distorted CVR used to produce the tabulation that is not seen by any party.

7 Adaptive Group Comparison Audits

We described a methodology to perform ballot comparison audits without the need to generate a global CVR for the entire election. As described in the introduction, no such CVR is necessary if one wishes to perform a batch comparison audit in settings where tabulated totals are available for the relevant batches. In this section, we show that a hybrid of these techniques is possible that permits tabulated batches to be broken into smaller untabulated collections that we call *groups*; these groups of ballots are then treated analogously to individual ballots in the adaptive audit. In particular, an audit can hand-count appropriately selected groups and compare these against an adaptively generated "group CVR" that declares totals for each group. This yields a trade-off between the size of the groups (and hence the effort involved in hand counting them) and the number of groups. Ballots do not need to be given identifiers in this procedure, though groups must be identifiable.

Batch comparison audits We begin by reviewing conventional batch comparison audits, the third major family of risk-limiting audits used in practice. We borrow notation from Definition 2. For an election E, a batch comparison audit consists of multiple iterations of the following experiment:

(1) A batch is selected with probability proportional to size.

(2) A full hand count is conducted for the batch.

(3) The observed discrepancy between the tabulated totals and the hand count is computed.

The envisioned hybrid audit procedure is as follows:

(1) A batch is selected with probability proportional to size.

(2) The batch is separated into ν groups and an untrusted "group CVR" is generated. This CVR reports the size, vote total for W, and vote total for L for each group in the selected batch. Thus the CVR consists of ν triples $(\mathsf{S}_{\beta,\mathsf{g}}, \mathsf{W}_{\beta,\mathsf{g}}, \mathsf{L}_{\beta,\mathsf{g}})$, one for each value of $\mathsf{g} \in [\nu]$.

(3) A group g is selected with probability proportional to its purported size, $\mathsf{S}_{\beta,\mathsf{g}}$.

(4) A full hand count is conducted for group g. Let $\mathsf{S}^{\mathrm{act}}_{\beta,\mathsf{g}}$, $\mathsf{W}^{\mathrm{act}}_{\beta,\mathsf{g}}$, and $\mathsf{L}^{\mathrm{act}}_{\beta,\mathsf{g}}$ denote the size and relevant totals.

(5) The observed discrepancy is
$$\mathsf{D}^{\mathcal{A}} := \frac{((\mathsf{W}^{\mathrm{cvr}}_{\beta,\mathsf{g}} - \mathsf{L}^{\mathrm{cvr}}_{\beta,\mathsf{g}}) - (\mathsf{W}^{\mathrm{act}}_{\beta,\mathsf{g}} - \mathsf{L}^{\mathrm{act}}_{\beta,\mathsf{g}}))}{\mathsf{S}_{\beta,\mathsf{g}}}.$$

Such a procedure may be preferable to batch comparison audits as one effectively identifies groups of ballots rather than individual ballots. Additionally, as the number of groups is typically much smaller than the number of ballots, it may be easier to identify and locate a particular group of ballots rather than identify an individual ballot. Of course, each comparison step in such an audit requires hand counting an entire group.

The sizes of groups declared in the group CVR is not assumed to be correct. Note, however that the notion of batch and the assumptions pertaining to batches—in particular that a correct manifest is supplied to the auditor—are common in the two approaches.

7.1 Adapting the Formalism

We now introduce a second *Auditor–Adversary* game for adaptive group comparison audits. The relevant notions of election, vote totals, and ballot manifest are identical to those of Section 4, though ballot identifiers are irrelevant for this approach. (Rather than formally redefine the notion of ballot collection to remove identifiers, we leave the notion unchanged and remark that they are unused.) The meaning of a CVR is adapted as indicated above so that it declares sizes and vote totals for groups in a batch (but contains no information about individual ballots). Figure 7 describes the adaptive batch RLA game between the auditor and adversary.

Definition 18 (Group Cast-Vote Record (CVR) syntax.)**.** *Let* $E = (\mathbf{B}, T)$ *be an election. A* Group Cast-Vote Record Table (CVR) *for batch* β *of* ν *groups is a sequence of tuples*
$$\mathrm{cvr}_\beta = ((\mathsf{S}^{\mathrm{cvr}}_1, \mathsf{W}^{\mathrm{cvr}}_1, \mathsf{L}^{\mathrm{cvr}}_1), \ldots, (\mathsf{S}^{\mathrm{cvr}}_\nu, \mathsf{W}^{\mathrm{cvr}}_\nu, \mathsf{L}^{\mathrm{cvr}}_\nu)),$$
where each coordinate is a natural number. We borrow general notation from Definition 5. We say that a CVR is well-formed *if* $\forall \mathsf{g} \in [\nu]$ *it holds that* $\max(\mathsf{W}^{\mathrm{cvr}}_\mathsf{g}, \mathsf{L}^{\mathrm{cvr}}_\mathsf{g}) \leq \mathsf{S}^{\mathrm{cvr}}_\mathsf{g}$.

