Simulations of Ballot Polling Risk-Limiting Audits

No Author Given

No Institute Given

Abstract. In this paper we present simulation results comparing the risk, stopping probability and number of ballots required over multiple rounds of ballot polling risk limiting audits (RLAs) Minerva, Selection-Ordered (SO) Bravo and End-of-Round (EoR) Bravo.
Bravo is the most commonly used ballot polling RLA and requires the smallest expected number of ballots when ballots are drawn one at a time and the (true) underlying election is as announced. In real audits, election officials draw multiple ballots at a time, and Bravo rules may be implemented as SO Bravo or EoR Bravo, with neither being optimal for large first round sizes. Minerva was recently proposed for use when the first-round stopping probability is large. It requires fewer first-round ballots, on average, than both SO Bravo and EoR Bravo for first round stopping probabilities of 0.9 when the (true) underlying election is as announced.
An open question, however, is how Minerva compares to SO Bravo and EoR Bravo over multiple rounds, and for lower stopping probabilities. In this paper, we present results from simulations of multiple round audits. We examine both: first-round stopping probabilities of .9, a common choice among election officials, and .25, which would be more favorable to Bravo.
We show that the simulation results are consistent with predictions of the R2B2 open-source library for ballot polling audits. Because the Minerva round schedule needs to be predetermined, we observe that the simulator is useful as a tool for understanding the behavior of Minerva with the fixed round schedule. Most importantly, we observe that both Bravo audits are more conservative than needed, while Minerva audits stop with fewer ballots, for both first round stopping probabilities, though using Minerva is not as advantageous for the smaller first round stopping probability.

Keywords: risk-limiting audit (RLA) · ballot polling audit · evidence-based elections · statistical election audit

1 Introduction

The literature contains numerous descriptions of vulnerabilities in deployed voting systems, and it is not possible to be certain that any system, however well-designed, will perform as expected in all instances. For this reason, *evidence-based elections* [14] aim to produce trustworthy and compelling evidence of the

correctness of election outcomes, enabling the detection of problems with high probability. One way to implement an evidence-based election is to use a well-curated voter-verified paper trail, compliance audits, and a rigorous tabulation audit of the election outcome, known as a risk-limiting audit (RLA) [8]. An RLA is an audit which guarantees that the probability of concluding that an election outcome is correct, given that it is not, is below a pre-determined value known as the risk limit of the audit, independent of the true, unknown vote distribution of the underlying election. Over a dozen states have seriously explored the use of RLAs—some have pilot programs, some allow RLAs to satisfy a general audit requirement and some have RLAs in statute.

This paper provides insight into the main approach to ballot polling RLAs, the Bravo audit [9], and the newer Minerva [19] ballot polling RLA, through the presentation of simulation results. While some properties of the two audits may be theoretically derived, for other properties theoretical results are not available. This paper examines the number of ballots drawn over multiple rounds of both audits, the related probabilities of stopping if the election is as announced, and the maximum risk of the audit.

1.1 Background

Of the multiple types of RLAs, this paper focuses on ballot-polling RLAs, which require a large number of ballots but do not rely on any special features of the election technology. In the general ballot-polling RLA, a number of ballots are drawn and tallied in what is termed a *round* of ballots [19]. A statistical measure is then computed to determine whether there is sufficient evidence to declare the election outcome correct within the pre-determined risk limit. Because the decision is made after drawing a round of ballots, the audit is termed a *round-by-round (R2)* audit. The special case when round size is one—that is, stopping decisions are made after each ballot draw—is a *ballot-by-ballot (B2)* audit.

The Bravo audit is designed for use as a B2 audit: it requires the smallest expected number of ballots when the true tally of the underlying election is as announced, and stopping decisions are made after each ballot draw. In practice, election officials draw many ballots at once, and the Bravo stopping rule needs to be modified for use in an R2 audit that is not B2. There are two obvious approaches. The B2 stopping condition can be applied once at the end of each round: End-of Round (EoR) Bravo. Alternatively, the order of ballots in the sample can be tracked by election officials and the B2 Bravo stopping condition can be applied retroactively after each ballot drawn: Selection-Ordered (SO) Bravo. SO Bravo requires fewer ballots on average than EoR Bravo but requires the work of tracking the order of ballots rather than just their tally.

Minerva was designed for R2 audits and applies its stopping rule once for each round. Thus it does not require the tracking of ballots that SO Bravo does. Zagórski *et al.* [19] prove that Minerva is a risk-limiting audit and requires fewer ballots to be sampled than EoR Bravo when an audit is performed in rounds, the two audits have the same pre-determined (before any ballots are drawn) round schedule and the underlying election is as announced. They also present

first-round simulations which show that MINERVA draws fewer ballots than SO BRAVO in the first round for first round sizes with a large probability of stopping when the (true) underlying election is as announced.

There are no results, either theoretical or based on simulations, regarding the number of ballots drawn over multiple rounds in a MINERVA audit with a pre-determined schedule. Because BRAVO does not need to work on a pre-determined round schedule, it can optimize the size of the next round based on the sample drawn so far. Thus an open question is whether the constraint of a predetermined round schedule limits the efficacy of MINERVA in future rounds, and there is no literature comparing the number of ballots drawn by MINERVA and SO BRAVO over multiple rounds. Note that the Average Sample Number (ASN) computations for BRAVO [9] apply only for B2 audits and are especially misleading as estimates of the number of ballots drawn over multiple rounds when first round sizes are large.

