

A Review of Robust Post-Election Audits

Various Methods of Risk-Limiting Audits
and Bayesian Audits

Edited by Liz Howard, Ronald L. Rivest, and Philip B. Stark PUBLISHED NOVEMBER 7, 2019

Table of Contents

Glossary	1
Introduction	3
Executive Summary	4
Risk-Limiting Audit Definition	5
Project Overview	6
Michigan Risk-Limiting Audit Pilot Results	9
Michigan RLA Partnership Background	10
Bureau of Elections Overview	11
City of Kalamazoo	13
City of Lansing	19
City of Rochester Hills	21
Risk-Limiting Audit Overview	23
Bayesian Audit Overview	26
K-Cut Overview	28
Conclusion	30
Acknowledgments	31

ABOUT THE BRENNAN CENTER FOR JUSTICE

The Brennan Center for Justice at NYU School of Law is a nonpartisan law and policy institute that works to reform, revitalize—and when necessary defend—our country’s systems of democracy and justice. The Brennan Center is dedicated to protecting the rule of law and the values of constitutional democracy. We focus on voting rights, campaign finance reform, ending mass incarceration, and preserving our liberties while also maintaining our national security. Part think tank, part advocacy group, part cutting-edge communications hub, we start with rigorous research. We craft innovative policies. And we fight for them—in Congress and the states, in the courts, and in the court of public opinion.

STAY CONNECTED TO THE BRENNAN CENTER

Visit our website at
www.brennancenter.org

Sign up for our electronic newsletters at
www.brennancenter.org/signup

© 2019. This paper is covered by the [Creative Commons Attribution-NonCommercial-NoDerivs license](https://creativecommons.org/licenses/by-nc-nd/4.0/). It may be reproduced in its entirety as long as the Brennan Center for Justice at NYU School of Law is credited, a link to the Center’s web pages is provided, and no charge is imposed. The paper may not be reproduced in part or in altered form, or if a fee is charged, without the Center’s permission. Please let the Center know if you reprint.

Glossary

Ballot Comparison (or Ballot-Level Comparison)

Audit: An audit in which individual ballots are sampled and the audit interpretation of each sampled ballot is compared with the voting system interpretation of the ballot.

Ballot Manifest: A list that indicates how the ballots in an election are organized and stored. For instance, a ballot manifest might list the ballot containers used for an election, the number of batches in each container, and the number of ballots in each batch.

Ballot-Polling Audit: An audit in which individual ballots are sampled and interpreted, much as in a public opinion poll, collecting evidence regarding whether the original outcome is correct (e.g., that the reported winner received the most votes). Ballot-polling audits do not depend upon (or use) the voting system interpretations of the ballots in the sample.

Cast Vote Record (CVR): Archival record of how the voting system interpreted all votes produced by a single voter. CVRs are generated by the voting system, but may be stored in electronic, paper, or other form.

Central Count System: Ballot collection and tabulation method where, once marked, ballots are centrally aggregated to be scanned at a central location. Vote-by-mail ballots are typically counted centrally.

Count-Down Method: A method of locating the individual ballots randomly selected to be reviewed. The method involves auditors counting individual ballots in a batch until they reach the selected ballot, e.g., the 512th ballot.

Diluted Margin: The smallest reported margin of victory (in votes) among contests being audited together as a group from the same sample, divided by the total number of ballots in the group of ballots from which the sample is to be drawn. (The ballots from which the sample is drawn must include every ballot cast in the contests under audit, and may include other ballots in addition.) Dividing by the number of ballots, rather than by the number of valid votes, allows for the possibility that the vote tabulation system mistook an undervote or overvote for a valid vote, or vice versa.

Hybrid RLA Method: An RLA method in which two or more different ways of using voter intent ascertained manually from the ballots selected are combined.

Imprinter: Equipment that can add a unique identification number to individual paper ballots before, during, or after they are scanned by the tabulator.

K-Cut Method: A method of approximating the random selection of ballots from a physical stack of ballots that involves doing k “cuts,” each involving moving a random portion of ballots from the top to the bottom of the stack, and then picking the ballot on top.

Precinct Count System: Ballot collection and tabulation method in which ballots are tabulated in the polling place. C.f., central count system.

Random Seed: A random seed is a number or string used to initialize a pseudorandom number generator, which is in turn used to generate a pseudorandom sample of ballots. Best practice selects the the random seed in a public ceremony involving 20 rolls of 10-sided dice.

Risk Limit: The largest probability that, if an outcome is wrong, the audit does not correct that outcome. For example, assume the reported outcome of an election contest is wrong, and the risk limit for the audit is 5%. In this instance, there is at most a 5% chance that the audit will not correct the outcome, and at least a 95% chance that the audit will correct the outcome. The risk limit is a number between 0 and 1 that quantifies the risk that the audit will not correct an incorrect outcome. The risk limit should be set before the audit is conducted, preferably in legislation.

Risk-Limiting Audit: A risk-limiting audit is any post-election procedure that has a known probability of correcting the reported outcome of an election contest if the reported outcome is incorrect, and no possibility of altering a reported election outcome that is correct. Here, “outcome” means the winner or winners or winning position (not the exact vote totals), and “incorrect” means that the outcome disagrees with what an accurate tabulation of voter intent ascertained manually directly from voter-verified paper ballot would show. A risk-limiting audit ensures that at the end of the canvass, the hardware, software, and procedures used to tally votes found the real winners. Risk-limiting audits do not guarantee that the electoral outcome is right, but they have a known minimum chance of correcting the outcome if it is wrong: see “risk limit.” Risk-limiting audits involve manually

examining some or all of the paper ballots validly cast in the contest.

Tool: Software calculator that performs various calculations required to conduct a risk-limiting audit. For example, a tool may be used to determine which ballots the audit should inspect manually and when the audit can stop.

Upset Probability: A notion specific to a Bayesian audit, defined as the probability that the reported winner(s) would be discovered not to win the audited contest if all ballots were to be examined. This notion is based on a model of voter preferences derived from the ballots sampled so far and a prior probability distribution on the voter preferences.

Audit Software Tools:

- **Ballot-Polling RLA Tool:** A tool designed by Dr. Philip Stark to conduct ballot-polling RLAs. <https://www.stat.berkeley.edu/~stark/Vote/ballotPollTools.htm>. This tool is open-source and available on the web and on Github.

- **Ballot Comparison RLA Tool:** A tool designed by Dr. Philip Stark to conduct ballot comparison RLAs. <https://www.stat.berkeley.edu/~stark/Vote/audit-Tools.htm>. This tool is open-source and available on the web and on Github.
- **BCTool:** A tool designed by Prof. Ronald L. Rivest and Mayuri Sridhar for Bayesian upset probability measurements for Bayesian hybrid audits. This tool is open-source and available on Github. <https://github.com/ron-rivest/2018-bctool>.
- **SUITE Tool:** A tool designed by Kellie Ottoboni and Dr. Philip Stark to assist the Michigan pilots conducting hybrid audits. The code is written in Python and displayed in an interactive Jupyter Notebook. The tool can be run interactively at https://mybinder.org/v2/gh/pbstark/CORLA18/master?filepath=code\%2Fsuite_toolkit.ipynb. This tool is open-source.

Introduction

In August of 2018, the Michigan Bureau of Elections and the city clerks of Kalamazoo, Lansing, and Rochester Hills partnered with the Brennan Center for Justice, Professor Ron Rivest and Mayuri Sridhar of MIT, Dr. Philip Stark and Kellie Ottoboni from the University of California, Berkeley, Jerome Lovato of the U.S. Election Assistance Commission, Verified Voting Foundation, and the Voting System Technical Oversight Program at Ball State University (the “RLA Team”) to conduct the first risk-limiting audit (“RLA”) pilot in Michigan.

While the main goal was to provide Michigan election officials with a hands-on learning experience about RLAs, the “gold standard” of post-election tabulation audits, participants gained broadly-applicable insights about best practices. We hope this report can serve as a resource to election officials across the country who are considering adding RLAs to their election security procedures or conducting similar pilots. We have designed it to be useful for election officials with varying levels of familiarity with RLAs.

The executive summary and the project overview provide a high-level overview of the project, introduce the different audit methods used, and list the voting systems and RLA tools used in each jurisdiction. A risk-limiting audit definition and a chart that summarizes the pilot audit results in each jurisdiction are also provided.

Next, election officials offer their insights on the process, the pilot and their goals moving forward. State

election officials provide the background to this unique partnership and add their perspective on the project. Importantly, they discuss lessons learned through piloting different audit methods and random ballot selection methods, and how they plan to move forward in the future.

Local election officials then provide a detailed overview of their individual experiences during the pilot. Each discusses their workload, communication strategy, implementation strategy and lessons learned. They also provide recommendations to state election officials.

Finally, two members of the RLA Team describe the foundational mathematics, specific procedures, sampling methods and RLA software tools.

If you have additional questions about the procedure, the pilot, or any other aspect of this partnership, as we hope you do, please do not hesitate to contact anyone involved.

Executive Summary

State and local election officials serve as our democracy’s last line of defense against malevolent foreign actors, equipment malfunctions, and human errors which may impact election results. Unfortunately, this responsibility often comes with few resources and heightened public scrutiny. One smart and effective tool available to election officials facing this reality is the post-election risk-limiting audit (“RLA”).

This special type of audit uses statistical methods and a manual review of paper ballots to check the accuracy of reported election outcomes. Specifically, RLAs are designed to provide assurance that the reported winner did in fact win the election, or in the alternative, to correct errors caused by cyberattacks, bugs, misconfiguration, or human error, if any combination of those altered the reported outcome. While the underlying math may be challenging for non-mathematicians to understand, the procedures to conduct such audits were shown to be straightforward.

In August of 2018, the Bureau of Elections agreed to partner with the RLA Team to conduct the first risk-limiting audit pilots in Michigan. The RLA Team worked directly with election officials to understand the relevant election administrative procedures and practices. They used this information to draft audit instructions (called audit protocols) for each locality. Each participating municipality had one day to conduct their pilot between December 3-5, 2018.

This approach produced many practical lessons. Most importantly, Michiganders gained confirmation that RLAs are possible in their state, which relies primarily on precinct-based voting on election day. Many other lessons

related to the procedure, voting systems, and messaging were drawn from this groundbreaking collaboration. These lessons include:

- Risk-limiting audits are an effective tool that can be implemented in Michigan, by Michigan election officials, using Michigan-certified voting systems.
- Risk-limiting audits are not procedurally difficult to implement by election officials.
- Risk-limiting audits can be implemented with minimal changes to pre-election and Election Day administrative procedures.

Overall, Michigan election officials were impressed with the results from the pilots, especially the potential to greatly improve post-election audit efficiency. While further work and additional pilots are necessary, this pilot equipped election officials with practical information necessary to make important election security policy decisions in Michigan. They currently plan to take the pilot to the next step by conducting RLAs at the county level after the May 2019 elections.

What Is a Risk-Limiting Audit?

A risk-limiting audit is any procedure that has a known chance of correcting the reported outcome if the outcome is wrong¹, and no chance of changing a correct reported outcome. The chance that the RLA does not correct an outcome that is wrong is the “risk limit.”

RLAs may be further characterized as “a method to ensure that at the end of the canvass, the hardware, software, and procedures used to tally votes found the real winners. Risk-limiting audits do not guarantee that the electoral outcome is right, but they have a large chance of correcting the outcome if it is wrong. They involve manually examining portions of an audit trail of (generally paper) records that voters had the opportunity to verify recorded their selections accurately.”²

From a procedural standpoint, an RLA is an audit of the reported results during which ballots cast in the election are examined. The procedure first requires election results and the predetermined risk limit (the maximum chance that the audit process will fail to correct an incorrectly-reported winner expressed as a percentage) to be entered into a formula (generally using a software calculator referred to as an “audit tool”). The audit tool then calculates the number of ballots to be reviewed, and randomly selects specific ballots or batches. Auditors then retrieve and review these randomly sampled ballots. The results

from this human review are entered into the audit tool. The audit tool calculates the measured risk based on a statistical analysis of these results.