Auditor (C)–Adversary ($\mathcal{A}$) game for election $E = (\mathbf{B}, T)$

(1) **Setup**.

(a) **Ballot and tabulation delivery (to $\mathcal{A}$).** The physical ballots $\mathbf{B}$ and the tabulation T are given to the adversary $\mathcal{A}$.

(b) **Ballot manifest and tabulation delivery (to C).** The ballot manifest $\mathsf{S}_E = (\mathsf{S}_1^{\text{act}}, \ldots, \mathsf{S}_k^{\text{act}})$ and the tabulation T are given to the auditor C.

(2) **Audit**. C repeatedly makes one of the following two requests of $\mathcal{A}$, or chooses to conclude the audit:

- **Group CVR request**. For some β, C requests a CVR for batch β. If the batch is not yet partitioned, $\mathcal{A}$ selects a natural number $\nu \geq 1$ and indelibly assigns each ballot $\mathbf{b} \in \mathbf{B}_\beta$ to a group $\mathrm{g} \in [\nu]$. Denote the partition of groups that arise from this assignment $\mathbf{B}_{\beta,1}, \ldots, \mathbf{B}_{\beta,\nu}$. $\mathcal{A}$ responds with a group CVR denoted $\mathtt{CVR}_\beta$.
- **Group request** For some batch β that has been partitioned into ν groups by $\mathcal{A}$, the auditor C requests the physical ballots for a particular group $\mathrm{g} \in [\nu]$. $\mathcal{A}$ responds with $\mathbf{B}^*_{\beta,\mathrm{g}} \subseteq \mathbf{B}_{\beta,\mathrm{g}}$.

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

Figure 7: The $\mathsf{RLA}_{\mathbf{Group},C,\mathcal{A}}(E)$ auditing game.

At certain points in the security game, the adversary must partition the ballots from a batch into groups. Once the batch is partitioned, this decision is immutable; the adversary may not change the partitioning later. Furthermore, when a group is requested by the auditor, we require that the adversary responds with a subset of the selected group. (Equivalently, one may think of the ballots as being indelibly assigned to groups in such a way that the auditor can determine the group to which a ballot is assigned and so detect any situation where the adversary might attempt to include in his response a ballot from another group.) Soundness for the above game is as in Definition 11: an auditor is *α-risk limiting* if for any invalid election E and any adversary $\mathcal{A}$,

$$\Pr_C[\mathsf{RLA}_{\mathbf{Group},C,\mathcal{A}}(E) = \textsc{Consistent}] \leq \alpha\,.$$

7.2 The Auditor

We now present an auditor for the adaptive group setting in Figure 8 (which adapts Figure 3). As before, to argue soundness, we consider an identity CVR transform function $\mathcal{T}_{\mathrm{Id}}$.

Next we show that BasicExperiment yields a $\mathsf{D}/|\mathsf{B}|$-dominating random variable $\mathsf{D}^{\mathcal{A}}$. Similarly to the treatment of Claim 1 for ballot comparison audits, we begin by focusing on the conditional distribution arising from fixing a particular batch β (in the first step of BasicExperiment). We let $\mathtt{BasicExperiment}_\beta$ refer to this experiment and let $\mathsf{D}_\beta^{\mathcal{A}}$ denote the random variable that arises at the conclusion of the experiment. As in the analysis of Claim 1, observe that g is independent of the partitioning and CVR generated by the adversary. The analysis of the full experiment BasicExperiment then follows by linearity of expectation (Claim 7). We implicitly work in the context of an arbitrary, but fixed, election E with the constraints and assumptions arising from the portion of the audit preceding the batch and group sampling iterations.

Claim 6. *Consider* $\mathtt{BasicExperiment}_\beta$ *in the context of an election* $E = (\mathbf{B}, T)$*. Then*

$$\mathbb{E}[\mathsf{D}_\beta^{\mathcal{A}}] \geq \mathsf{D}_\beta/\mathsf{S}_\beta\,.$$

Proof. Let $\mathbf{B}_1, \ldots, \mathbf{B}_\nu$ be the partition of ballots created by the adversary for batch β and let cvr be the CVR returned by the adversary. We prove the claim for an arbitrary, fixed choice of cvr and $(\mathbf{B}_\mathrm{g})_{\mathrm{g}=1}^{\nu}$; the claim then holds for any distribution over these values. Recall that $\sum_{\mathrm{g}=1}^{\nu} |\mathbf{B}_\mathrm{g}| = \mathsf{S}_\beta$. Note that if CheckConsistent = Error then $\mathsf{D}_\beta^{\mathcal{A}} = 2$. The