Both BRAVO and MINERVA have been integrated into election audit software *Arlo* [16], and, as such, available for use in real election audits. Both have been used in real election audits. For this reason, it is very important to understand their properties over multiple rounds.

1.2 Our Results

We show the following:

1. Even when the first round stopping probability is as small as 0.25, the number of ballots required for MINERVA is smaller than that required by SO BRAVO and EoR BRAVO is by a fraction of . Compare this to:
2. For a first round stopping probability of 0.9, when consequent MINERVA rounds are the same size (multiplying factor 1), consequent conditional stopping probabilities are about 0.75 and 0.74 respectively for rounds two and three. When the multiplying factor is 1.5, the conditional stopping probabilities for rounds two and three are 0.91 and 0.83 respectively.
3. Our simulations may be used to study the stopping probabilities and maximum risks of the different audits for different stopping probabilities for the first round, we provide examples.

1.3 Organization

Section 2 describes related work and 3 our open source software. The experiments we performed are described in section 4 and sections 5 and 6 present our results. Section 7 has our conclusions.

2 Related work

The BRAVO audit [9] is a well-known ballot polling audit which has been used in numerous pilot and real audits. When used to audit a two-candidate election,

it is an instance of Wald's sequential probability ratio test (SPRT) [17], and inherits the SPRT property of being the most efficient test (requiring the smallest expected number of ballots) if the election is as announced. The model for Bravo and the SPRT is, however, that of a sequential audit: a sample of size one is drawn, and a decision of whether to stop the audit or not is taken. Real election audits invest in drawing a large number of ballots before making the decision. It is possible to apply Bravo to the sequence of ballots if the sequential order is retained. This is not, however, the most efficient possible use of the drawn sample because information in consequent ballots is ignored when applying Bravo to ballots that were drawn earlier in the sample.

We do know a great deal about the properties of Bravo. The risk limiting property of Bravo follows from the similar property of the SPRT. Stopping probabilities for Bravo may be computed as described by Zagórski *et al.*[19,18]. The results of the computations match simulation results reported by Lindeman *et al.* [9, Table 1].

The Minerva audit [19,18] was developed for large first round sizes which enable election officials to be done in one round with large probability. It uses information from the entire sample, and has been proven to be risk limiting when the round schedule for the audit is determined before the audit begins. That is, information about the actual ballots drawn in the first round cannot be used to determine future round sizes. First-round sizes for a 0.9 stopping probability when the election is as announced have been computed for a wide range of margins and shown to be smaller than those for both EoR and SO Bravo. First round simulations of Minerva [18] demonstrate that its first-round properties—regarding the probabilities of stopping when the underlying election is tied and when it is as announced—are as predicted for first round sizes with stopping probability 0.9.

Ballot polling audit simulations have been used to familiarize election officials and the public with the approach [13]. McLaughlin and Stark [11,10] compare the workload for the Canvass Audits by Sampling and Testing (CAST) and Kaplan-Markov (KM) audits using simulations. Blom *et al.* demonstrate the efficiency of their ballot polling approach to audit instant runoff voting (IRV) using simulations [6]. Huang *et al.* present a framework generalizing a number of ballot polling audits and compare their performance (round sizes and stopping probabilities) using simulations [7]. This work was prior to the development of Minerva, and focuses on the comparison between Bayesian audits [12] and Bravo, essentially studying the impact of the prior of the Bayesian RLA.

3 Software

In this section we describe the software implementing ballot polling audits, termed the R2B2 library, and the simulator software used for this research. All the software is released as open source under the MIT License.

3.1 R2B2 Library

The R2B2 Python library [5] provides a framework for the exploration of round-by-round and ballot-by-ballot RLAs. The goal in designing R2B2 is two fold:

1. Provide an elegant Python library which can be easily imported and used in any other code base.
2. Provide an interactive set of tools which can be utilized 'out-of-the-box' for experimenting with and learning about risk-limiting audits.

Design The high-level design of R2B2 is an object-oriented view of election audits. The three main object classes, `Election`, `Contest`, and `Audit`, serve to group data into logically independent structures.

The `Election` contains the information that comprises an entire election, most importantly, the total number of ballots cast in the election and the list of `Contest`s from the election. At the moment `Election` does not offer functionality beyond grouping `Contest`s.

The `Contest` contains the information related to a single contest such as the ballots cast in that contest, the candidates, the type of contest, and the reported tally. Providing a structure to hold this information independent of any particular audit is especially useful for exploratory work.

The `Audit` contains information related to the audit parameters for a single contest, such as the risk limit, sampling method, and `Contest` to audit. It is important to note the `Audit` is an Abstract Base Class upon which specific RLAs are built. It only contains the parameters and attributes common to the RLAs of this paper and provides a set of methods that can be called by any audit implementation. The functionality of `Audit` can be divided into two basic groups: *interactive* and *bulk*.

