# Simulations of Ballot Polling Risk-Limiting Audits

No Author Given

[1] Department of Computer Science, The George Washington University*
odbroadrick@gmail.com
[2] Department of Computer Science, Stanford University grantmcc@stanford.edu
[3] Wroclaw University of Science and Technology** and
[4] Votifica
filip.zagorski@gmail.com

**Abstract.** A Risk-Limiting Audit (RLA) draws paper election ballots in subsequent samples called rounds. After each round, a rigorous error criterion decides whether to stop (confirming the announced outcome) or continue (draw another round). The most commonly used RLA, Bravo (Stark), requires the fewest ballots when round sizes are 1. Recent RLA Minerva requires fewer ballots than Bravo but requires round sizes to be fixed before the audit begins. We present Providence, and audit that has similar efficiency to Minerva and supports choosing round sizes during the audit. We prove Providence is Risk-Limiting and resistant to an adversary who can choose subsequent round sizes given previous samples. We present simulations in which Providence has similar efficiency to Minerva. Finally, we describe the pilot use of Providence by the Rhode Island Board of Elections. Our implementation of Providence in the R2B2 software library provides standard functionality plus a function which gives the probability with which an announced loser will appear to be winning after a round, a potentially useful metric to election officials when choosing round sizes.

**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* [12] 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

* Supported in part by NSF Award 2015253
** Author was partially supported by Polish National Science Centre contract number DEC-2013/09/D/ST6/03927

audit of the election outcome, known as a risk-limiting audit (RLA) [5]. 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 in the US 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 focuses on ballot-polling RLAs, which require a large number of ballots relative to comparison RLAs but do not rely on any special features of the election technology. Since comparison RLAs are not always feasible, ballot-polling audits remain an important resource and have been used in a number of US state pilots (California, Georgia, Indiana, Michigan, Ohio, Pennsylvania and elsewhere). In the general ballot-polling RLA, a number of ballots are drawn and tallied in what is termed a *round* of ballots [17]. 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.* [17] 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. Broadrick *et al.* provide further simulations that show Minerva requires fewer ballots over multiple rounds and for lower stopping probability.

Both Bravo and Minerva have been integrated into election audit software *Arlo* [14], and, as such, are available for use in real election audits. Both have been used in real election audits [13, 17]. For this reason, it is important to understand their properties over multiple rounds.

### 1.1 Our Contributions

We present Providence, and provide the following:

1. Proof that Providence is an RLA and resistant to an adversary who can choose subsequent round sizes with knowledge of previous samples
2. Simulations of Providence, Minerva, SO Bravo, and EoR Bravo which show that Providence has similar efficiency to Minerva, both greater than either implementation of Bravo
3. Results and analysis from the use of Providence in a pilot audit in Rhode Island
4. Open source implementation of Providence including the novel metric Probability of Misleading(name?)

### 1.2 Organization

Section 2 describes related work. The experiments we performed are described in section **??**, and sections **??** and **??** present our results. Section 6 has our conclusions.

## 2 Related work

The Bravo audit [6] 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) [15], 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 large numbers of ballot, called rounds, before making stopping decisions because sequentially sampling individual ballots has significant overhead (unsealing storage boxes and searching for individual ballots). It is possible to apply Bravo to the sequence of ballots in a round 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 estimated as implemented in [14]; this method is due to Mark Lindeman and uses quadratic approximations. A later method for stopping probability estimates presented by Zagórski *et al.* [16,17] uses a similar technique for narrow margins and a separate algorithm for wider margins, the results of which match simulation results reported by Lindeman *et al.* [6, Table 1].

The Minerva audit [16, 17] 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 inform 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 are smaller than those for EoR and SO Bravo. First round simulations of Minerva [16] 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 [11]. McLaughlin and Stark [7, 8] 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 [2]. Huang *et al.* present a framework generalizing a number of ballot polling audits and compare their performance (round sizes and stopping probabilities) using simulations [4]. This work was prior to the development of Minerva, and focuses on the comparison between Bayesian audits [10] and Bravo, essentially studying the impact of the prior of the Bayesian RLA. Some workload measurements have been made [3]. While total ballots sampled can give naive workload estimates [9], Bernhard presents a more complex workload estimation model [1].

## 3 Providence

### 3.1 Resistance against an adversary choosing round sizes

## 4 Simulations

We run simulations assuming announced outcomes are correct to give insight into stopping probability and number of ballots drawn.

## 5 Pilot

A pilot audit was performed in Providence, Rhode Island in February 2022 from November 2021 special elections. The audited contest was a question on School Construction and Renovation Projects which had an announced 25.67% margin. The risk-limit was 10%. A large first round size of 140 with probability of stopping 95% was selected in order to give many ballots for more interesting analysis afterwards.

## 6 Conclusion

## References

1. Bernhard, M.: Election Security Is Harder Than You Think. Ph.D. thesis, University of Michigan (2020)

2. 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
3. Cause, C., VerifiedVoting, Center, B.: Pilot implementation study of risk-limiting audit methods in the state of rhode island. https://www.brennancenter.org/sites/default/files/2019-09/Report-RI-Design-FINAL-WEB4.pdf
4. 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
5. Lindeman, M., Stark, P.B.: A gentle introduction to risk-limiting audits. IEEE Security & Privacy **10**(5), 42–49 (2012)
6. Lindeman, M., Stark, P.B., Yates, V.S.: BRAVO: Ballot-polling risk-limiting audits to verify outcomes. In: EVT/WOTE (2012)
7. 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
8. McLaughlin, K., Stark, P.B.: Workload estimates for risk-limiting audits of large contests (2011), https://www.stat.berkeley.edu/~stark/Preprints/workload11.pdf
9. Ottoboni, K., Bernhard, M., Halderman, J.A., Rivest, R.L., Stark, P.B.: Bernoulli ballot polling: A manifest improvement for risk-limiting audits. International Conference on Financial Cryptography and Data Security pp. 226–241 (2019)
10. Rivest, R.L., Shen, E.: A Bayesian method for auditing elections. In: EVT/WOTE (2012)
11. 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
12. 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
13. Virginia Department of Elections: Results of risk-limiting audit of nov. 3, 2020 general election in virginia. https://www.elections.virginia.gov/rla-results_nov-3-2020/
14. VotingWorks: Arlo, https://voting.works/risk-limiting-audits/
15. Wald, A.: Sequential tests of statistical hypotheses. The Annals of Mathematical Statistics **16**(2), 117–186 (1945)
16. 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

17. 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