Auditor $C[\mathcal{T}, (\mathsf{Stop}, \mathsf{R})]$ for an election E

(1) Receive ballot manifest and tabulation:

$$\mathsf{S}_E^{\text{act}} = (\mathsf{S}_1^{\text{act}}, \ldots, \mathsf{S}_k^{\text{act}}); \qquad T = (\mathsf{S}_1^{\text{tab}}; \mathsf{W}_1^{\text{tab}}, \mathsf{L}_1^{\text{tab}}), \ldots, (\mathsf{S}_k^{\text{tab}}; \mathsf{W}_k^{\text{tab}}, \mathsf{L}_k^{\text{tab}}))\,.$$

(2) For $\beta = 1$ to k: (a) $\mathsf{S}_\beta^{\text{tab}} := \mathsf{S}_\beta^{\text{act}}$;
(b) $\mathsf{W}_\beta^{\text{tab}} := \min(\mathsf{W}_\beta^{\text{tab}}, \mathsf{S}_\beta^{\text{act}})$;
(c) $\mathsf{L}_\beta^{\text{tab}} := \min(\mathsf{L}_\beta^{\text{tab}}, \mathsf{S}_\beta^{\text{act}})$.

(3) Let $\quad S^{\text{act}}, S^{\text{tab}} := \sum_{\beta=1}^{k} \mathsf{S}_\beta^{\text{tab}} = \sum_{\beta=1}^{k} \mathsf{S}_\beta^{\text{act}}$.

$$\mu := \frac{\sum_{\beta=1}^{k}(\mathsf{W}_\beta^{\text{tab}} - \mathsf{L}_\beta^{\text{tab}})}{\mathsf{S}^{\text{act}}}\,.$$

(4) If $\mu \leq 0$ return `Inconclusive`.
(5) Initialize iter $= 0$.
(6) Repeat until $\mathsf{Stop}_\mu(\mathsf{D}_1, \ldots, \mathsf{D}_{\mathsf{iter}}) = 1$:
  (a) Increment iter := iter + 1.
  (b) Perform $\mathsf{D}_{\mathsf{iter}}$:= `BasicExperiment`
(7) If $\mathsf{R}_\mu(\mathsf{D}_1, \ldots, \mathsf{D}_{\mathsf{iter}}) = 1$ return `Consistent`
  else return `Inconclusive`.

`BasicExperiment`:

(1) Select batch β with probability $\mathsf{S}_\beta^{\text{tab}}/\mathsf{S}^{\text{tab}}$.
(2) Request CVR for batch β. Response denoted cvr_β.
(3) Apply the transform: $\text{cvr}_\beta := \mathcal{T}(\mathsf{S}_E^{\text{act}}, T, \text{cvr}_\beta)$.
(4) Pick g with probability $\mathsf{S}_{\beta,\mathrm{g}}/\mathsf{S}_\beta^{\text{tab}}$.
(5) If `CheckConsistent`$(\mathsf{S}_E^{\text{act}}, T, \text{cvr}_\beta)$ = `Error`, Return 2.
(6) Ask adversary for ballot group g from batch β.
(7) Let $\mathbf{B}_{\beta,\mathrm{g}}$ denote the returned ballots.
(8) If $|\mathbf{B}_{\beta,\mathrm{g}}| \neq \mathsf{S}_{\beta,\mathrm{g}}$, return 2.
(9) Let $\mathsf{W}^{\text{act}}, \mathsf{L}^{\text{act}} \in \mathbb{N}$ denote the vote totals of the ballots returned by the adversary.
(10) Return $((\mathsf{W}_\mathrm{g}^{\text{cvr}} - \mathsf{L}_\mathrm{g}^{\text{cvr}}) - (\mathsf{W}^{\text{act}} - \mathsf{L}^{\text{act}}))/\mathsf{S}_{\beta,\mathrm{g}}$.

`CheckConsistent`$(\mathsf{S}_E^{\text{act}}, T, \text{cvr}_\beta)$:

(1) If cvr_β is not well formed (Def. 18) return `Error`.
(2) If $\mathsf{S}_\beta^{\text{cvr}}, \mathsf{S}_\beta^{\text{act}}, \mathsf{S}_\beta^{\text{tab}}, \sum_\mathrm{g} \mathsf{S}_{\beta,\mathrm{g}}^{\text{cvr}}$ are not all equal, return `Error`.
(3) If $\sum_\mathrm{g} \mathsf{W}_{\beta,\mathrm{g}}^{\text{cvr}} \neq \mathsf{W}_\beta^{\text{tab}}$ or $\sum_\mathrm{g} \mathsf{L}_{\beta,\mathrm{g}}^{\text{cvr}} \neq \mathsf{L}_\beta^{\text{tab}}$, return `Error`.
(4) Return `OK`.

$\mathcal{T}_{\text{Id}}(\mathsf{S}_E^{\text{act}}, T, \text{cvr}_\beta)$:

(1) Return cvr_β.