In contrast to traditional post-election audits, where ballots cast in or on a fixed percentage of individual voting precincts or voting machines are reviewed, the percentage of ballots reviewed during a risk-limiting audits depends on the margin of victory in the contest being audited. The smaller the margin of victory, the larger the number of ballots that the audit will have to reviewed, all else equal. The predetermined risk limit also inversely impacts the number of ballots to be reviewed.

RLAs have the advantage of being both effective and efficient because they adjust the workload to get just enough evidence that contest results are correct, if contest results are indeed correct. (If the contest results are incorrect, an RLA has a large chance of leading to a full manual tally to correct the results.)

1 Here, wrong means that the winner(s) are not the winner(s) that a full tabulation based on manually determining voter intent from the paper would find.

2 Mark Lindeman and Philip B. Stark, “A Gentle Introduction to Risk-limiting Audits,” *IEEE Security and Privacy Special Issue on Electronic Voting* (2012): 1. Available at <https://www.stat.berkeley.edu/~stark/Preprints/gentle12.pdf>.

Project Overview

After learning that Michigan election officials were researching various post-election audits, the Brennan Center for Justice approached the Michigan Bureau of Elections Director, Sally Williams, and offered to work in partnership with the Bureau to pilot RLAs in multiple jurisdictions in Michigan. The RLA Team was quickly assembled and the Bureau selected three municipal clerks to participate in the pilot: Tina Barton, clerk, City of Rochester Hills; Scott Borling, clerk, City of Kalamazoo; and Chris Swope, clerk, City of Lansing.

Michigan is an ideal state to pilot RLAs for many reasons. Local and municipal officials share election administration responsibilities and they represent jurisdictions of vastly different sizes and electorates. In addition, the secretary of state has a significant amount of discretion over the post-election audit procedures conducted at the local level.

The RLA Team worked directly with the local election officials to gain an understanding of state law, local election administration procedures and practices and various regulations. They also shared information about risk-limiting audits, Bayesian audits, the underlying mathematics, and audit methodologies.

One question many Michigan election officials asked was, “Will my voting system support a risk-limiting audit?” Because Michigan uses only voting systems that employ paper ballots, the answer was simple: yes. However, each county may use any one of three state-certified voting systems. To accommodate differences among voting

systems, there are multiple RLA *methods* from which election officials can choose. The RLA method selected is generally determined by the voting system’s capability to produce cast vote records and on how the jurisdiction organizes voted ballots. Because each municipality in this pilot used different voting systems, different RLA methods were used by different localities.

Not only did Michigan election officials pilot different RLA methods, but they also piloted different methods of Bayesian audits. Bayesian audits are a close cousin of RLAs and are explained by Mayuri Sridhar in greater detail in the Bayesian Audit Overview Section below. RLAs and Bayesian audits involve the same steps; the difference between them is what they guarantee. In particular, an RLA guarantees that if the reported outcome is incorrect, the audit has a known, pre-specified chance of correcting the outcome. (Bayesian audits do not offer that guarantee.) A description of the different audit methods piloted in Michigan appears below:

Method	Overview	Requirements
Ballot Comparison	Auditors manually review randomly selected paper ballots; for each ballot, their visual interpretation of each vote is recorded and compared to the voting machine’s record of how that ballot was originally tallied by the audit Tool (software calculator).	1) Voting system must retain a record of how the ballots were counted (a Cast Vote Record/ “CVR”) in the order scanned; 2) Retention of ballots in the order scanned; and 3) Ballot manifest (log of all ballots cast and their storage location).
Ballot-Polling	Auditors manually review randomly selected paper ballots; the results are aggregated and analyzed in stages by the audit Tool (software calculator).	Ballot manifest (log of all ballots cast and their storage location).
Hybrid	Auditors review randomly selected ballots. These ballots are sorted into two different groups: ballot comparison and ballot-polling. The results are analyzed as described above.	1) Voting and ballot storage system that meets the requirements for a Ballot Comparison audit, and 2) Voting system that meets the requirements for a Ballot-Polling audit. E.g., using a DS450 to tabulate absentee ballots and a DS200 to tabulate ballots cast in precinct.

As Kellie Ottoboni explains in the Risk-Limiting Audit Overview Section, the ballot comparison method is the most efficient method in that fewer ballots must be reviewed to provide statistical assurance that the originally reported election outcome is accurate, if the originally reported outcome is indeed accurate. This means if a jurisdiction uses a voting system that supports ballot comparison audits (i.e., the voting system maintains a cast vote record and there is a way to connect each ballot with its cast vote record) election officials in that jurisdiction will be able to conduct the most efficient audits. If the voting system in use doesn't meet the ballot comparison audit requirements listed above, then so long as paper ballots are used, a ballot-polling audit can be conducted.

For elections in which different types of equipment are employed (e.g., if absentee ballots are tabulated centrally and Election Day ballots are tabulated in-precinct) it may be possible to use the Hybrid method. Hybrid audits may increase the efficiency of the audit by allowing the use of the ballot comparison method to review ballots cast on equipment which meets the requirements above and the use of the ballot-polling method to review the remainder. One lesson learned during this pilot is that Hybrids do not always result in increased efficiency, however. In places like Michigan, where most of the ballots are currently cast in polling places on election day and are only auditable via the ballot-polling method, the Hybrid method may result in auditing more ballots than if the ballot-polling method had been used for all ballots.

As each certified system in Michigan has different features, these features determined the audit method used. The Hart Verity Voting system used in Rochester Hills does not meet the requirements for a ballot comparison audit described above. Therefore, we piloted a ballot-polling audit in Rochester Hills. Since the ES&S DS450 (used for absentee ballots in Kalamazoo) and the Dominion ICC High Speed Scanner (used for absentee ballots in Lansing) satisfy the requirements for the ballot comparison audit, and their in-precinct voting equipment satisfies the requirements for a ballot-polling audit (but not for the more efficient ballot comparison audit), we piloted Hybrid audits in Kalamazoo and Lansing.

For each method, we piloted both the RLA and the Bayesian audit. This approach did not result in additional work for election officials because the Bayesian calculations can be conducted using the same data obtained in the RLA audit procedure. The results from the review of the randomly selected ballots were entered into two different software calculators: the SUITE tool for the RLA calculations and the BCTool for the Bayesian calculations. Based on the results entered, the SUITE tool provided the RLA risk limit and the BCTool provided the Bayesian "upset probability."

In Kalamazoo, we made an important discovery about

the voting equipment in use: the DS450, used for tabulating absentee ballots, is equipped with an imprinter. The Kalamazoo Clerk worked directly with the vendor and found that, with the insertion of an ink cartridge, the imprinter could imprint unique identifiers on every ballot during the tabulation process. This functionality had a significant positive impact on the efficiency of the pilot in Kalamazoo. Further, it had a significant impact on the assessment of all the voting equipment used in the pilot.

After the audit method was selected, the next step was to choose the sampling method (the method used to randomly select ballots for review during the audit). Prior to the Michigan project, the main sampling method used in RLAs was the count down method, where auditors count down through a batch of ballots to find a particular ballot that has been randomly selected by the audit tool. In 2018, Rivest and Sridhar created a new sampling method, called *k*-cut.³ Testing in Michigan proved *k*-cut to be more efficient than counting down when stacks of ballots are large. *K*-cut is not appropriate for ballot comparison audits at this time.

Additional considerations arose as this pilot was planned during a live Michigan election, specifically the 2018 November General Election. Accordingly, an important element of this pilot was an effective communication plan to share information with voters and the public in general about the pilot. A key goal of election officials who implement RLAs is to justify greater voter confidence in our electoral system. To do this in Michigan, election officials worked to educate the press and the voters about the process and its important role in election security.

The RLA Team and Judd Choate, the Elections Director for the Colorado Secretary of State, assisted the Michigan Bureau and individual clerks with their RLA pilot communication strategy. While the state and local election officials believed that announcing the pilot was an important opportunity to educate voters about RLAs and their importance, they also believe that the existing RLA talking points can be much improved.

With these lessons in mind, the Bureau plans to expand the RLA pilot to include county election officials and work directly with voting system vendors to improve audit support functionality.

3 Mayuri Sridhar and Ronald L. Rivest, "k-Cut: A Simple Approximately-Uniform Method for Sampling Ballots in Post-Election Audits" available at <https://arxiv.org/abs/1811.08811>.

Jurisdiction	Vendor	Voting System	Pilot Type
Kalamazoo	ES&S	EVS 6010: DS200 (in-precinct); DS450 (absentee)	Hybrid Ballot-Polling (in-precinct ballots) and Ballot Comparison (absentee ballots)
Lansing	Dominion	DVS 5.0S: ICP Tabulator (in-precinct); ICC High Speed Scanner (absentee)	Hybrid Ballot-Polling (in-precinct ballots) and Ballot Comparison (absentee ballots)
Rochester Hills	Hart	Verity Voting 2.2.2 Verity (in-precinct); Verity Central (absentee)	Ballot-Polling

Michigan RLA Pilot Results

	Kalamazoo	Lansing⁴	Rochester Hills
Vendor	ES&S	Dominion	Hart
Audit Method/s Used	Hybrid Ballot-Polling (in-precinct ballots) and Ballot Comparison (absentee ballots)	Hybrid Ballot-Polling (in-precinct ballots) and Ballot Comparison (absentee ballots)	Ballot-Polling
Predefined Risk Limit	5%	5%	5%
Total Ballots Cast	27,666	21,328	36,666
Votes cast for winner	20,699	10,309	22,999
Votes cast for the next highest vote recipient	5,569	7,694	12,343
Margin of victory	55%	12%	29%
Percentage of Ballots with usable CVR	19%	50%	0%
Number of ballots audited	40	260	76
Percentage of ballots audited	.14%	1.20%	0.20%
Attained Risk	3.70%	51.00%	2.10%
Calculated Bayesian Upset Probability	0.03%	9.95%	0.3%

⁴ For pilot purposes, certain ballots and batches were excluded from consideration. Also, the risk-limit and upset probability were calculated based on the ballot-polling method.

Michigan RLA Partnership Background

Virginia Vander Roest

While risk-limiting audits are new to Michigan, post-election audits are not. Michigan first implemented procedural post-election audits in 2013. These audits are performed by county clerks and primarily focus on verifying that local election officials have properly implemented many important election security measures. Although some traditional percentage-based audits were added in November 2017, the current post-election procedures do not include a process aimed at verifying election outcomes. The Michigan Bureau of Elections recently prioritized the research of audit procedures that could provide this check on election outcomes.

In May of 2018, the Brennan Center asked the Michigan Bureau of Elections to collaborate on a project conducting pilot risk-limiting audits for the November general election. Always open to opportunities to improve election administration in Michigan, the Bureau agreed. The Bureau's first step in the RLA pilot process was to identify three local jurisdictions to participate in the project and audit results from the November 2018 election. The Bureau approached the clerks in three jurisdictions who eagerly accepted the challenge: the City of Rochester Hills, Oakland County; City of Lansing, Ingham County; and the City of Kalamazoo, Kalamazoo County.