The interactive implementation allows users to execute an audit step-by-step as it might progress during a live election audit through the following:

- The `run()` method begins an interactive audit executing where users are prompted for round sizes and the counts of winner ballots found in the sample and in return are given information about the current state of the audit and whether the stopping condition(s) have been met.
- Two distributions representing the null and alternative hypotheses are maintained and allow for computation of the audits per-round risk and stopping probability schedules.
- Before each round, the audit will recommend possible next round sizes given different criteria, such as a set of desired stopping probabilities.

The bulk implementations allows users to generate a larger set of data from an audit such as:

- A set of stopping conditions given a set of round sizes.
- A set of risk levels given a set of round size and winner ballots pairs.
- A list of all stopping conditions from the minimum to the maximum round size.

Usage R2B2 makes understanding and exploring election audits simple for the user with no Python knowledge while simultaneously providing a comprehensive set of tools for the experienced Python developer.

Using R2B2 is as simple as using any other Python library: simply import the library and all of the functionality is at your finger tips. Not only does this allow users to write their own Python scripts for exploring RLAs, it also allows R2B2 to be plugged in to any other Python library. See the following Jupyter Notebooks for information on the usage of R2B2: Basic Usage [3], Generating Graphs [4].

R2B2 also provides a significant amount of functionality 'out-of-the-box' for educational or exploratory use. For those who wish to learn about RLAs without having to write any code themselves, R2B2 provides a command line tool for both interactive auditing and generating audit results and statistics for larger data sets.

3.2 Simulation Software

As described above R2B2 has implementations of several ballot polling risk-limiting audits as well as a simulator, all written in Python. For each of these audits, the software can compute the stopping condition for a given sample and estimates of the next round size to achieve a desired stopping probability. For a given audit and random seed, the simulator draws random samples, with replacement, using a pseudorandom number generator, given the number of votes for each candidate, and the number of invalid votes, in the underlying election (these need not be chosen to be as announced).

When the number of candidates is more than two, the audit is carried out pairwise for each candidate pair, and votes for all other candidates are considered invalid votes.

After drawing a simulated sample of ballots, the simulator evaluates the given audit's stopping condition for this sample. If the audit stops, the simulation stops, and if the audit continues, the simulation draws another round. The abstract simulator class does not prescribe any one method for choosing round sizes. We implement several classes to support various round size choices: round sizes from an estimate to achieve a desired probability of stopping, predetermined round sizes, and pseudorandomly-generated round sizes.

3.3 Testing

The R2B2 software is used to compute stopping conditions and next round estimates. It is intended for use by us and other researchers, and designed for this purpose. We have also independently implemented all the functionality in matlab [2] (the two codebases are written by different individuals) and have extensively checked the results of the two codebases. Additionally, for use in regular election audits by election officials, we have written an add-on [1] to the *Arlo* risk-limiting audit software, the results of which have also been extensively checked against the other two codebases.

4 Experiments

In this section, we motivate and describe the experiments. We consider a two candidate plurality contest, and assume that ballots are sampled with replacement, as is common in the literature.

We first present relevant definitions.

Definition 1. *An audit $\mathcal{A}$ takes a sample of ballots X as input and gives one of the following decisions*

1. *Correct: the audit is complete*
2. *Uncertain: continue the audit*

All of the audits discussed in this paper are modeled as binary hypothesis tests. Under the alternate hypothesis, H_a, the announced outcome is correct. That is, the true underlying ballot distribution is given by the announced ballot tallies. Under the null hypothesis, H_0, the true outcome is a tie (or the announced winner lost by one vote, and the number of ballots is large enough that the probability of drawing a ballot for the winner is that of drawing one for the loser).

The maximum risk of an audit is the probability that an audit stops, given that the underlying election is a tie. (Vora show that this is the maximum risk [15].)

Definition 2 (Risk). *The maximum risk R of an audit $\mathcal{A}$ is*

$$R(\mathcal{A}) = \Pr[\mathcal{A}(X) = Correct \mid H_0]$$

This leads us to the following definition of an α-RLA.

Definition 3 (Risk Limiting Audit (α-RLA)). *An audit $\mathcal{A}$ is a Risk Limiting Audit with risk limit α iff*

$$R(\mathcal{A}) \leq \alpha.$$

We present measures of stopping probability in the j^{th} round of the audit, given that the underlying election is as announced.

Definition 4 (Stopping Probability). *The stopping probability S_j of an audit $\mathcal{A}$ in round j is*

$$S_j(\mathcal{A}) = \Pr[\mathcal{A}(X) = Correct \text{ in round } j \ \wedge \mathcal{A}(X) \neq Correct \text{ previously} \mid H_a]$$

Experimentally, using our simulations, S_j would be estimated by the fraction of audits that stop in round j.

Note that $\sum_j S_j(\mathcal{A}) = 1$. We can also consider the cumulative stopping probability:

Definition 5 (Cumulative Stopping Probability). *The cumulative stopping probability C_j of an audit $\mathcal{A}$ in round j is*

$$C_j(\mathcal{A}) = \sum_{i=1}^{j} S_j$$

Experimentally, using our simulations, C_j would be estimated by the fraction of audits that stop in or before round j.