Figure 8: The auditor $C_{\mathcal{T},(\mathsf{Stop},\mathsf{R})}$ for adaptive group comparison.

claim is clearly true in this case since $\mathsf{D}_\beta/\mathsf{S}_\beta \le 2$ by definition. We work with the assumption `CheckConsistent = OK`, and hence $\sum_{g=1}^{\nu} \mathsf{S}_{\beta,g} = \sum_{g=1}^{\nu} |\mathbf{B}_g| = \mathsf{S}_\beta$, for the remainder of the proof.

In general, for a partition $(\mathbf{A}_1, \ldots, \mathbf{A}_\nu)$ of the ballots in $\mathbf{B}_\beta$ and a family of ballot subsets $(\mathbf{A}_1^*, \ldots, \mathbf{A}_\nu^*)$ with the property that $\forall g$, $\mathbf{A}_g^* \subset \mathbf{A}_g$, we let $\mathsf{D}_\beta((\mathbf{A}_g)_{g=1}^{\nu}; (\mathbf{A}_g^*)_{g=1}^{\nu})$ denote the random variable arising from the experiment if the adversary initially forms the partition given by $\mathbf{A}_g$, sends cvr to C, and then answers any request for group g with $\mathbf{A}_g^*$. We let $(\mathbf{B}_g^*)_{g=1}^{\nu}$ be the set family determined by the adversary $\mathcal{A}$ so that by definition $\mathsf{D}_\beta^{\mathcal{A}} = \mathsf{D}_\beta((\mathbf{B}_g)_{g=1}^{\nu}; (\mathbf{B}_g^*)_{g=1}^{\nu})$. The sets $\mathbf{B}_g^*$ might not cover all the ballots in $\mathbf{B}_\beta$.

We now show that there exists a partition of ballots $(\mathbf{B}_g^{\min})_{g=1}^{\nu}$ with the property that $\forall g$, $|\mathbf{B}_g^{\min}| = \mathsf{S}_g$ and, moreover, $\mathsf{D}_\beta^{\mathcal{A}} \ge \mathsf{D}_\beta((\mathbf{B}_g^{\min})_{g=1}^{\nu}; (\mathbf{B}_g^{\min})_{g=1}^{\nu})$ (with certainty over choice of g). (Note that in this experiment the same set system is used for the initial partition and the answers of the adversary to group requests.) To define the partition $(\mathbf{B}_g^{\min})_{g=1}^{\nu}$:

- We say that a group g is viable $|\mathbf{B}_g^*| = \mathsf{S}_g$. In this case, define $\mathbf{B}_g^{\min} = \mathbf{B}_g^*$. Let $\mathbf{B}_{\text{viable}} = \bigcup_{g|g \text{ is viable}} \mathbf{B}_g^*$.
- The sets $\mathbf{B}_g^{\min}$ for nonviable g are defined to form an arbitrary partition of the remaining ballots $\mathbf{B}_\beta \setminus \mathbf{B}_{\text{viable}}$ with the size constraints $\forall$ nonviable g, $|\mathbf{B}_g^{\min}| = \mathsf{S}_g$. Note that this is always possible because $\sum \mathsf{S}_{\beta,g} = |\mathbf{B}_\beta|$.

Any size mismatch (when the subset of ballots returned by the adversary for a request for group g does not have size $\mathsf{S}_{\beta,g}$) results in a maximal, default discrepancy of 2. It follows that

$$\mathsf{D}_\beta^{\mathcal{A}} = \mathsf{D}_\beta((\mathbf{B}_g)_{g=1}^{\nu}; (\mathbf{B}_g^*)_{g=1}^{\nu}) \ge \mathsf{D}_\beta((\mathbf{B}_g^{\min})_{g=1}^{\nu}), (\mathbf{B}_g^{\min})_{g=1}^{\nu}) \,.$$

Specifically, note that g is drawn according to the same distribution in the two experiments and, for any viable g, these two random variables take the same value; for any nonviable g the first takes the default value of 2, while the second is

$$\frac{(\mathsf{W}_g^{\text{cvr}} - \mathsf{L}_g^{\text{cvr}}) - (\mathsf{W}_g^{\text{act}} - \mathsf{L}_g^{\text{act}})}{\mathsf{S}_g} \le 2\,,$$

where the actual vote totals here are with respect to $(\mathbf{B}_g^{\min})$. Then one has that