While these clerks share many characteristics, such as being well-respected and influential members of the Michigan election administration community and serving mid-size cities which use an Absent Voter Counting Board (central count) with high speed tabulators to count their absentee ballots, there is one important difference. Each uses a different one of the three Michigan-certified voting systems. Assessing all the different voting equipment in use in Michigan was an important aspect of this pilot. The pilots' success in all three jurisdictions enabled Bureau staff to confidently recommend expanding the project to include county election officials in the May 2019 elections.

One concern held by many Michigan election officials prior to this pilot was related to the fact that Michigan's election administration system is based on in-precinct voting. This system is very different than the vote-by-mail system used in Colorado, the one state widely renowned for conducting and pioneering RLAs. Further, Colorado

has 64 local election officials while Michigan has 1,520 election officials. Is it possible to implement RLAs in a state that conducts elections so differently than Colorado? To squarely answer this question, the Bureau decided to partner with experts and pilot the procedure in Michigan.

In September of 2018, Bureau staff and the participating clerks attended an RLA Partnership kickoff meeting with the RLA Team. Weekly conference calls started immediately thereafter. As each jurisdiction used a different voting system, each individual clerk worked closely with their vendors to gather information on the system capabilities, including the ability to access and review the cast vote record (CVR). Overall, the jurisdictions found they did not need to change procedures. The pre-election and Election Day procedures for precinct-based voting did not change and the absent voter counting board process only required extra diligence to ensure ballot order was maintained in those jurisdictions that had access to a CVR.

In addition to weekly calls, each clerk attended and observed the State of Colorado at Denver City/County's risk limiting audit in November 2018. The ability to observe an RLA in person was an exceptionally important part of the learning process. This observation day provided the Michigan project team with the crucial opportunity to connect the information they had collected with the preparations that were underway – witnessing an audit allowed them to put that information together in a practical way. In early December, the Michigan team held its first RLA pilot.

Bureau of Elections Overview

Virginia Vander Roest

The partnership with the RLA Team provided for a rich and successful learning environment that allowed us to test several different RLA processes and theories within each jurisdiction. We found early on that the single RLA method employed in Colorado (a vote-by-mail state in which ballots are counted centrally), the ballot comparison method, is simply not feasible in a state like Michigan in which the majority of ballots are scanned and tabulated on Election Day in precincts. This reality necessitated a pilot which incorporated new techniques and Michigan's unique election administration model that relies predominantly on a precinct-count system.

Our pilot did just that. We piloted various different methods of RLAs and Bayesians and found what worked in Michigan. The pilot was a success in that it provided proof that risk-limiting audits and/or Bayesians audits are possible in Michigan. This process equipped staff with sufficient experience and expertise to confidently recommend the continuation of pilots across the state at the county level after the May 2019 elections.

Lessons Learned

This project resulted in several lessons learned. Counterintuitively, we learned the most about the process in the jurisdiction facing the greatest number of obstacles and real-world issues. Evaluating the process by actually conducting pilots is invaluable. With this hands-on experience, the Bureau now has the practical knowledge necessary to expand the pilot to the county-level and use evidence gathered in Michigan by Michigan election officials to determine the most effective and efficient audit.

BALLOT COMPARISON V. BALLOT-POLLING

The ballot comparison auditing method requires retaining and storing absentee ballots in the order scanned and tabulated. Ensuring strict compliance with this requirement proved challenging when the absentee voting equipment did not imprint unique IDs on the ballots during tabulation. Without a unique ID on the ballot, it was not possible to confirm that the ballot reviewed by auditors, e.g., Ballot No. 312, was in fact Ballot No. 312 in the CVR, the file which contains the record of how Ballot No. 312 was tabulated. Errors can occur during the count down method process during which the auditors count the ballots one-by-one until they reach Ballot No. 312. Errors can also occur when election officials store the ballots in the ballot bag and either fail to indicate the top v. bottom of the stack of ballots or inaccurately identify Ballot No. 1.

In the City of Kalamazoo, a unique numeric identifier ("ID") was imprinted on each absentee ballot as it was scanned and tabulated. The voting equipment

included the ID number on the CVR that lists how each vote was counted. In this way, the imprinted ID numbers provided a helpful verification that the correct ballot was pulled. In contrast, a unique ID was not imprinted on absentee ballots during scanning in Lansing because the voting system does not currently provide this functionality. Without this unique ID imprinted on the ballot, there was no opportunity to confirm that the absentee ballot retrieved by the auditors was the accurate ballot. As the ballot comparison method was used for absentee ballots in Lansing, it is imperative to retrieve the correct ballot.

The CVRs for each voting system evaluated need to be improved in a few ways. Specifically, CVRs need to be easier to save as an Excel or .csv file. Further, imprinter functionality for central count voting systems used to process absentee ballots is necessary to best support effective ballot comparison audits in jurisdictions and states which do not batch absentee ballots.

As the ballot-polling method does not require storing ballots in the same order scanned and tabulated, this method proved easier to effectively administer in jurisdictions using the Dominion ICC G1130 High Speed Scanner System Version 5.0 which does not include imprinter functionality in Michigan.

BALLOT COMPARISON AND BALLOT-POLLING HYBRID – SUITE TOOL

A big difference between the ballot comparison method and ballot-polling method is the efficiency of the process. As explained by Kellie Ottoboni, efficiency in this context is measured by the number of ballots that must be reviewed to provide sufficient statistical confidence that the election outcome is accurate (assuming that the reported election outcome is indeed accurate). The ballot-polling audit requires more ballots to be randomly selected and reviewed by hand than a ballot comparison audit. One interesting discovery made during the pilot was that, although we assumed that the Hybrid audit would reduce the number of ballots to be reviewed in

comparison to the number that would be required for a straight ballot-polling audit, that was not the case when the SUITE tool was used in our pilots. The Hybrid methodology actually called for a higher number of ballots to be reviewed than a ballot-polling audit.

RLA MODEL VS BAYESIAN MODEL

Using both statistical models for the audit proved to be helpful as calculation tools were used for the first time. The Bayesian Model showed a lower level of risk at each of the three audits and generally seems to require fewer ballots to reach a 5% upset probability than it takes to reach a 5% risk limit in an RLA when using the polling method.

BALLOT COUNTING METHOD VS *k*-CUT METHOD

An important element in post-election audits employing statistical methodology is the random selection of ballots. In Michigan, we employed two different methods of randomly selecting ballots: 1) Software calculator (“tool”) selection method coupled with a hand count down process, and 2) *k*-cut method. All three pilot jurisdictions considered all ballots cast in a single precinct to be an individual batch. Ballots cast absentee were similarly batched by precinct in Lansing and Kalamazoo, they were batched into groups of 50 in Rochester Hills. This approach resulted in batches as small as approximately 50 ballots to batches as large as approximately 1800 ballots.

Due to the large number of ballots generally found in precinct-based batches, the *k*-cut method proved to be much more efficient than counting to find a specific number in the batch, especially when the ballot number to find was higher than 200. The continued evaluation of that process would be valuable, even if its usage is limited. As the current *k*-cut methodology requires the batch to be “cut” six times, an elimination of a cut, coming from a better understanding of the math, would further increase efficiency. As the study of *k*-cut methodology and practice grows, it may be found to be mathematically sound to further increase efficiency by decreasing the number of cuts required.

Moving Forward

While we learned a lot in the three days of pilot audits, continued pilots are needed to identify the best auditing methods to use. Currently, our post-election procedural and traditional percentage-based audits are generally conducted by county clerks. While not a perfect comparison, the traditional percentage-based audit in Kalamazoo required the county clerk to hand-count approximately 30,000 ballots; the RLA pilot in the City of Kalamazoo required the review of exactly 40 ballots. If RLAs are

adopted in lieu of traditional percentage-based audits, it could save county clerks a significant amount of time. Accordingly, the next logical step in this exploratory process is a county-level RLA pilot audit. Some jurisdictions and counties will conduct elections in May of 2019 providing an opportunity to identify a few counties able to conduct a pilot audit of the entire county. As ballots are generally stored at the local, not county level, this will present one significant challenge in the county-wide audit.

We will be better able to specify RLA pilot valuation metrics when we obtain the list of counties in which May elections will be held. Even though that list is not yet available, we know that the size of these counties will vary drastically. Size is important because it often dictates absentee ballot processing procedures. For example, jurisdictions in smaller counties generally send their absentee ballots to the precincts to be tabulated, which will prevent the pilot of a ballot comparison audit. Continued evaluation of the use of the RLA model versus Bayesian model will be needed. Among other things, the Bureau plans to pay particular attention to the number of ballots needed in ballot-polling audits of races with a small margin of victory.

Another key factor to ensuring RLA scalability is the ability to continue to provide opportunities to election officials from across the state to observe the process and provide input. As more jurisdictions and counties participate, it will be necessary to form a committee to assist in ensuring that all have the tools they need to effectively conduct a rigorous and procedurally efficient RLA. As election officials from across the state were welcomed to participate and/or observe the pilots, many of these officials also provided constructive and valuable feedback that will instruct the procedure going forward.

Looking ahead, the development and/or procurement of a tool (or software suite) to effectively organize the process will be needed. All voting systems need to create CVR records which are easily accessible, usable and do not require the election administrator user to convert JSON files into user-friendly and widely-used formats, such as Excel files. Further, voting equipment with imprinter functionality best supports RLAs and this functionality should be expanded to all voting equipment. Imprints on ballots cast in-precinct would greatly increase the potential for efficient statewide ballot comparison audits.

Continued collaboration with academics to further improve the process, ensure the availability of necessary tools, and provide ongoing evaluation will be an essential element of success. Risk-limiting audits will continue to evolve around the country in different election environments, and learning from each other expedites the implementation process and continues to increase efficiency even once RLAs are adopted.

City of Kalamazoo

Scott Borling, Kalamazoo City Clerk

Overview

On Wednesday, December 5, 2018 the Kalamazoo City Clerk's Office hosted a pilot Risk Limiting Audit (RLA) of the results of the General Election held on November 6th. The audit took place at the Kalamazoo City Hall, began at approximately 9:30 a.m., and ended at approximately 2:00 p.m. Approximately 30 people were in attendance, including city staff, audit team members, pilot project partners, and observers. The Kalamazoo RLA was a *hybrid* RLA that combined the results of a ballot-polling audit and a ballot comparison audit. A ballot-polling audit was used to examine and evaluate ballots cast and tabulated in the city's 27 voting precincts on Election Day. A ballot comparison audit was used to examine and evaluate the absentee voter ballots cast prior to Election Day and processed on Election Day by an Absent Voter Counting Board. In addition, the same sample of ballots that was used for the RLA was also used for a Bayesian audit, which was conducted simultaneously. The target race was the race for governor, but all three statewide races (governor, secretary of state, and attorney general) were tallied. A 5% risk-limit was used.

General Information About the November 6, 2018 Election in the City of Kalamazoo

- 27 voting precincts at 19 polling locations
- Official number of registered voters on Election Day: 53,856
- Absent Voter Counting Board (AVCB) opened at 8:00 a.m. on Election Day.
- One (1) DS450 high speed tabulator from ES&S to tabulate the absentee ballots.
- Votes cast absentee: 5,294

Preparation for the Audit

Pre-election preparations for the Pilot RLA were considerable, as it took several weekly phone meetings with the project partners, phone and email conversations with the city's voting system vendor, and a few hours of equipment testing to conclude that very little in the way of pre-election preparation was necessary. In short, most of the pre-election preparation was learning about RLAs; trying to understand how the audit process would work with Michigan election laws, processes, and procedures; and discovering the capabilities of the City's voting system, including the Election Management System (EMS), the

software used to program the tabulators and receive, aggregate, and report the results. The following "lessons learned" during pre-election preparations shaped the way the pilot RLA would be executed in Kalamazoo:

- The DS450 high-speed tabulator used to process absentee voter ballots was able to imprint a serial number on each ballot as it was tabulated. These numbers provided links to the Cast Vote Records (CVR's) produced by the EMS. The imprinted numbers enabled the use of a ballot comparison audit for absentee ballots without requiring the city's AVCB to maintain the strict ballot order needed by other systems to match ballots with their correct CVRs. The only required change to AVCB procedures was the need to reorient the ballots if they had to be rescanned, so the new CVR number could be imprinted on a clean corner.
- Because the city's precinct-based tabulators do not have an imprinting feature and there is no way to keep Election Day ballots in order after they are tabulated, a ballot-polling audit would be used to audit the results of ballots cast and tabulated on Election Day in the city's 27 voting precincts.