Finally, we are also interested in the probability that an audit will stop in round j given that it did not stop earlier:

Definition 6 (Conditional Stopping Probability). *The conditional stopping probability of an audit $\mathcal{A}$ in round j is*

$$\chi_j(\mathcal{A}) = \Pr[\mathcal{A}(X) = \textit{Correct in round } j \mid H_a \wedge \mathcal{A}(X) \neq \textit{Correct previously}]$$

Experimentally, using our simulations, χ_j would be estimated by the ratio of the audits that stop in round j to those that "entered" round j, i.e. those that did not stop before round j.

We simulated audits for a risk limit of 0.1 using margins from the 2020 Presidential election, limiting ourselves to pairwise margins for the two main candidates of 0.05 or larger. Round sizes increase roughly proportional to the inverse square of the margin, so smaller margins are computationally much more expensive to simulate. For each of these states, we simulated $10,000 = 10^4$ audits assuming the underlying election was as announced (H_a), and an additional $10,000 = 10^4$ audits assuming the underlying election was a tie (H_0).

We ran simulations for: (a) .9 probability of stopping in the first round, enabling election officials to be done in the first round with very high probability if the election is as announced and (b) .25 probability of stopping in the first round which is more favorable to Bravo. We ran our simulations for up to five rounds.

For rounds after the first one, we chose the round schedule as follows. For both versions of Bravo, we chose a single round schedule: each round size has the same conditional stopping probability as the first one. As the proof of the risk-limiting property of Minerva assumes that its round schedule is determined before any ballots are drawn, we could not use this approach for Minerva round sizes. Instead, we chose to compare two fixed round schedules for Minerva: one where the additional number of ballots drawn in a round is the same as in the previous round (multiplying factor of 1.0) and the second where the multiplying factor is 1.5. We consider the case of drawing samples of the same size because it may reflect a practical way to continue an audit; if election officials have selected some first round size within reasonable logistical bounds, drawing the same number of ballots in subsequent rounds may be practical. We also consider round sizes with samples increasing by a multiple of 1.5 because this version is integrated into *Arlo*, and the multiplying factor was chosen as it roughly ensures a .9 conditional stopping probability in the second round for a first round stopping probability of .9.

5 Stopping Probability and Risk

5.1 Stopping Probability as a Function of Round and Margin

For both SO and EoR Bravo simulations, our software estimated round sizes that would give $\chi_j(\mathcal{A}) = 0.9$ and used those for the simulations. In Figure 1, we display the proportion of EoR Bravo audits that stopped in the j^{th} round to all audits which had not stopped before the j^{th} round, for $j = 1, 2, 3$. Though we carried out the simulations for 5 rounds we show only the first three rounds of the simulations because very few audits, $(.1)^{j-1} \cdot (10^4)$ on average, make it to the j^{th} round for $j \geq 4$. In Figure 2, we display the same proportions for SO Bravo audits. In both cases, these proportions are estimates of the true value of $\chi_j(\mathcal{A})$ for $j = 1, 2, 3$ as a function of margin. We see that, especially in earlier rounds for which the values are more representative of true audit behavior because fewer simulated audits have stopped, our round size predictions are accurate (the proportions are close to 0.9).

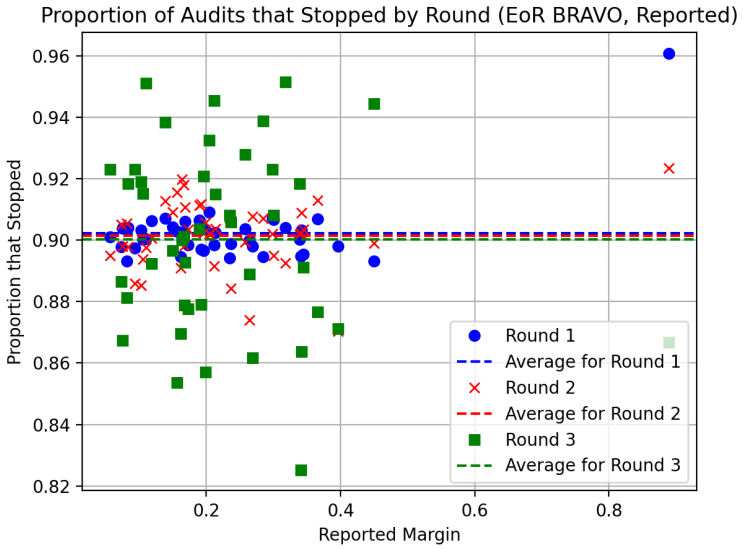

Fig. 1. This plot shows, for each state margin, when the underlying election is as announced, the number of EoR Bravo audits that stopped in the j^{th} round, as a fraction of all EoR Bravo audits which had not yet stopped before the j^{th} round for $j = 1, 2, 3$ and $S_1 = 0.9$.