$$\begin{aligned}
\mathbb{E}\left[\mathsf{D}_\beta^{\mathcal{A}}\right] &\ge \mathbb{E}\left[\mathsf{D}_\beta^{\mathcal{A}}((\mathbf{B}_g^{\min})_{g=1}^{\nu}), (\mathbf{B}_g^{\min})_{g=1}^{\nu})\right] \\
&= \sum_{g=1}^{\nu} \frac{\mathsf{S}_g}{\mathsf{S}_\beta}\left(\frac{(\mathsf{W}_g^{\text{cvr}} - \mathsf{L}_g^{\text{cvr}}) - (\mathsf{W}_g^{\text{act}} - \mathsf{L}_g^{\text{act}})}{\mathsf{S}_g}\right) \\
&= \frac{(\mathsf{W}_\beta^{\text{cvr}} - \mathsf{L}_\beta^{\text{cvr}})}{\mathsf{S}_\beta} - \sum_{g=1}^{\nu} \frac{1}{\mathsf{S}_\beta}(\mathsf{W}_g^{\text{act}} - \mathsf{L}_g^{\text{act}}) \\
&= \frac{(\mathsf{W}_\beta^{\text{cvr}} - \mathsf{L}_\beta^{\text{cvr}})}{\mathsf{S}_\beta} - \sum_{g=1}^{\nu} \frac{1}{\mathsf{S}_\beta}\left(\sum_{\mathbf{b}\in\mathbf{B}_g^{\min}} (\mathsf{W}_{\mathbf{b}}^{\text{act}} - \mathsf{L}_{\mathbf{b}}^{\text{act}})\right) \\
&= \frac{1}{\mathsf{S}_\beta}\left((\mathsf{W}_\beta^{\text{cvr}} - \mathsf{L}_\beta^{\text{cvr}}) - \sum_{\mathbf{b}\in\mathbf{B}_\beta} (\mathsf{W}_{\mathbf{b}}^{\text{act}} - \mathsf{L}_{\mathbf{b}}^{\text{act}})\right) \\
&= \frac{1}{\mathsf{S}_\beta}\left((\mathsf{W}_\beta^{\text{tab}} - \mathsf{L}_\beta^{\text{tab}}) - (\mathsf{W}_\beta^{\text{act}} - \mathsf{L}_\beta^{\text{act}})\right) = \frac{\mathsf{D}_\beta}{\mathsf{S}_\beta}\,.
\end{aligned}$$

This completes the proof of Claim 6. □

Showing that this extends to the overall discrepancy follows exactly as in Claim 2:

Claim 7. *The expectation of* $\mathsf{D}^{\mathcal{A}}$ *over a single iteration satisfies*

$$\mathbb{E}[\mathsf{D}^{\mathcal{A}}] = \sum_\beta \left(\frac{\mathsf{S}_\beta^{\text{act}}}{\mathsf{S}^{\text{act}}} \cdot \mathbb{E}[\mathsf{D}_\beta^{\mathcal{A}}]\right) \ge \sum_\beta \left(\frac{\mathsf{S}_\beta^{\text{act}}}{\mathsf{S}^{\text{act}}} \cdot \frac{\mathsf{D}_\beta}{\mathsf{S}_\beta^{\text{act}}}\right) = \frac{\mathsf{D}}{\mathsf{S}}\,.$$

Furthermore, one can easily show that CVR transforms do not affect whether the auditor is risk-limiting as in Lemma 1.

Why group sizes don't have to be trusted. Our techniques for trusting an adversarial declaration of group sizes do not extend to an adversarial declaration of batch sizes which must still be counted or verified by a trustworthy component. There are two key differences in the group setting:

(1) Group size is only hand-counted if selected, and

(2) An iteration is marked with $\mathsf{D} = 2$ on any size mismatch.

In principle in an adaptive ballot comparison audit, one could add these two steps of first-hand counting the entire batch and rejecting if the true size is not equal to the declared size. However, we expect this to be drastically more work and likely to introduce more errors given the larger size of batches. One could use this technique for small batches, for example, ballots at a precinct that contain votes for valid write-in candidates are often tabulated separately.

8 Conclusion

This article presents a formal model of comparison risk-limiting audits and a new class of risk-limiting audits called adaptive comparison audits. The formal model allows us to answer critical procedural questions such as showing that the labeling of ballots need not be trusted. Adaptive comparison audits provide efficiency improvements as one only produces a CVR for batches selected for audit.

Acknowledgments

These results were developed as part of a collaboration with the Office of the CT Secretary of State and, additionally, were supported in part by a grant from that office.

Discussions with Mark Lindeman, Philip B. Stark, Lynn Garland, and anonymous reviewers improved the narrative and technical treatment. A.R. is supported by a research grant from IOG and NSF grant #1801487; B.F. is supported by NSF Grants #2232813 and #2141033 and the Office of Naval Research.

References

[1] T. Antonyan, T. Bromley, L. Michel, A. Russell, A. Shvartsman, and S. Stark. Computer assisted post election audits. State Certification Testing of Voting Systems National Conference, 2013.

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

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

[4] M. Blom, J. Budurushi, R. L. Rivest, P. B. Stark, P. J. Stuckey, V. Teague, and D. 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.

[5] J. Bretschneider, S. Flaherty, S. Goodman, M. Halvorson, R. Johnston, M. Lindeman, R. L. Rivest, P. Smith, and P. B. Stark. Risk-limiting post-election audits: Why and how, 2012.