The imprinting function on the high-speed tabulator requires no special programming and no changes to the scanner settings. The function is always "on," and the only way to turn it "off" is to remove the ink cartridge. Because the imprinting feature on the high-speed scanner had never been used in Michigan and was not specifically tested during the state certification process, representatives from the Bureau of Elections observed a test of the imprinter during the regular pre-election logic and accuracy testing of the city's absentee ballots. The Bureau granted permission to use the imprinter a few days later.

After Election Day the preparations included the creation of the ballot manifest, the gathering of results data, and the exporting of Cast Vote Records from the EMS. The ballot manifest⁵ was created using the number

⁵ Risk-Limiting Audits – Practical Application, Jerome Lovato, U.S. Election Assistance Commission (2018) 4 (defining "ballot manifest" as "A list that indicates how the ballots in an election are organized and stored. For instance, a ballot manifest might list the ballot containers used for an election, the number of batches in each container, and the number of ballots in each batch."), https://www.eac.gov/assets/1/6/Risk-Limiting_Audits_-_Practical_Application_Jerome_Lovato.pdf.

of ballots cast as indicated in the poll books for each precinct. Each Election Day precinct and absent voter precinct was a separate batch, and provisional ballots that were counted were considered a batch regardless of the precinct. This resulted in 55 batches (resulting from 27 Election Day precincts and 27 absentee precincts, adding provisional ballots as well).

Results data was obtained from the county clerk and forwarded to the project partners who were preparing the SUITE RLA tool. The aggregated results for the entire city were provided as well as results split between absent voters and Election Day voters.

To obtain the CVR's the City Clerk and Deputy City Clerk met with the County Clerk in his office where the EMS software is housed. As the County Clerk had never used the EMS (election management system) to access the CVRs, the vendor helped with the export process. CVR data was exported into two spreadsheets: one that contained ballot metadata, including the serial number imprinted on the ballot; and one that contained the selections for each ballot, as determined by the voting system. At this point there was some confusion as information provided by the vendor prior to the election was based on a newer version of the software, and some of the field names were different. But the vendor clarified the information. After a spot check of the data, the spreadsheets were forwarded to the project partners who were loading the CVR information into the SUITE RLA tool.

Execution of the Audit

The audit itself was carried out in three phases: seed selection; the retrieval of ballots; and the review of ballots. The decision was made to keep the retrieval and review of ballots separate from each other to allow everyone involved – city staff, the audit teams, project team members, and observers – to focus on each step in the audit process. This segregation of steps facilitated management of the audit and created a controlled environment where the process was clear and deliberate. In addition, this approach was designed to promote transparency, as observers were not forced to choose which parts of the process to watch. Finally, keeping the ballot review and retrieval processes separate allowed the city to use a relatively small number of volunteers without sacrificing speed or accuracy.

Four audit teams, with two auditors per team, both retrieved and reviewed the ballots. The audit teams were comprised of five local clerks from Kalamazoo County, two precinct chairpersons from the City of Kalamazoo, the Kalamazoo County Clerk, and a member of the Rhode Island Board of Elections who was present to observe the process. These volunteers were chosen because of their prior experience conducting hand recounts, which made them familiar with proper ballot handling techniques and counting ballots while being observed.

Seed Selection Ceremony

The seed selection ceremony was held in the City Commission Chambers. A series of die-rolls using 10-sided dice was used to generate the 20-digit “seed number” the SUITE RLA tool needed to create the list of ballots to be reviewed. As audit participants and observers arrived they were asked to put their names on small pieces of paper. The names were placed in a bowl and randomly drawn. As each name was drawn that person was invited to come forward and roll a die. The casting of the dice resulted in the following seed number:

5 9 0 4 8 5 3 5 5 2 4 6 2 2 1 2 0 0 4 6

The seed number was entered into the SUITE RLA tool, which identified 33 ballots for review. As there was ample time the number of ballots was increased to 40 to grow the sample size and give participants more experience with the process. Of the 40-ballot sample, 32 Election Day ballots were identified for review in the ballot polling audit, and 8 ballots were identified for review in the ballot comparison audit. The same 40-ballot sample was used for the Bayesian audit.

Retrieval of Ballots

The list of ballots to be retrieved was divided into four lists, one for each audit team. The lists were transferred to a USB flash drive from the computer running the SUITE RLA tool, and they were printed directly from the USB drive on a non-networked printer/copier stationed in the City Commission Chambers. The lists contained four data fields:

- Sampled Ballot (the number of the ballot within the entire strata)
- Batch Label (in most cases the precinct number)
- Which Ballot in Batch (the number of the ballot within the specified batch, also referred to as the Count Number)
- Number of Times Sampled (no ballot was sampled more than once)

The ballot retrieval process took place in the Community Room, a large conference room adjacent to the City Commission Chambers. Because the batches of absent voter ballots were stored together in two sealed containers, one audit team focused on retrieving those eight ballots for the sake of efficiency. The other three audit teams retrieved the 32 Election Day ballots. The retrieval lists were arranged so that multiple teams were not given ballots to retrieve from the same batches.

The audit teams used two different methods for retrieving ballots: the “count-down” method, and the *k*-cut method. With both methods the first step was to verify the seal number on the ballot container against the seal number on the Ballot Container Certificate. After the container was opened the audit team removed the identified batch of ballots from the ballot container and formed them into an organized stack.

COUNT DOWN METHOD

The count-down method is one method used to retrieve randomly selected ballots. When using this method, the RLA tool selects a random ballot from the ballot manifest. This ballot is identified using the information used in creating the ballot manifest, often a specific batch number, and the number of a ballot in the identified batch (for example, Ballot 412 in Batch 29.) In this example, the auditors found the ballot bag in Batch 29 was stored. They removed the ballots from the bag, carefully ensuring that the ballots remained in the same order stored. They would then start counting the ballots from the top of the stack until they reached ballot number 412.

When ballot 412 was removed from the stack, it was replaced with a bright pink slip sheet containing the date, batch or precinct number, and the count number for that specific ballot. A removable label was placed on the ballot that contained the precinct number, the word “Count” to indicate the count-down method was used, and the count number. The labeled ballot was placed in a tray, and the audit team moved on to the next ballot on the list. If the ballot retrieval list called for another ballot from the same batch, the counting continued from the point where the first ballot in the batch was removed. For the batches of absentee ballots, the slip sheets and labels also included the 4-digit CVR numbers imprinted on the ballots.⁶

K-CUT METHOD

The *K*-cut method involved cutting the batch of ballots like a deck of cards six times and selecting the ballot on top of the batch after the sixth cut. To determine the point at which the batch was cut each time, an app was used to randomly generate a number between 1 and 100. That number was converted to a percentage, and the audit team estimated the point in the batch that matched the percentage, starting from the top. For example, if the random number generated was “33”, the cut point was a point in the stack approximately 33% down from the top. To make the cut, the ballots from the cut point up (the top portion) were lifted off the stack and placed on the table next to the stack from the cut point down (the

bottom portion). The bottom portion was then placed on top of the top portion, completing the cut. This process was done six times, and the ballot on top after the sixth cut was removed, labeled and placed in a tray. A slip sheet was placed on top of the batch, indicating which ballot was removed and the date. If the ballot retrieval list called for more than one ballot from a batch, the six-cut process described above was performed again, and the resulting top ballot labeled and removed. Ballots retrieved using the *k*-cut method were labeled with the precinct number, the word “K Cut,” and the count number. Even though the count number was not actually used to retrieve the ballot, it was included on the label so that the ballot could be matched with the corresponding ballot on the retrieval list.

One technique that made the *k*-cut method easier was using a narrow table that allowed one audit team member to stand on each side, in line with the stack of ballots. The auditors could hold onto opposite ends of the ballot stack and lift to make the cuts. This technique was particularly useful with large batches and when the cut point percentage was high, resulting in the need to lift hundreds of ballots at times while keeping the stack in order.

After the ballot/ballots were retrieved from a batch, the audit teams placed the batch back in the ballot container, updated and signed the Ballot Container Certificate, and resealed the container.

In general, the count-down method was used when the retrieval list called for a ballot with a Count # less than 300, and the *k*-cut method was used when the count number was greater than 300. All absentee ballots were retrieved using the count-down method, as the batches of absentee ballots were smaller on average than the batches of Election Day ballots.

Review of Ballots

After the ballots were retrieved the four audit teams convened in the City Commission Chambers to review the ballots. One team reviewed and counted the eight absentee ballots using a tally sheet designed for the ballot comparison audit. The other three teams reviewed and counted the 32 Election Day ballots using tally sheets designed for the ballot-polling audit. Both tally sheets contained the candidates for the three statewide offices as listed on the ballots (governor, secretary of state, and attorney general) and spaces under each office to indicate if a write-in vote was cast or if the contest was over/under voted. The difference between the two tally sheets was the comparison audit sheet had an additional column for the CVR number and could be used to tally four ballots, while the ballot comparison sheet did not have a CVR column and could be used to tally six ballots.

The audit teams reviewed the ballots using a call-and-switch method where one team member interpreted the ballot and called out the result for each of the statewide

⁶ If an absentee ballot needed to be rescanned during the tabulation process, multiple CVR numbers were printed on the ballot. The highest number is the number that would have a matching CVR in the EMS.

offices (governor, secretary of state, and attorney general), and the other team member marked the tally sheet with those results. Then the team members exchanged the ballot and tally sheet, and the member who had marked the tally sheet called out the results from the ballot, and the other team member reviewed and checked the tally sheet for accuracy.

After all the ballots had been reviewed, interpreted, and tallied, the ballot information and results from the absent voter ballots were entered into the SUITE RLA tool, which used the imprinted CVR number to match and compare the results from the audited ballots with the corresponding CVR data that had been loaded into the software prior to the audit event. For the Election Day ballots, the results from the tally sheets were manually totaled and the totals entered into the SUITE RLA tool. The results from all ballots were entered into the Bayesian audit tool as well.

The result of the Pilot RLA was a 3.7% risk limit for the Hybrid comparison/polling RLA, which was less than the target risk limit of 5%. For the Pilot Bayesian Hybrid Audit the upset probability was 0.03%. At the close of the audit the original ballots and tally sheets were sealed in a separate container.

Issues/Challenges and Solutions/ Lessons Learned

LESSON LEARNED: RLAs ARE NOT DIFFICULT; IMPRINTING MAKES BALLOT COMPARISON AUDITS INCREDIBLY EASY

From a procedural perspective, RLAs are not difficult. Ballot-polling audits require no pre-election preparation and the execution of the audit is like a hand recount but with significantly less work. Ballot comparison audits can require significant organizational work for the teams processing ballots on or prior to Election Day, but that work is almost eliminated when CVR numbers are imprinted on the ballots and used to match the ballots with their cast vote records.