Figure 3 and Figure 5 show the same proportions for Minerva round multipliers of 1.0 and 1.5 respectively. We see that the first round size estimates were fairly accurate, with first round stopping probabilities being very close to .9. For subsequent rounds, the multipliers of 1.0 achieved smaller stopping probabilities, as it was not chosen so as to obtain $\chi_j(\mathcal{A}) = 0.9$. The 1.5 multiplier is a good

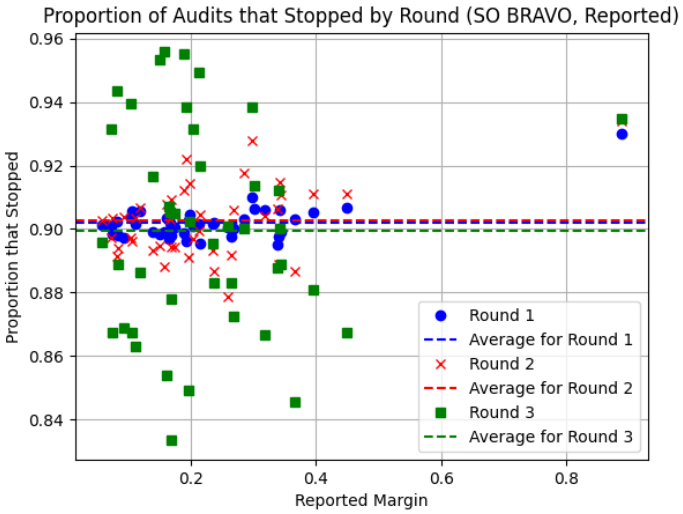

Fig. 2. This plot shows, for each state margin, when the underlying election is as announced, the number of SO Bravo audits that stopped in the j^{th} round, as a fraction of all EoR Bravo audits which had not yet stopped before the j^{th} round for $j = 1, 2, 3$ and $S_1 = 0.9$.

estimate for $j = 2$, but the stopping probability for $j = 3$ is slightly smaller than 0.9. Note that we chose a simple multiplier for future rounds, but one could make more accurate round size estimates before the audit begins.

Finally, we can perform a similar study for $S_1 = 0.25$. See **??** for an example, Minerva with round mutiplier 1.5.

5.2 Maximum Risk as a Function of Round and Margin

We also study the proportion of audits that stopped when the underlying election was a tie. This proportion should approach a value less than the risk limit, 0.1, as more audits are performed.

We observe that the risk of EoR Bravo is roughly an order of magnitude less than the risk limit. These results are as expected, because EoR Bravo is known to be too conservative [19].

In Figure 6 we show only the results for the 13 states for which our simulations with an underlying tied election have completed. To estimate the next round size that achieves a desired stopping probability, the SO Bravo software generates the probability distribution on the number of ballots in the sample ballot by ballot (see [19]) since the stopping condition needs to be evaluated for each individual ballot drawn. Because the underlying tied election causes audits to move on to larger rounds, the simulations are computationally expensive. SO Bravo is proven to be a Risk-Limiting Audit, and we observe in Figure 6, that

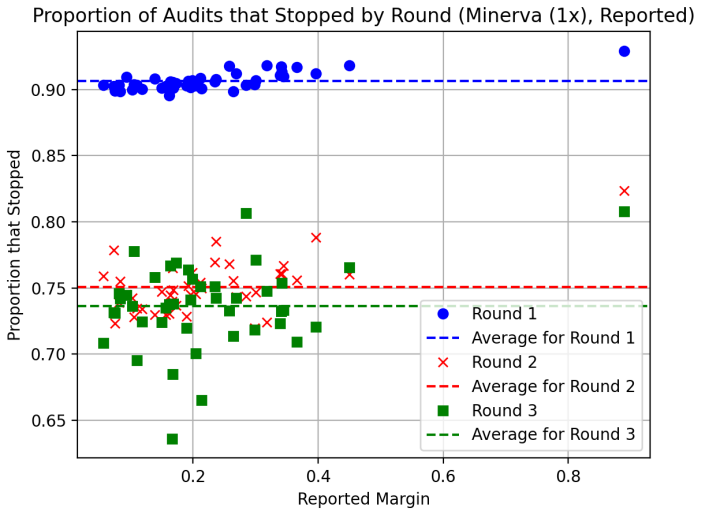

Fig. 3. This plot shows, for each state margin, when the underlying election is as announced, the number of MINERVA audits that stopped in the j^{th} round, as a fraction of all MINERVA audits which had not yet stopped before the j^{th} round for $j = 1, 2, 3$, round size multiple of 1.0 and $S_1 = 0.9$.

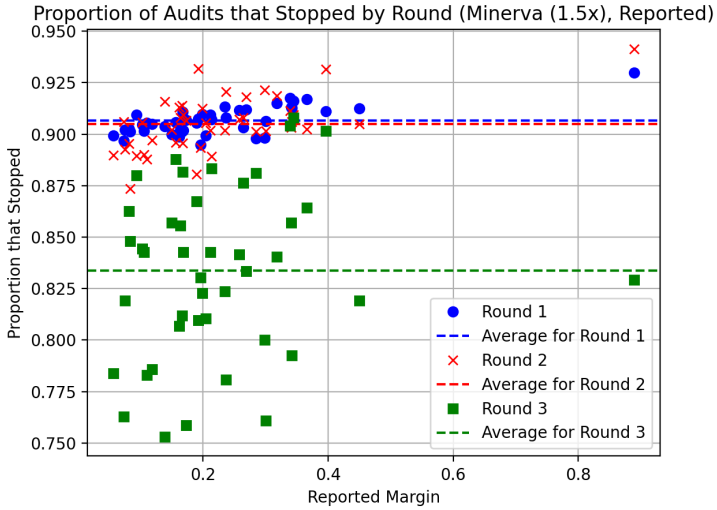

Fig. 4. This plot shows, for each state margin, when the underlying election is as announced, the number of MINERVA audits that stopped in the j^{th} round, as a fraction of all MINERVA audits which had not yet stopped before the j^{th} round for $j = 1, 2, 3$, round size multiple of 1.5 and $S_1 = 0.9$.