[6] S. Checkoway, A. Sarwate, and H. Shacham. Single-ballot risk-limiting audits using convex optimization. In *Proceedings of the 2010 International Conference on Electronic Voting Technology/Workshop on Trustworthy Elections*, EVT/WOTE'10, page 1–13, USA, 2010. USENIX Association.

[7] Election Assistance Commission. Voluntary voting system guidelines, 2005.

[8] B. Fisch, D. Freund, and M. Naor. Physical zero-knowledge proofs of physical properties. In *Annual Cryptology Conference*, pages 313–336. Springer, 2014.

[9] B. C. for Justice and R. I. R. W. Group. Pilot implementation study of risk-limiting audit methods in the state of Rhode Island, 2019.

[10] L. Garland, N. McBurnett, J. Morrell, M. K. Schneider, and S. Singer. Principles and best practices for post-election tabulation audits, 2018.

[11] C. R. L. A. W. Group. Risk-limiting audit recommendations for connecticut, 2022.

[12] J. L. Hall, L. W. Miratrix, P. B. Stark, M. Briones, E. Ginnold, F. Oakley, M. Peaden, G. Pellerin, T. Stanionis, and T. Webber. Implementing risk-limiting post-election audits in California. In USENIX, editor, *2009 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE '09)*, Montreal, Canada, 2009. USENIX, USENIX.

[13] M. J. Higgins, R. L. Rivest, and P. B. Stark. Sharper p–values for stratified election audits. *Statistics, Politics, and Policy*, 2(1), 2011.

[14] M. Lindeman. Rhode island presidential risk-limiting audit, november 19-24, 2020 (brief report), 2020.

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

[16] M. Lindeman, P. B. Stark, and V. S. Yates. Bravo: Ballot-polling risk-limiting audits to verify outcomes. In *EVT/WOTE*, 2012.

[17] N. McBurnett. rlacalc - calculate statistical parameters for risk-limiting post-election audits. Jun 2022.

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

[19] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

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

[21] C. S. of State. 2020 general election risk-limiting audit discrepancy report, 2020.

[22] F. D. of State. Precinct-level election results, 2020.

[23] K. Ottoboni, M. Bernhard, J. A. Halderman, R. L. Rivest, and P. 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.

[24] K. Ottoboni, P. B. Stark, M. Lindeman, and N. McBurnett. Risk-limiting audits by stratified union-intersection tests of elections (SUITE). In *International Joint Conference on Electronic Voting*, pages 174–188. Springer, 2018.

[25] A. Russell, L. Michel, B. Fuller, M. Desmarais, J. Wohl, W. Reller, and S. Ahmad. Statistical analysis of post-election audit data for the november 03, 2020 presidential elections, 2021.

[26] M. Sridhar and R. L. Rivest. k-cut: A simple approximately-uniform method for sampling ballots in post-election audits. In *International Conference on Financial Cryptography and Data Security*, pages 242–256. Springer, 2020.

[27] P. Stark. Tools for comparison risk-limiting election audits, July 2022.

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

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

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

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

[32] P. 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.

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

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

[35] P. B. Stark. Alpha: Audit that learns from previously hand-audited ballots, 2022.

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

[37] D. Williams. *Probability with Martingales*. Cambridge University Press, 1991.

A Calculation of CVR Generation Percentages

In this section, we discuss the reported percentages of CVR generated with the adaptive ballot comparison method. We use Connecticut and Florida as case studies for three reasons: (1) elections are managed by each municipality with no voting equipment that is capable of producing CVRs with identifiers, (2) they represent different population sizes and number of precincts with Florida having approximately 6000 precincts and Connecticut having approximately 700, and (3) there is a large variance in municipality size. Furthermore, Connecticut uses a semi-automated transitive tabulator [1] to produce CVRs after the fact for some fraction of municipalities.

Our experimental framework adopts the Kaplan-Markov test presented in Definition 12 with $\gamma = 1.1$ and "a bit of rounding" [15]. In particular, ballot sample sizes were obtained from Neal McBurnett's tool, `rlacalc` [17], using the following data: (1) For Connecticut, the number of ballots used is 1,823,857, which is the number of votes cast in 2020 CT presidential election. (2) For Florida, the population of ballots is 11,067,456, which is the number of votes cast in the 2020 FL presidential election.

The number of precincts and voters for each town is pulled from the Connecticut Secretary of State's website and Florida's precinct-level election results. Ballots were split among towns by reserving 5% of votes as absentee and then splitting the remaining 95% evenly into the number of precincts in that town. This means that for a town the number of batches is always one more than the number of precincts. 100 simulations are conducted of the following experiment:

(1) Randomly distribute ballots to precincts according to their size.

(2) Randomly pick (with replacement) sample size ballots among all ballots. For all batches with a picked ballot mark the batch as picked

(3) Compute the total fraction of ballots in batches that are picked divided by the total number of ballots.