CHALLENGE: ELECTION MANAGEMENT SYSTEMS AND DECENTRALIZED ELECTION ADMINISTRATION

Michigan administers its elections on the local level, with city and township clerks running the elections for their jurisdictions in coordination with their county clerks. One of the divisions of responsibility in this enterprise is the separation of the voting system, which is owned and operated by the cities and townships, from the software that programs the equipment, which is owned and operated by the county clerks. That same software (generically referred to as the Election Management System or EMS) collects and aggregates the election results and is needed to output most of the reports and data the voting system can produce, including the Cast Vote Record (CVR). As a result of this division of responsibilities, important

elements of executing the Pilot RLA were out of the local clerk's control and beyond the local clerk's knowledge. But county clerks cannot conduct RLAs by themselves either, as the local clerks have custody of the ballots – the audit trail an RLA examines.

LESSONS LEARNED:

RLAs cannot work in Michigan without collaboration between county and local clerks. The pilot RLA did not teach a new lesson in this area as much as it reinforced an existing reality: that county, city, and township clerks must work together for the election system in Michigan to work for voters. A second lesson in this area is that county and local clerks need to learn more about the data available through the EMS software and how that data can be used to improve transparency in the electoral process. Vendors can play a role through training opportunities, and the Bureau of Elections can play a role by providing direction and best practices for how tools like CVRs and ballot images can be used to improve election administration processes.

ISSUE: WRITE-IN VOTES AND THE BALLOT COMPARISON METHOD

In a ballot comparison audit, the ballots in the sample are compared to their Cast Vote Records, which show how the voting system interpreted each ballot and counted their votes. In Kalamazoo's AVCB there was no written record of how votes for valid write-in candidates were counted for *specific ballots*. The CVR will show that the voting system tallied a "write-in" vote for a particular contest, but human examination is required to determine if the name written on the ballot is the name of a valid write-in candidate. If a written-in name appears to be the name (or variation) of a valid write-in candidate, that name is recorded on a tally sheet for that precinct as it appears on the ballot. Other votes for the same name are indicated with tally marks, and variations of the name are recorded on the tally sheets (1 per precinct) in the same manner.

The tally sheets are included in the poll book and submitted to the County Board of Canvassers on election night. It is the Board of Canvassers that determines whether the names recorded on the write-in tally sheets are names of valid write-in candidates, and it is the Board of Canvassers that manually adds those totals to the official election results, usually days after the election. Even if the AVCB created a record of specific ballots with write-in votes and names tallied for each specific ballot, it would be impossible to know how a specific write-in vote on a specific ballot was handled by the Board of Canvassers. The count for that ballot could not be verified. For the purposes of the pilot RLA this write-in issue was a non-issue, as there were no valid write-in votes cast in contests being examined among the absentee ballots.

But it could potentially be an issue for ballot comparison audits in elections and contests with a competitive write-in campaign.

SOLUTIONS:

A potential solution for ballot comparison audits is for the AVCB to create a record of ballots with write-in votes and the name(s) tallied for each specific ballot; and for the Board of Canvassers to use the same report to record the disposition of those write-in votes. An alternate solution is to use the ballot-polling method which does not rely on a comparison of individual ballots with the CVRs to conduct an RLA and a Bayesian audit. In both of these audits, the auditors' interpretation of the write-in vote on individual ballots would be based on the State of Michigan's definitions for valid votes. In elections and races where there is a strong write-in candidate, it might be wise to use the ballot-polling RLA or a ballot-polling Bayesian audit.

ISSUE: ALIGNING THE RLA PROCESS WITH THE CANVASS AND RECOUNT PROCESSES

In "A Gentle Introduction to Risk-limiting Audits," the authors state, "As long as the audit does not yield sufficiently strong evidence [that a full hand count would confirm the original (voting system) outcome], more ballots are manually inspected, potentially progressing to a full hand tally of all the ballots."⁷ Later the article states, "A risk-limiting audit amends the outcome if and only if it leads to a full hand tally that disagrees with the original outcome." The notion of an RLA amending the reported election results might be technically possible, but this cannot happen in practice unless there is a legal basis that grants authority for the RLA results to replace the reported results. One of the issues that needs to be addressed if Michigan implements RLAs is reconciling the RLA process with the existing recount process and achieving alignment between the two. At what point does an RLA become a recount? Should there be an automatic trigger, or does a human agent need to intervene? Should a good audit result preclude a candidate or party from requesting a recount? These are questions that need to be answered.

A related question that needs to be answered is whether an RLA should take place prior to the canvass and certification of the election results or after the canvass. Conducting an RLA prior to the certification of results would allow the audit results to be factored into the certification decision. In addition, if the audit grew into a full recount, the recount would be a proactive step

on the part of election officials to ensure the results were correct in the first place rather than requiring candidates to initiate a recount.

LESSON LEARNED: THE K-CUT METHOD OF BALLOT RETRIEVAL FOSTERS EFFICIENCY AND ACCURACY

Regardless of whether the *k*-cut is found to be a *mathematically valid* way to randomly select ballots, it is a *faster* procedure and less fatiguing compared to the "count-down" method when dealing with large batches of ballots. In addition, the *k*-cut method eliminates the risk of counting errors when dealing with a large stack of ballots to count through.

Recommendations

- Michigan should implement RLAs on a statewide basis. An RLA can provide a level of assurance, with a strong statistical foundation, that reported election results are not incorrect. Moreover, the strength of the statistical evidence provided by an RLA can approach the level of evidence provided by a full hand recount, but with only a fraction of the effort. RLAs would complement Michigan's existing security measures by reviewing the results of individual contests.
- The ballot-polling audit is the RLA type that should be implemented in Michigan. The level of ballot organization required for a comparison audit cannot be achieved on Election Day using the precinct-based tabulators on which most Michigan ballots are currently tabulated. Ballot-polling audits require no preparation prior to, during, and after an election beyond the ballot handling and chain of custody procedures local clerks should already be following. This makes it an ideal method for implementation across the varied environments and capacities represented by election jurisdictions in the state, of which there are more than 1,500.
- Under the Secretary of State's authority and mandate to establish a post-election audit program, the role of the Michigan Bureau of Elections should be to establish standard processes and procedures for RLAs across the state. In addition, the Bureau should create standardized forms and instructions that use language and terminology that county and local clerks in Michigan will understand. Finally, the Bureau is in the best position to acquire the software needed to conduct RLAs.
- RLAs in Michigan should be implemented at the county-level. Beyond the need for statutory compli-

⁷ Mark Lindeman and Philip B. Stark, "A Gentle Introduction to Risk-limiting Audits," *IEEE Security and Privacy Special Issue on Electronic Voting* (2012): 1. Available at <https://www.stat.berkeley.edu/~stark/Preprints/gentle12.pdf>.

ance, there are practical reasons why this should be a county-based implementation. From an administrative standpoint, the Bureau could more easily manage a process conducted by 83 county election officials than a process conducted by the more than 1,500 municipal election officials. County clerks own and operate the EMS software and are able to coordinate audits with the local clerks in their counties.

- The conversation between election administrators, academic specialists, and advocacy groups must continue. There needs to be a national forum where these stakeholders can come together to develop and improve standards and best practices for post-election audits, including the RLA and Bayesian processes. It is in the best interests of election administrators to support and understand the technical demands of the academic specialists, as strong, rigorous procedures and statistical methods will be better able to detect incorrect results and will be more defensible to skeptics. Election administrators need to be committed to continual improvement in all election processes, including post-election audits. In turn, academic specialists and advocacy groups need to support election officials by understanding the demands and realities of managing elections in the field. And election administrators need to be able to implement processes and methods that are accepted as valid by the specialists even as those specialists are continually working to test and improve those methods and processes.
- Risk Limiting Audits and Bayesian Audits should be used together. The fact that the two audits can share the same ballot sample in most cases means they can be executed simultaneously with little extra work for election administrators. And their methodological differences can actually be a source of strength when applied to the same data. Bayesian audits are explained in the Bayesian Audit Overview section below.
- Election administrators, academic specialists, and advocacy groups must continue to work on the language and terminology used to communicate with the public regarding RLAs. This starts with election administrators attempting to learn, even if imperfectly, some of the math and statistical principles behind RLAs. Can election officials execute RLAs without understanding the statistical foundations? Yes. But election officials are in the best position to help the academic specialists communicate to the public about RLAs. Moreover, engaging with the specialists to improve the RLA process requires a basic level of understanding about the process. For their part the academic specialists need to develop better, simpler, more visual, engaging ways to communicate the principles behind risk-limiting audits and what the audit results indicate. If the principles behind RLAs remain conceptually inaccessible to the average citizen, the process will not be transparent even if the audit tools use open-source software and the statistical models are published for everyone to see.

City of Lansing

Chris Swope, Lansing City Clerk

Overview

On Tuesday, December 4, 2018, the Lansing City Clerk's Office hosted a Pilot Risk Limiting Audit (RLA) of the results of the General Election held on November 6. The audit took place at the Clerk's Office, began at approximately 9:00 a.m., and ended at approximately 6:00 p.m. Approximately 75 people were in attendance, including city staff, audit team members, pilot project partners, multiple Michigan election officials, and observers. The Lansing RLA was a *hybrid* RLA that combined the results of a ballot-polling audit and a ballot comparison audit. In addition, the same sample of ballots that was used for the RLA was also used for a Bayesian audit, which was conducted simultaneously. The target race was the race for Court District 54A Judge, but all three statewide races (governor, secretary of state, and attorney general) were tallied.

General Information About the November 6, 2018 Election in the City of Lansing

- 45 voting precincts at 33 polling locations
- Official number of registered voters: 81,746
- Absentee Voting Counting Board ("AVCB") opened at 7:00 a.m.
 - AVCB used one Dominion ICC G1130 High Speed Scanner System to tabulate absentee ballots
- Votes cast absentee: 10,914

Clerk Overview

Today's reality requires election administrators to be even more vigilant regarding the integrity of our elections. Auditing election results is necessary to ensure public confidence in the election process. Risk-limiting audits (RLAs) have the potential to provide a more effective manner of audit by increasing confidence that ballots are being tabulated correctly. The City of Lansing was delighted to participate in pioneering this new tool in an effort to further secure Michigan elections.

Although we had no idea what to expect from this process and participating in the pilot program required some extra time and attention during an already busy election season, overall this project was a great learning experience for all involved and an important step forward for Michigan.

Clerk Preparation and Workload

Checklists outlining the tasks necessary before and during audit were not provided and would be very helpful in the future. Also, receiving information on the approximate number of ballots to be pulled and examined sooner would have been useful in preparing and staffing the audit properly.

We were briefed and prepared to keep the absentee ballots processed on Election Day in the same order that they were put into the high-speed scanner for tabulation. However, this was difficult at times during a very long and busy Election Day.

While interesting, the academic, mathematical component of RLAs is far less important to clerks than the practical considerations, such as staffing needs and training materials. For example, a flow chart of the process, diagrams of how to set up teams, and list of equipment and supplies needed would have been much more helpful to election administration than lessons in statistical theories. An overall idea of how things will function, with a short explanation of the mathematical theory behind it, would suffice.

Communications and Messaging

RLAs could increase efficiency in the audit process and reduce the financial resources and time necessary to conduct a post-election audit. RLAs have the possibility to minimize the number of ballots that must be examined during an audit. Under Michigan's current post-election audit criteria, a full hand recount is required in randomly selected precincts. RLAs could serve as an important compliment to the already existing audit procedures in Michigan. Pairing our current election security measures with a robust post-election audit process designed to serve as a check on election results is a common-sense step forward.