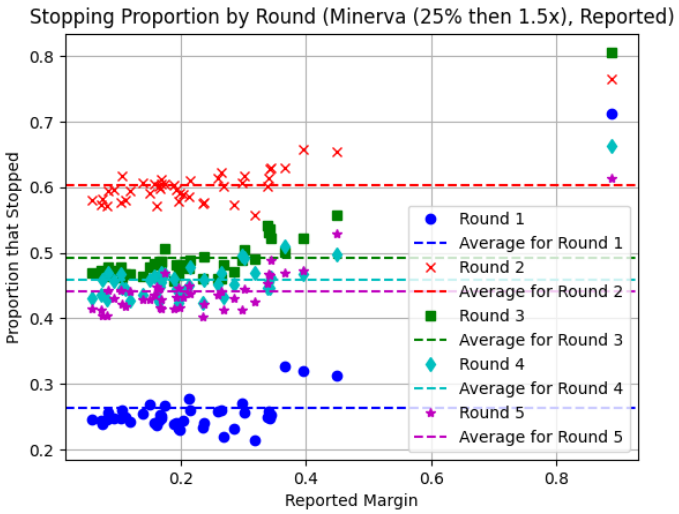

Fig. 5. This plot shows, for each state margin, when the underlying election is as announced, the number of MINERVA audits that stopped in the j^{th} round, as a fraction of all MINERVA audits which had not yet stopped before the j^{th} round for $j = 1, 2, 3$, round size multiple of 1.5 and $S_1 = 0.25$.

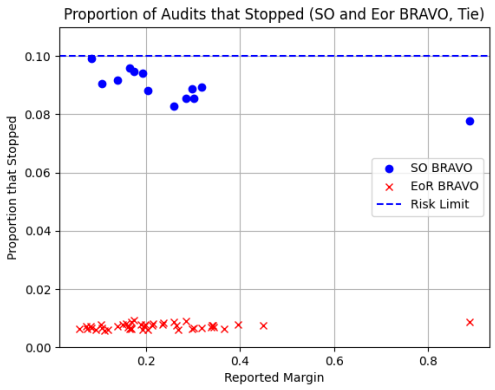

Fig. 6. This plot shows the fraction of EoR BRAVO audits (all states with margins at least 0.05) and SO BRAVO audits (the 13 states for which our simulations are complete so far) that stopped in any of the 5 rounds when the underlying election was a tie.

the risk of SO Bravo is much nearer the risk limit than that of EoR Bravo, as expected.

Figures 7 and Figure 8 show that fewer than 0.1 of the audits stopped when the underlying election was a tie, for round multiples of 1.0 and 1.5 respectively, as would be expected for an RLA with risk limit 0.1. Unlike EOR Bravo, the experimental risks here are much closer to the risk limit, showing that Minerva stops on average with a less conservative risk; Minerva is sharper.

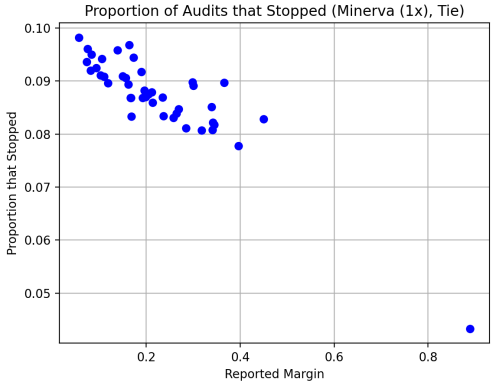

Fig. 7. This plot shows, for each state margin, the fraction of Minerva audits with a round size multiple of 1.0 that stopped in any of the 5 rounds when the underlying election was a tie.

6 Number of Ballots

In this section we present our data on the expected number of ballots drawn as the number of rounds increases, and on the fraction of audits that stop (an estimate of cumulative stopping probability, C_j) for the states of Texas, Missouri and Massachusetts, with margins of 0.057, 0.157 and 0.342 respectively. Interestingly, we observe that the advantage Minerva has for a first round size with stopping probability $S_1 = 0.9$ is retained for $S_1 = 0.25$.

We observe that the behavior of both Minerva audits is similar, and that the plot for SO Bravo is to the right (more ballots) and below (lower probability of stopping) those for Minerva, even for a stopping probability as low as 0.25. We observe that the plot for EoR Bravo shows the worst performance, which is not surprising. We observe similar behavior across margins (see Figures 10 and 11), though the improvement due to Minerva reduces margins get larger. We see also that the behavior is similar to that seen for $S_1 = 0.9$ (see Figure 12).