This last fraction is reported as the fraction of CVR generated. We report the average value of number of distinct picked batches and fraction of generated CVR are summarized in Table 1. The full simulation software is available at this Github repository. The full simulation code can also (1) distribute overstatement and understatement errors, and (2) compute risk and stopping time. However, this functionality was not used to create Table 1.

$\mathcal{T}_{\mathsf{Over},\mathrm{Force}}(\mathsf{S}_E^{\mathrm{act}}, T, \mathrm{cvr}_\beta)$:

(1) If cvr_β is not properly formed tuple according to Definition 5 output `Error`.
(2) While there exist two rows i and j where $i < j$ and both have identifier ι, replace the identifier in row j with an unused identifier in $\{\bot_t\}$.
(3) If $\mathsf{S}_\beta^{\mathrm{cvr}} \neq \mathsf{S}_\beta^{\mathrm{act}}$, then
    (a) While $\mathsf{S}_\beta^{\mathrm{cvr}} < \mathsf{S}_\beta^{\mathrm{act}}$ add a new row to cvr_β with an unused identifier in $\{\bot_t\}$ and zeroes for all votes.
    (b) While $\mathsf{S}_\beta^{\mathrm{cvr}} > \mathsf{S}_\beta^{\mathrm{act}}$ remove the last row of cvr_β.
(4) Place all rows with $\iota \in \{\bot_t\}$ at the end of the CVR.
(5) For all ι where $\mathsf{W}_{r_\iota} = 1, \mathsf{L}_{r_\iota} = 1$ set $\mathsf{W}_{r_\iota} = 0$.
(6) If $\mathsf{W}_\beta^{\mathrm{cvr}} \neq \mathsf{W}_\beta^{\mathrm{tab}}$.
    (a) While $\mathsf{W}_\beta^{\mathrm{cvr}} < \mathsf{W}_\beta^{\mathrm{tab}}$
        i. While $\mathsf{L}_\beta^{\mathrm{cvr}} > \mathsf{L}_\beta^{\mathrm{tab}}$, find the last row r such that $\mathsf{L}_r = 1$ set $\mathsf{W}_r = 1, \mathsf{L}_r = 0$.
        ii. Find the last row r such that $\mathsf{W}_r = 0, \mathsf{L}_r = 0$ set $\mathsf{W}_r = 1, \mathsf{L}_r = 0$.
    (b) While $\mathsf{W}_\beta^{\mathrm{cvr}} > \mathsf{W}_\beta^{\mathrm{tab}}$
        i. While $\mathsf{L}_\beta^{\mathrm{cvr}} < \mathsf{L}_\beta^{\mathrm{tab}}$, find the last row r such that $\mathsf{W}_r = 1$ set $\mathsf{W}_r = 0, \mathsf{L}_r = 1$.
        ii. Find the last row r such that $\mathsf{W}_r = 0, \mathsf{L}_r = 0$ set $\mathsf{W}_r = 0, \mathsf{L}_r = 1$.
(7) If $\mathsf{L}_\beta^{\mathrm{cvr}} \neq \mathsf{L}_\beta^{\mathrm{tab}}$. Set $i := \mathsf{S}_\beta^{\mathrm{cvr}}$.
    (a) While $\mathsf{L}_\beta^{\mathrm{cvr}} < \mathsf{L}_\beta^{\mathrm{tab}}$: find the last row r such that $\mathsf{W}_r = 0, \mathsf{L}_r = 0$ set $\mathsf{W}_r = 1, \mathsf{L}_r = 0$.
    (b) While $\mathsf{L}_\beta^{\mathrm{cvr}} > \mathsf{L}_\beta^{\mathrm{tab}}$: find the last row r such that $\mathsf{W}_r = 0, \mathsf{L}_r = 1$ set $\mathsf{W}_r = 0, \mathsf{L}_r = 0$.

Figure 9: CVR transform function that ensures consistency and no overvotes.

B Auditor and transform without overvotes

In Section 3 we presented an auditor that allows "overvotes" [15]. An overvote means that a CVR row or ballot that has marks for both the winner and loser is considered valid. It is also possible for $\mathcal{T}_{\mathrm{Force}}$ to create overvotes.

Here we present an alternative auditor and transform function that does not allow or create overvotes. The auditor differs from Figure 3 in exactly two places:

(1) Step (2)b which sets $\mathsf{W}_{\mathsf{iter}}^{\mathrm{tab}} := \min(\mathsf{W}_{\mathsf{iter}}^{\mathrm{tab}}, \mathsf{S}_{\mathsf{iter}}^{\mathrm{act}})$ is moved after Step (2)c and replaced with $\mathsf{W}_{\mathsf{iter}}^{\mathrm{tab}} := \min(\mathsf{W}_{\mathsf{iter}}^{\mathrm{tab}}, \mathsf{S}_{\mathsf{iter}}^{\mathrm{act}} - \mathsf{L}_{\mathsf{iter}}^{\mathrm{tab}})$. This ensures that the sum of $\mathsf{W}_{\mathsf{iter}}^{\mathrm{tab}} + \mathsf{L}_{\mathsf{iter}}^{\mathrm{tab}} \leq \mathsf{S}_{\mathsf{iter}}^{\mathrm{act}}$.