Challenges and Solutions

Not surprisingly, human error resulted in some challenges during the ballot comparison portion of the audit. Specifically, human error impacted the retention of the absentee ballots in the ordered tabulated. As explained above, this step is essential for an effective ballot comparison audit. This audit method involves the random selection of a ballot from the ballot manifest by the SUITE tool (for example, Ballot No. 312 in Batch 23 may be the selected random ballot). While the CVR obtained by election officials prior to the audit identifies how the voting equipment tabulated Ballot No. 312, this information is not shared with the auditors prior to the hand review to

prevent bias when they interpret the markings on the ballot. To retrieve this ballot, audit officials first collected the ballot bag with the ballots from Precinct 23. The bag was then opened and the ballots stacked in the bag were removed and set on the audit table. Audit officials would then count-down the stack of ballots until they reached Ballot No. 312 in the stack.

Though obvious in retrospect, clearly designating the top of the ballot stack is essential to ensure that audit officials use the same No. 1 ballot as the election officials who processed and batched the absentee ballots. Here, we learned that retaining ballots in the same order scanned is of limited value if audit officials are not able to determine which ballot is properly designated No. 1 because the top and bottom of the stack of ballots is not clearly designated.

Further, it was difficult for our teams to count to the correct ballot using the count down method. Clearer directions and training will alleviate some of these issues.

Lansing's equipment does not currently have the necessary technology to imprint on the ballot. Without imprinting, proper paper management will remain a challenge. It was difficult to verify if our teams had pulled the right ballot for comparison to the Cast Vote Record (CVR).

Another challenge was related to the information needed for the pilot, specifically the cast vote record. As a city clerk, I administer elections using a voting system purchased by the County Clerk. She is responsible for various election administration duties, including aggregating election results from across the county. After contacting the vendor, I discovered that as a city clerk, I do not have access to the CVR, only the county clerk does. This reality, and the novelty of my request as CVRs have not been needed previously in Michigan, caused some delays in obtaining usable CVR data. In the future, we will need to plan accordingly with the County Clerk and the vendor to retrieve the CVRs in a timely manner.

While counting to the correct ballot can be difficult and caused issues during the ballot comparison portion, using the *k*-cut method during the ballot-polling portion was much easier. The only issue is the size and weight of

the ballots. It was suggested for the longer, heavier ballots that a team of two lift from each end to perform the *k*-cut.

Finally, we did not establish a clear procedure on tallying results with our ballot-polling teams. Some teams were pulling ballots and tallying as they went, while others were pulling the ballots and tallying all at the end. This should be determined ahead of time and incorporated into the training.

Recommendations

An immediate attempt to implement a statewide RLA may result in some confusion. Many would have several reasonable questions about the nature and procedure of the RLAs, as well as practical considerations about their cost in time and money. Some of these questions can only be answered by conducting or observing an RLA pilot. Additional pilots at municipalities of various sizes should be performed in order to fully understand the impact RLAs will have statewide and to allow local and county clerks to attend or participate. Observing an RLA provides the best insight into how it works. We realized the procedure was not as complicated as we thought once we saw it in action.

Timing may be the most significant factor in implementing an RLA. Election officials work on election matters for several months before and after each election, and have few resources to implement new processes during that time period. It may be best to pilot this during odd-year or smaller elections.

In addition, a uniform checklist and uniform terminology should be put in place before rolling out the next phase. Using common terminology and avoiding the more technical aspects of creating the procedure would be wise. While the process used to create the system is fascinating, most of it went over the heads of the participants. These changes will help streamline the process and provide greater clarity to new participants.

Local municipalities, under the direction of the State of Michigan Bureau of Elections and with the assistance of their respective County Clerks, should conduct the RLA.

Rochester Hills

Tina Barton, Rochester Hills City Clerk

Overview

On Monday, December 3, 2018, the Rochester Hills City Clerk's Office hosted a Pilot Risk Limiting Audit (RLA) of the results of the General Election held on November 6th. The audit took place at the Rochester College, began at approximately 9:00 a.m., and ended at approximately 5:00 p.m. Approximately 40 people were in attendance, including city staff, audit team members, including multiple Michigan election officials, pilot project partners and observers. The Rochester Hills RLA was a ballot-polling RLA. In addition, the same sample of 76 ballots that was used for the RLA was also used for a Bayesian audit, which was conducted simultaneously. The target race was Proposal 3, but all three statewide races (governor, secretary of state, and attorney general) were tallied. A 5% risk-limit was used.

General Information About the November 6, 2018 Election in the City of Rochester Hills

- 32 voting precincts at 22 polling locations
- Official number of registered voters: 53,784
- AVCB opened at 7:00 a.m.
 - AVCB used 4 Cannon High Speed Scanners each with their own Verity Central computers (Hart Equipment) to tabulate absentee ballots
- Absentee ballots cast and counted: 12,924

Purpose

As the City Clerk for Rochester Hills, I chose to take part in the Risk Limiting Audit because the audit practice that Michigan currently uses places a significant emphasis on actions taken before Election Day, with a minimal number of precincts throughout the state undergoing a full hand recount. The Risk-Limiting Audit (RLA) adds a check on the outcome of the election to the current post-election audit process. The validity of the results is equally as important as the pre-Election Day responsibilities.

In December 2016, in an article discussing low voter turnout, the Washington Post flagged that “a Gallup poll two weeks before Election Day found that only one-third of Americans (35 percent) were ‘very confident’ that their

vote would be counted accurately.”⁸ Two years later, the Pew Research Center reported that “55% (of voters surveyed) said they were not too confident or not at all confident that elections would be secure.”⁹

Many states, including Michigan, currently have some sort of post-election audit process in place. Michigan's audit is considered a “performance-based audit.” This audit requires election officials to affirm that various election administration procedures were properly conducted. For example, clerks must verify that a logic and accuracy test was completed, notices were published properly and correct seal numbers were recorded. What these audits are lacking is Paul Harvey; they are lacking “the rest of the story.” A risk-limiting audit provides the rest of the story about whether votes were counted as cast. Did the program recognize the ballot markings correctly? Did the winner actually win? Are we confident in our count?

I cannot stop cyber threats. I cannot give a 100% guarantee that a program was not hacked. However, I can pull the paper ballots from a certified sealed container and perform an audit on the results. This is why I chose to take part in the Michigan Risk-Limiting Audit Pilot.

Communication and Messaging

All three clerks agreed that the initial message sent out to voters and the press should be in the form of a joint press release. By coordinating our messaging, we not only learned from one another, but we also ensured that we were providing the same accurate information to voters and the media. The press release gave an overview of what an RLA is and which communities would be participating in the pilot. Once we agreed to the press release language, I sent the release to my local press contacts. Additionally, I used social media to help ensure that voters and the public were provided with information about the pilot. I specifically approached this outreach from an “are

⁸ Pippa Norris, Holly Ann Garnett, and Max Grömping, “Why Don't More Americans Vote? Maybe Because They Don't Trust U.S. Elections” *Washington Post*, Dec. 26, 2016, <https://www.washingtonpost.com/news/monkey-cage/wp/2016/12/26/why-dont-more-americans-vote-maybe-because-they-dont-trust-u-s-elections/>.

⁹ “Most Voters Have Positive Views of Their Midterm Voting Experiences,” *Pew Research Center*, Dec. 17, 2018, <http://www.people-press.org/2018/12/17/most-voters-have-positive-views-of-their-midterm-voting-experiences/>.

you curious about RLAs?” angle.

When local and state officials make public announcements about RLAs, I recommend including a link to the [video](#) recently released by the Colorado Secretary of State that provides an easy-to-understand primer on RLAs to all RLA-related announcements and press releases.¹⁰ This video also explains how Colorado election officials successfully worked with one another and other interested parties to develop and implement the RLA process. It is a great voter education tool that can also be used for press outreach.

Preparation

The State of Michigan Bureau of Elections contacted our office in the summer of 2018 regarding the possibility of piloting an RLA in Michigan. Soon after, Brennan Center Counsel Liz Howard made a personal visit to discuss the process. From those first general conversations, we had approximately ten phone conferences, we met in-person as a group, and all three communities made a visit to Denver to observe their RLA after the November election. The preparation for the RLA included an education component covering the math underlying both procedures, how the risk is determined, and how ballots should be processed and stored. We did not prepare for the audit any differently than we do for the current election and audit process.

¹⁰ “Colorado’s Risk-Limiting Audit,” Colorado Secretary of State Office, published June 15, 2018, on YouTube, <https://www.youtube.com/watch?v=ysG4pFFmQ-E>.

Process

The City of Rochester Hills processed over 36,600 ballots on Election Day. This is a combined total of absentee ballots, as well as precinct ballots. Proposal 3 was chosen to audit due to the high margin of difference between the “yes” and “no” votes, which stood at 29%. This information was entered into the SUITE RLA Tool which calculated that 76 ballots were needed for the pilot.

Rochester Hills uses Hart equipment. The scanner unit for the precinct sits on a large black box. Ballots go through the scanner and fall randomly into the box. Our absentee ballots are counted on a high-speed machine. Each batch that is processed is done so by precinct number, batch number, scanner ID, and then counted into groups of 50 ballots for each batch. As neither our in-precinct nor our absentee voting equipment created a usable cast-vote record (“CVR”), the ballot-polling method was used for all ballots in the Rochester Hills pilot.

Approximately 40 election officials attended our pilot – local, county, and state offices were represented. Local clerks from Oakland County were chosen to serve on the teams that performed the audit. Bureau staff provided a brief RLA presentation at the beginning of the pilot. After the Seed Selection Ceremony, the random seed was entered in RLA SUITE tool.

As part of our audit pilot, we tried out a new method to randomly select ballots to review for the audit – the *k*-cut method. MIT’s Prof. Ron Rivest and Mayuri Sridhar created the *k*-cut method.

Risk-Limiting Audit Overview

Kellie Ottoboni

Math and Tools

These pilots were the first time the SUITE method for hybrid RLAs has been used in practice. SUITE is a general method for conducting RLAs using *stratified samples* of ballots, where ballots are divided into distinct non-overlapping groups (or *strata*) and samples of ballots are drawn independently from each group. Audits can essentially be conducted separately, but in parallel, for each stratum of ballots, with the results from all *strata* combined at the end to produce a single measured risk for the entire contest. In particular, SUITE is useful for auditing contests that include jurisdictions with heterogeneous voting equipment. Ballot polling can be conducted alongside ballot-level comparisons of ballots cast on equipment that produces ballot-level CVRs. The math requires minor modifications to existing ballot polling and ballot-level comparison RLAs. In Kalamazoo and Lansing, there were two *strata*: 1) AVs (absentee ballots for which CVRs can be linked to specific ballots) and 2) Election Day ballots (ballots cast for which CVRs cannot be linked to specific cast ballots). When using just one stratum, the SUITE tool reverts to the single RLA method of choice. (In Rochester Hills, we used the SUITE tool to facilitate the ballot-polling method for all ballots.)

Kellie Ottoboni and Dr. Philip Stark created the software tool for SUITE. The code is written in Python and displayed in an interactive Jupyter Notebook. The tool can be run interactively at https://mybinder.org/v2/gh/pbstark/CORLA18/master?filepath=code/suite_toolkit.ipynb.

This was intended to be a prototype user interface for these pilot RLAs rather than an industrial-strength tool; the hope is that software developers will be able to base a proper program on our prototype.

The SUITE tool calculates the necessary pieces of information for each step of the audit:

- It estimates the initial sample sizes needed in each stratum to stop the audit, assuming that the reported results were correct and that one-vote overstatements occur at a very low rate in the ballot comparison stratum.
- It uses the SHA256 PRNG from the [cryptorandom package](#) to draw the random samples of ballots in each stratum.
- It reads in a ballot manifest for each stratum as separate CSV files, then determines which ballots in which batch need to be pulled based on the samples

that were drawn previously, and finally exports this data to a CSV file to be printed.