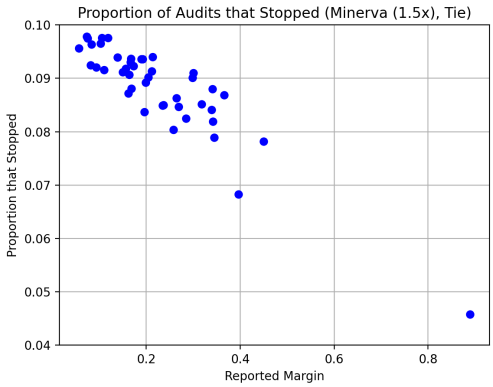

Fig. 8. This plot shows, for each state margin, the fraction of MINERVA audits with a round size multiple of 1.5 that stopped in any of the 5 rounds when the underlying election was a tie.

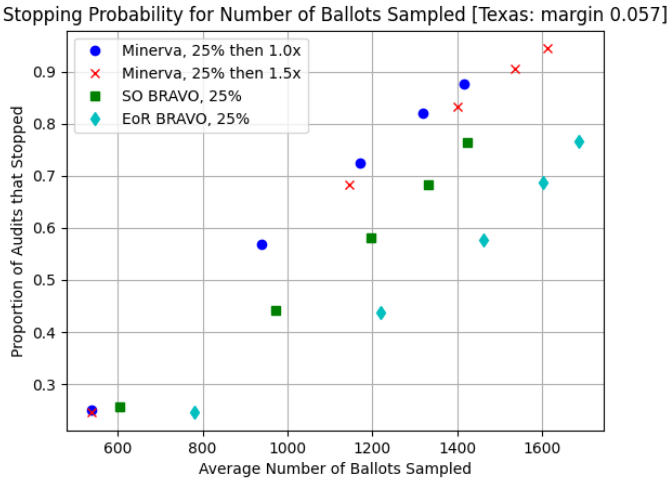

Fig. 9. This plot shows the cumulative fraction of audits that stopped as a function of average number of sampled ballots for all four audits we studied, for the state of Texas, margin 0.057, and first round stopping probability $S_1 = 0.25$.

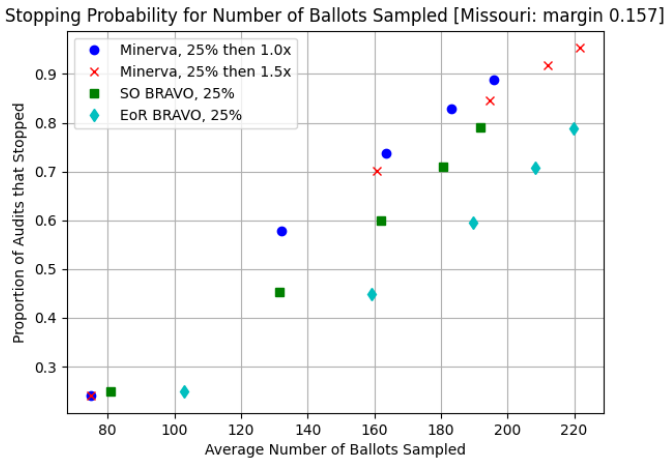

Fig. 10. This plot shows the cumulative fraction of audits that stopped as a function of average number of sampled ballots for all four audits we studied, for the state of Missouri, margin 0.157, and first round stopping probability $S_1 = 0.25$.

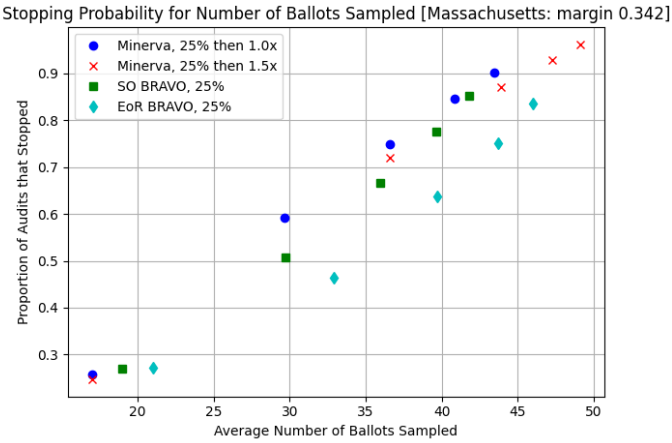

Fig. 11. This plot shows the cumulative fraction of audits that stopped as a function of average number of sampled ballots for all four audits we studied, for the state of Massachusetts, margin 0.342, and first round stopping probability $S_1 = 0.25$.