(2) A check is added to `CheckConsistent` as follows: If there exists a row with identifier ι in cvr_β such that $\mathsf{W}_\iota = 1$ and $\mathsf{L}_\iota = 1$ return `Error`. This step is added before the step that returns `OK`. Let $\texttt{CheckConsistent}_{\mathsf{Over}}$ denote the modified procedure.

The main changes are in the transform function shown in Figure 9 here the transform never creates a row where both winner and loser are 1. Differences are highlighted in Blue.

Claim 8. *Figure 9 always completes and outputs a CVR such that* $\texttt{CheckConsistent}_{\mathsf{Over}}$ *returns* `OK`.

Proof. Importantly, after Step (3) in the modified Figure 3 it is true that for all batches k,

$$\mathsf{W}_k^{\mathrm{tab}} + \mathsf{L}_k^{\mathrm{tab}} \leq \mathsf{S}_k^{\mathrm{act}}.$$

Furthermore, after Step (3) in Figure 9 it is true that $\mathsf{S}_k^{\mathrm{act}} = \mathsf{S}_k^{\mathrm{tab}} = \mathsf{S}_k^{\mathrm{cvr}}$. We now show that Steps (6) and (7) in Figure 9 eventually lead to a CVR consistent with the tabulation without overvotes. At each iteration of Step (6) one of four conditions must be true:

(1) $\mathsf{W}_k^{\mathrm{cvr}} = \mathsf{W}_k^{\mathrm{tab}}$,

(2) $\mathsf{W}_k^{\mathrm{cvr}} > \mathsf{W}_k^{\mathrm{tab}}$,

(3) $\mathsf{L}_k^{\mathrm{cvr}} > \mathsf{L}_k^{\mathrm{tab}}$, or

(4) There is a row in the CVR with identifier ι such that $\mathsf{W}_\iota^{\mathrm{cvr}} = 0, \mathsf{L}_\iota^{\mathrm{cvr}} = 0$.

To see that the four cases are complete, if the first three cases are not true then $\mathsf{W}_k^{\mathrm{cvr}} < \mathsf{W}_k^{\mathrm{tab}}, \mathsf{L}_k^{\mathrm{cvr}} \le \mathsf{L}_k^{\mathrm{tab}}$. This means that

$$\mathsf{W}_k^{\mathrm{cvr}} + \mathsf{L}_k^{\mathrm{cvr}} < \mathsf{W}_k^{\mathrm{tab}} + \mathsf{L}_k^{\mathrm{tab}} \le \mathsf{S}_k^{\mathrm{tab}} = \mathsf{S}_k^{\mathrm{cvr}}.$$

That is, there are fewer than $\mathsf{S}_k^{\mathrm{cvr}}$ 1s in the CVR and there must be some row with both winner and loser set to 0.

In each of the above cases, Step (6) either finds a row to change or completes. Furthermore, note that $\mathsf{W}_k^{\mathrm{cvr}}$ monotonically approaches $\mathsf{W}_k^{\mathrm{tab}}$ so it only requires at most $|\mathsf{W}_k^{\mathrm{tab}} - \mathsf{W}_k^{\mathrm{cvr}}|$ steps to complete.

For Step (7) note that in addition to the above properties it now holds that $\mathsf{W}_k^{\mathrm{cvr}} = \mathsf{W}_k^{\mathrm{tab}}$. Of course, if $\mathsf{L}_k^{\mathrm{cvr}} > \mathsf{L}_k^{\mathrm{tab}}$ one can always change a row with $\mathsf{L}_k^{\mathrm{cvr}} = 1$ and $\mathsf{W}^{\mathrm{cvr}} = 0$ to be both 0. Now suppose that $\mathsf{L}_k^{\mathrm{cvr}} < \mathsf{L}_k^{\mathrm{tab}}$, then it holds that

$$\mathsf{W}_k^{\mathrm{cvr}} + \mathsf{L}_k^{\mathrm{cvr}} = \mathsf{W}_k^{\mathrm{tab}} + \mathsf{L}_k^{\mathrm{cvr}} < \mathsf{W}_k^{\mathrm{tab}} + \mathsf{L}_k^{\mathrm{tab}} \le \mathsf{S}_k^{\mathrm{tab}} = \mathsf{S}_k^{\mathrm{cvr}}.$$

That is, there are fewer $\mathsf{S}_k^{\mathrm{cvr}}$ 1s in the CVR and there must be some row with both winner and loser set to 0. This completes the proof of Claim 8. □