- It takes the statistics from each sample (the number of 1-vote and 2-vote overstatements and understatements in the comparison stratum, and the number of votes seen for each candidate in the polling stratum) as input and runs the risk calculation for each pair of reported winners and reported losers.
- If the data do not provide sufficiently strong evidence that the reported winners really won, it estimates how many more ballots need to be sampled in each stratum to confirm the reported outcome, assuming that the rates of discrepancies and the rates of votes for each candidate will continue to reflect those seen in the initial samples.
- It runs each of these steps again for a second round of sampling.
- It logs each step of the process in JSON files.

Procedures

This section describes the audits in each city and the differences between them. Since the goal of these pilots was to gain hands-on experience, instruct local election officials, compare procedures, and identify bottlenecks in the processes, we opted to reduce the population of ballots under audit to each city. The sampling frame for a true RLA includes all ballots cast for a particular contest; none of the contests under audit were entirely contained in the sampling frame of ballots. Each audit pretended that the ballots we had access to comprised the entire contest. Reducing the sampling frame made these pilots feasible to conduct in a single day and illustrated RLA procedures to a wide audience.

ROCHESTER HILLS

The Rochester Hills pilot RLA took place on December 3. About 40 county and local election officials from around the state attended to observe the audit process. We audited Proposal 18-3, a statewide proposal to add new voting policies to the Michigan constitution. In Rochester Hills, “Yes” votes won with a 29% margin.

Rochester Hills uses Hart InterCivic Verity tabulators, which do not provide CVRs that can be used for ballot comparison audits. Thus, we used only ballot polling for the audit. We did not need to use different software for

this: The SUITE tool reverts to pure ballot polling when the CVR stratum size is 0 ballots.

The SUITE tool estimated that we'd need to look at 76 ballots to achieve a risk limit of 5% if the reported results were correct. We had 5 two-person audit boards pull the ballots from ballot bags and sample them either by counting down to the ballot that the SUITE tool sampled or by using *k*-cut to select a ballot from the stack. Audit boards used a mix of counting and *k*-cut to sample ballots: if the SUITE tool sampled a ballot that would require counting more than 200 ballots to find it, the team used *k*-cut instead.

It took approximately 2.5 hours for the audit boards to pull the ballots and record them, and another half hour for a separate team to tally them.

The sample contained 50 “yes” votes and 26 “no” votes. The measured risk based on the sample was 2%.

LANSING

The Lansing pilot RLA took place on December 4. We audited the 54-A District Court Judge race using a 9% risk limit. Cynthia Ward won over Ayanna Neal by about a 10% margin. Though this contest was contained within the city of Lansing, we reduced the sampling frame of ballots for reasons stated below, making this a pilot of RLA procedures rather than a true RLA.

First, there were several ballot bags that could not be opened. Due to a pending recount of a Lansing school board election, some ballots were not available for review during the scheduled RLA pilot. Specifically, any ballot bag containing ballots from Precinct 45 could not be opened. These bags contained ballots from other precincts as well. One way to handle this would have been the “phantoms to zombies” approach, treating each inaccessible ballot in the sample as a vote for the reported loser.¹¹ Because the point of the pilot was to illustrate procedures rather than to obtain a precise risk measurement, we opted to remove ballots from the ballot bags containing Precinct 45 from the sampling frame.

Second, the SUITE tool estimated that the initial sample size needed would be over 500. Based on observing the sampling in Rochester Hills the previous day, we knew that it would not be possible to examine 500 ballots before the close of business day. To reduce the sample size needed, we reduced the sampling frame: It was therefore decided to remove approximately 30 Election Day precincts from the ballot manifest and the reported vote totals.

(Another approach could have been to draw a fixed number of ballots, say 100, from the entire population and simply measure the attained risk of the sample. This approach would not have attained the desired 9% risk

limit, but would have been equally illustrative.)

This changed the fraction of ballots with a CVR from 23.6% to 50%, reducing the workload estimate for hybrid SUITE. This did not substantially affect the reported margin between Ward and Neal amongst Election Day votes. The initial sample size estimate decreased from over 500 to 260, with 130 ballots from the absentee votes and 130 from Election Day votes.

Lansing used 6 two-member audit boards. It took about 4 hours to pull and record all the ballots, then another hour to compare AV ballots to the CVR and tally the Election Day sample. The audit boards used counting down to sample AV ballots, because AV ballots were linked to their corresponding CVR by position: the batches of AV ballots and the CVR file were both supposed to contain ballots in the order in which they were scanned. They used a mix of counting and *k*-cut to sample Election Day ballots.

Among the 130 sampled AV ballots, we identified 15 discrepancies between the paper ballots and the CVR. It is unlikely that these were true discrepancies; we believe that the sampled paper ballots were matched to the incorrect CVR because of either counting errors in sampling the ballots or noncompliance with one of the many new paper ballot management policies implemented in an effort to retain the ballots in the same ordered scanned for this pilot. With so many AV discrepancies, the SUITE risk was 100%.

For comparison, we pooled the AV and Election Day samples for a pure ballot-polling audit. One AV ballot was sampled three times, reducing the overall ballot polling sample size to 258 ballots. In total, there were 116 votes for Ward, 94 votes for Neal, and 48 invalid ballots or write-ins. The initial sample size estimate for ballot polling was 285 ballots, just higher than the sample size actually drawn. Overall, the risk calculated using SUITE ballot polling was 87%. (The risk calculation using BRAVO ballot polling is 38%. This difference likely occurs because there were so many votes for neither Ward nor Neal in the sample.) In a true RLA, this would have led to a second round of sampling ballots.

KALAMAZOO

The Kalamazoo pilot RLA took place on December 5. We audited the governor's race,

pretending that the governor's race was entirely contained in Kalamazoo city. In this contest, Gretchen Whitmer beat runner-up Bill Schuette by a margin of 54%.

The SUITE tool estimated that we would need to pull 33 ballots to achieve a 5% risk limit. Based on our experiences the previous two days, we knew this would not take long. To ensure that we would meet the risk limit and to give each audit board extra hands-on experience, we increased the total sample size to 40 ballots, 8 from the AV ballots and 32 from Election Day. It took 4 two-member audit boards about 1.5 hours to sample and tally the ballots.

11 Jorge H. Banuelos and Philip B. Stark, “Limiting Risk by Turning Manifest Phantoms into Zombies,” arXiv:1207.3413, Jul. 14, 2012. Available at <https://arxiv.org/abs/1207.3413>.

The AV ballots were imprinted with ID numbers, so the comparison audit involved looking up the imprinted ID in the CVR file. They did not rely on preserving the order of ballots as in Lansing, so a mix of counting and *k*-cut were used to sample both AV ballots and Election Day ballots.

There were no discrepancies in the AV sample. In the Election Day sample, there were 23 ballots for Whitmer, 8 ballots for Schuette, and 1 for Gelineau.

Lessons Learned

Ballot-level comparison audits are the most statistically *efficient* type of RLA: of all existing RLA methods, they require examining the fewest number of ballots if the reported margin is correct. We conjectured that the SUITE hybrid RLA would also be efficient, since it uses ballot-level comparison for a portion of the ballots. However, we found that it is actually less efficient than

ballot polling alone when the majority of ballots do not have a CVR. In Lansing and Kalamazoo, the proportion of ballots with a CVR was below 25% and the hybrid RLA required sampling a few more ballots than ballot polling alone would have.

RLAs are meant to be flexible; the only absolutely necessary element required to conduct an RLA is a trustworthy paper trail. The choice of auditing method comes down to efficiency. Election officials should weigh efficiency of the statistical method – that is, the number of ballots that a method requires to be examined - against the efficiency or feasibility of the process. For instance, we learned that for Lansing, ballot polling might be the most efficient option: ballot-level comparison is possible but difficult and SUITE is not necessarily more efficient than ballot polling.

Bayesian Audit Overview

Mayuri Sridhar

Math/Procedures

Bayesian audits were first proposed by Rivest and Shen in 2012. These audits use a statistical model of voter preferences on the cast ballots (based on an assumed “prior” probability distribution of voter preferences and data from the ballots seen so far in a random sample of the ballots) to answer the question:

If one were to examine all of the of the ballots, what is the probability that the winner would be someone other than the reported winner?

This is known as the upset probability.

Bayesian audits are very similar to risk-limiting audits, which provide the following subtly different assurance:

If the reported winner is not the same as the winner that would be seen if all of the cast ballots were to be manually examined, then the risk-limiting audit has a high probability of manually examining all of the ballots.

Bayesian audits can be run “in parallel” with standard risk-limiting audits. That is to say, the risk-limiting audit can do its sampling and compute the attained risk (the chance that an incorrect outcome would escape detection by the audit) and the Bayesian audit can use the same sample to compute its estimated upset probability.

These two types of audits thus have nearly identical structures and very similar goals. It is easy to have the Bayesian “piggy-back” on the sampling work being done by the RLA and to do both audits at the same time.

The Bayesian audit uses a mathematical model based on an assumed “prior” distribution of voter preferences and the ballots seen in the sample to estimate the probability that if all ballots were examined the winner would be found to be other than the reported winner. This is the called the *upset probability*. The model computes these “posterior” probabilities using a model based on the Dirichlet-Multinomial statistics. For more details, see Bayesian Tabulation Audits: Explained and Extended, <https://arxiv.org/abs/1801.00528>.

(As a technical aside for those familiar with the Bayesian method, which can be ignored by other readers: The method used a “prior” of a pseudocount of 1 for each candidate for a ballot-polling audit. For ballot-comparison audits the method used a pseudocount of 50 for the case that the reported choice for a ballot was equal to the

actual choice as observed in the audit, and a pseudocount of 1 otherwise. This prior is chosen because it is unbiased with respect to the candidates, and relatively weak—the effect of the prior on the upset probability estimates is quickly swamped by the data in the audit sample.)

In short, the Bayesian method runs a number of simulation trials on the computer based on assumptions about voter preferences, expanding the sample until it has the size of the set of cast ballots, using the statistical model to guide the expansion. The fraction of the trials for which the reported winner doesn’t win the trial is an estimate of the upset probability.

The Bayesian audit takes as input an upset probability limit, just as a risk-limiting audit takes as input a risk limit. If the estimated upset probability is less than the upset probability limit, then the Bayesian audit may stop and accept the reported winner as being the actual winner. The Bayesian audit also requires as input an assumed “prior” probability distribution for voter preferences. Different priors generally result in different upset probabilities.

Tools

Procedurally, the Bayesian audit and the Risk-Limiting Audit follow the same general structure. In particular, both audits require a ballot manifest, an initial sample size estimate, and a random seed. Let us say that the initial sample size is 100 ballots. Then, we require software which takes as input the ballot manifest and the seed and creates a sampling plan, which describes which ballots to sample. Once the sampling is done and the votes on the ballots have been interpreted, we can compute whether the audit stopping condition is satisfied. For Bayesian audits, this involves software to run simulations and estimate the Bayesian upset probability. From here, if the audit stopping condition has not been met, then we would use software to estimate how many additional ballots we need to sample and then escalate the audit. In practice, during the pilots, we did not consider escalating the audit.

The Bayesian audit, like an RLA, needs to randomly sample the cast paper ballots. Indeed, the sampling method is exactly the same as for an RLA, and the Bayesian audit and the RLA can utilize the same audit data obtained by manual examination of the randomly chosen ballots. Which ballots are examined in the sample is determined by a pseudorandom number generator that takes an input a “random seed” – a 20 digit number.