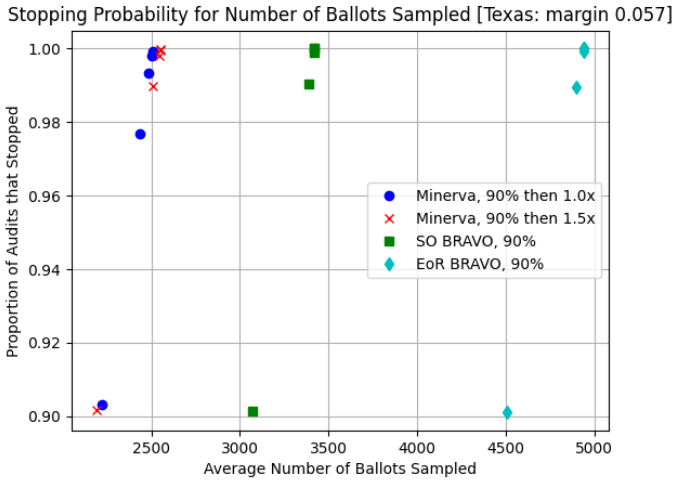

Fig. 12. This plot shows the cumulative fraction of audits that stopped as a function of average number of sampled ballots for all four audits we studied, for the state of Texas, margin 0.057, and first round stopping probability $S_1 = 0.9$.

7 Conclusions and Future Work

References

1. Anonymized: Athena - risk limiting audit (round-by-round), https://github.com/xxxx
2. Anonymized: brla_explore, https://github.com/xxxx
3. Anonymized: R2B2 Basics, https://github.com/xxxx
4. Anonymized: R2B2 Comparing Audits with Graphs, https://github.com/xxxx
5. Anonymized: The R2B2 (Round-by-Round, Ballot-by-Ballot) library, https://github.com/xxxx
6. Blom, M.L., Stuckey, P.J., Teague, V.J.: Ballot-polling risk limiting audits for IRV elections. In: Krimmer, R., Volkamer, M., Cortier, V., Goré, R., Hapsara, M., Serdült, U., Duenas-Cid, D. (eds.) Electronic Voting - Third International Joint Conference, E-Vote-ID 2018, Bregenz, Austria, October 2-5, 2018, Proceedings. Lecture Notes in Computer Science, vol. 11143, pp. 17–34. Springer (2018). https://doi.org/10.1007/978-3-030-00419-4_2, `https://doi.org/10.1007/978-3-030-00419-4_2`
7. Huang, Z., Rivest, R.L., Stark, P.B., Teague, V.J., Vukcevic, D.: A unified evaluation of two-candidate ballot-polling election auditing methods. In: Krimmer, R., Volkamer, M., Beckert, B., Küsters, R., Kulyk, O., Duenas-Cid, D., Solvak, M. (eds.) Electronic Voting - 5th International Joint Conference, E-Vote-ID 2020, Bregenz, Austria, October 6-9, 2020, Proceedings. Lecture Notes in Computer Science, vol. 12455, pp. 112–128. Springer (2020). https://doi.org/10.1007/978-3-030-60347-2_8, `https://doi.org/10.1007/978-3-030-60347-2_8`

8. Lindeman, M., Stark, P.B.: A gentle introduction to risk-limiting audits. IEEE Security & Privacy **10**(5), 42–49 (2012)
9. Lindeman, M., Stark, P.B., Yates, V.S.: BRAVO: Ballot-polling risk-limiting audits to verify outcomes. In: EVT/WOTE (2012)
10. McLaughlin, K., Stark, P.B.: Simulations of risk-limiting audit techniques and the effects of reducing batch size on the 2008 california house of representatives elections (2010), https://www.stat.berkeley.edu/users/vigre/undergrad/reports/McLaughlin_Stark.pdf
11. McLaughlin, K., Stark, P.B.: Workload estimates for risk-limiting audits of large contests (2011), https://www.stat.berkeley.edu/~stark/Preprints/workload11.pdf
12. Rivest, R.L., Shen, E.: A Bayesian method for auditing elections. In: EVT/WOTE (2012)
13. Stark, P.B.: Simulating a ballot-polling audit with cards and dice. In: Multidisciplinary Conference on Election Auditing, MIT (december 2018), http://electionlab.mit.edu/sites/default/files/2018-12/eas-ballotpollingsimulation.pdf
14. Stark, P.B., Wagner, D.A.: Evidence-based elections. IEEE Secur. Priv. **10**(5), 33–41 (2012). https://doi.org/10.1109/MSP.2012.62, https://doi.org/10.1109/MSP.2012.62
15. Vora, P.L.: Risk-limiting Bayesian polling audits for two candidate elections. CoRR **abs/1902.00999** (2019), http://arxiv.org/abs/1902.00999
16. VotingWorks: Arlo, https://voting.works/risk-limiting-audits/
17. Wald, A.: Sequential tests of statistical hypotheses. The Annals of Mathematical Statistics **16**(2), 117–186 (1945)
18. Zagórski, F., McClearn, G., Morin, S., McBurnett, N., Vora, P.L.: The Athena class of risk-limiting ballot polling audits. CoRR **abs/2008.02315** (2020), https://arxiv.org/abs/2008.02315
19. Zagórski, F., McClearn, G., Morin, S., McBurnett, N., Vora, P.L.: Minerva– an efficient risk-limiting ballot polling audit. In: 30th USENIX Security Symposium (USENIX Security 21). pp. 3059–3076. USENIX Association (Aug 2021), https://www.usenix.org/conference/usenixsecurity21/presentation/zagorski