To generate a random seed, we used 10-sided dice to generate a 20-digit random number. We used the SUITE software, designed by Kellie Ottoboni and Dr. Philip Stark,

to estimate the initial sample size, based on the reported results of the election. The SUITE software was also used to produce a sampling plan for the first stage of the audit. For the Bayesian upset probability calculations, we used BCTool, which is a tool designed by Ronald L. Rivest and Mayuri Sridhar for Bayesian upset probability measurements for Bayesian hybrid audits. This tool is open-source and available on Github.

Results/Lessons Learned

We note that we did not use the Bayesian results to drive the audit. That is, the audit's stopping rule was not based on a requirement on how low the upset probability was; rather, we measured and reported the Bayesian upset probability after the first round of sampling. All audits used the priors described in the *Math/Procedures* section.

For the pilot in Rochester Hills, Michigan, we used a ballot-polling audit, which required a sample of 76 ballots to audit a proposition. We found that 50 ballots in our sample were for "Yes" and 26 ballots were for "No". The selected precincts had a total of 36,666 votes. Running this through the Bayesian audit tool provided a 0.27% Bayesian upset probability for the prior that was used.

For the pilot in Lansing, Michigan, we used a hybrid audit to audit a judgeship. We drew 258 ballots, since the race had a relatively small margin (approximately 10%). However, we found that there were many discrepancies where the machine interpretation of a ballot did not match the auditor's manual interpretation. We note that Lansing did not have ballot imprinting and the only way to find the cast vote record (CVR) for a ballot was by position. That is, we assumed that the ballots stayed in order in the tabulators and that the paper ballot at the 5th position in the stack would correspond to the

5th CVR. However, maintaining the order of the paper ballots is quite tricky and we would highly recommend implementing ballot imprinting for those seeking to run a comparison audit. The Bayesian hybrid upset probability was measured to be 10.28% for the prior used. Due to the large number of discrepancies, we chose to measure the Bayesian ballot-polling upset probability as well, which came out to 9.95% for the prior used.

For the pilot in Kalamazoo, Michigan we used a hybrid audit to audit the governor's race. We note that Kalamazoo had ballot imprinting on their DS450, which made the comparison audit significantly easier. That is, we would choose a paper ballot based on the sampling plan and find the corresponding CVR, based on the imprint on the ballot. We found no discrepancies between the machine interpretations and the manual interpretations in Kalamazoo. We drew an initial sample of 40 ballots, 32 from the ballot-polling stratum and 8 from the ballot-comparison stratum. The Bayesian ballot-comparison audit provided an upset probability of 0.03% for the prior used.

While the RLA and the Bayesian audit were quite similar in structure and purpose, it appeared that the measured upset probability is about seven times smaller than the measured risk. Other experiments have found that the risk of a Bayesian audit can be up to 11 times larger than the upset probability. Larger ratios might occur in more complicated elections and for other social choice functions. Thus, for now, it may be appropriate in future audits to use an upset probability that is about ten times smaller than the desired risk limit for "comparable" results, at least for some priors. (Comparable is in quotes here, since the RLA and the Bayesian audit are actually estimating somewhat different quantities.)

K-Cut Overview

An important step in all RLAs and Bayesian audits is the selection of random ballots. Thus, the auditors must select and employ a sampling method. In Michigan, two different sampling methods were employed: 1) countdown method, and 2) k -cut method. When using the countdown method, auditors count down to a specific ballot in a batch identified by the Tool. For example, when the Tool selects Ballot #412 from Precinct 17, the auditors retrieve the ballot bag for Precinct 17, remove the ballots from the ballot bag and place the ballots in a stack on a table. The auditors then began to count the ballots, one at a time, starting from the top of the stack, to the four-hundred and twelfth ballot. This ballot is subsequently manually reviewed during the audit. When using k -cut the auditors take the stack of ballots from Precinct 17 and cut the stack (in a similar manner to a deck of cards) six times. The ballot on top of the stack after being cut 6 times is the ballot which is subsequently manually reviewed during the audit.

For the Kalamazoo and Lansing pilots, the ballots cast on Election Day were segregated from the ballots cast absentee. This segregation was important because, as noted above, different audit methods were used on these different ballot universes or *strata*. Different ballot selection methods were used to randomly select ballots from different universes in different pilots.

Here, we measure efficiency by the amount of time it takes the auditors to retrieve the random ballot. The most efficient sampling method will vary based on the total batch size and the ballot number selected by the Tool. For the Michigan pilot, we tried various combinations of the countdown method and the k -cut method. In Michigan, the k -cut method was generally more efficient when the batch sizes were in excess of 200 and the Tool selected a ballot number greater than 200.

Math/Procedures

The k -cut method is a procedure which can be used during an audit to select a random ballot. The purpose of k -cut is to improve the efficiency of sampling ballots. In particular, the sampling plan for an RLA defines which ballots to find and manually interpret during the audit. For example, the first ballot that we need to find might be ballot #526 from the election day ballots in Precinct #1. To find this exact ballot, we need to remove all the election-day ballots from Precinct #1, stack them together, and count 526 ballots down from the top of the stack. Recovering even a single ballot from precincts which have thousands of ballots can be tedious when using the countdown method to select random ballots for review in an audit.

The k -cut procedure is designed to simplify the process of randomly choosing a ballot from a stack. In particular, if the sampling plan requires us to find ballot #526 from election-day ballots in Precinct #1, we start the same way.

We remove all the election-day ballots from Precinct #1 and stack them together. However, instead of counting to find ballot #526, we make k cuts in the stack and choose the ballot that ends up on top. For any single cut, we use a random number generator to generate a random number between 1 and 99. Let's say we generate the number 32. Then, the audit team would try to remove approximately 32% of the ballots off the top of the stack, where this approximation can be eyeballed. Then, they place the remaining ballots on top of the ballots that have been removed to complete the cut.

We can then repeat this process k times, and choose the ballot on top as the randomly chosen ballot from this stack. In practice, we recommend choosing $k=6$.¹²

Results/Lessons Learned

We found that the k -cut (with $k=6$) procedure worked quite well in improving the sampling efficiency of the audit. Here, sampling efficiency is determined by the amount of time auditors need to select a random ballot for the audit. In particular, in Rochester Hills, we found that making 6 cuts and choosing the top ballot took approximately 60 seconds and counting a single ballot took about 1 second. This implies that k -cut would be more efficient, when the ballot position in the stack is at least 60, which was quite common. In Rochester Hills, we recommended that k -cut is used when the ballot position was at least 200.

— Special thanks to Mayuri Sridhar for her work on this section.

¹² For further details on the procedure, see Mayuri Sridhar and Ronald L. Rivest, "k-Cut: A Simple Approximately-Uniform Method for Sampling Ballots in Post-Election Audits," arXiv:1811.08811v3, Jan. 2, 2019. Available at <https://arxiv.org/abs/1811.08811>.

In Lansing, to save time, we used *k*-cut for all election-day ballots and the counting technique for all the absentee ballots. We timed 12 iterations of 6 cuts, on varying stack sizes and calculated the average time for 6 cuts to be 65 seconds. We noticed that when the stack contained more than 1,000 ballots, *k*-cut took closer to 80 seconds.

We repeated this experiment in Kalamazoo, which had several large stacks of ballots. For large stacks (which contained at least 1,000 ballots and up to 1,600), making 6 cuts took an average of 137 seconds, which is almost twice as long as the average time for smaller stacks. In Kalamazoo, counting ballots took an average of 1.7 seconds per ballot over five measurements. We note that when many ballots need to be counted, the pace slowed down. That is, when approximately 100 ballots need to be counted, the average time was about 1 second per ballot. When the team had to count at least 300 ballots,

the average time was closer to 1.8 seconds per ballot. We also timed the audit team's entire process from start to finish. That is, we measured the time it took for a team to open the bag containing all the ballots, obtain a ballot using *k*-cut, and close the bag up. This overall process took slightly over 4 minutes in total.

We note that the initial design of *k*-cut relied on the audit team to make "random" cuts, to the best of their ability, without the use of the random number generator. We believe that the use of the random number generator will improve how random each cut is, which might allow us to reduce our value of *k* from 6 to 4, which will further increase the efficiency of this procedure. We ran some tests in Lansing to test this conjecture and the preliminary results seem quite promising, although further experimentation is needed. We would like to thank the cities in Michigan for allowing us to pilot this procedure and better understand its usability and possible improvements.

Conclusion

Risk-limiting audits are a smart and effective tool that election officials should consider adding to their election security toolkit. These audits can complement existing election security measures and procedures by providing a check on election outcomes. They are designed to provide assurance that a reported winner did win an election or, in the alternative, to detect election irregularities such as cyberattacks or human error that may have altered an election outcome.

By conducting these pilots, Michigan election officials gained hands-on experience with the procedure and confirmation that it can work in Michigan's precinct-based election administration system with Michigan-certified voting systems. This experience provided invaluable lessons about RLAs and equipped election officials with

the experience and confidence necessary to take the next steps as they consider statewide RLAs: 1) expanding the pilot RLA pilot project to include county election officials, and 2) working with voting system vendors to improve RLA support functionality.

ACKNOWLEDGMENTS

The Michigan election officials, Scott Borling, Clerk, City of Kalamazoo, Tina Barton, Clerk, City of Rochester Hills, Chris Swope, Clerk, City of Lansing and Virginia (“Ginny”) Vander Roest, Training and Communications Manager, Michigan Bureau of Elections and the RLA Team, Drs. Jay Bagga and Bryan Byers, Voting System Technical Oversight Program (VSTOP), Ball State University, Wilfred Codrington, Andrea Cordova, Chris Deluzio and Liz Howard, Brennan Center for Justice, Jerome Lovato, U.S. Election Assistance Commission, Professor Ron Rivest and Mayuri Sridhar, Massachusetts Institute of Technology (supported by the Center for Science of Information (CSOI), an NSF Science and Technology (STC), under grant agreement CCF-0939370 and Verified Voting Foundation) and Dr. Philip Stark and Kellie Ottoboni, University of California, Berkeley and Verified Voting Foundation gratefully acknowledge the many contributors to the various organizations involved in this project. This report – and the pilots – were a true team effort. The authors gratefully acknowledge the colleagues who collaborated in preparing this report and assisting with the pilots, most importantly: Shelby Moss, Deputy Clerk, City of Kalamazoo, Leanne Scott, Deputy Clerk, City of Rochester Hills, Robin Stites, Elections Supervisor, City of Lansing and

Sally Williams, Director of Elections, Michigan Bureau of Elections. We sincerely appreciate the election officials from across the country who assisted with and supported this effort: Brenda Cabrera, General Registrar and Director of Elections, City of Fairfax, VA, Judd Choate, Director, Division of Elections, Colorado Secretary of State’s Office, Brennan Center’s Edgardo Cortes, former Commissioner of Elections, Virginia, Debra Johnson, Clerk and Recorder, Denver, CO, Neal Kelley, Registrar of Voters, Orange County, CA, Democracy Funds’ Jennifer Morrell (former Arapahoe County, CO election official), Richard Rydecki, Assistant Administrator, Wisconsin Elections Commission, Dwight Shellman, County Support Manager, Colorado Secretary of State’s Office and Steven Taylor, Warehouse and Logistics Manager, Rhode Island Board of Elections. We also thank the dozens of Michigan municipal and county election officials who staffed or observed the pilots. Brennan Center Deputy Director Lawrence Norden provided instrumental guidance, direction and edits. Monica Crane Childers offered ongoing feedback, hands-on expertise, assistance and edits. Matthew Bernhard, University of Michigan, provided on-the-ground technological support and expertise.