CCC
Computing Community Consortium
Catalyst

Research Opportunities in Evidence-Based Elections

A Computing Community Consortium-led white paper[1] Published January 2022

Authors: Josh Benaloh (Microsoft Research), Philip B. Stark (University of California, Berkeley), Vanessa Teague (Australian National University), Melanie Volkamer (Karlsruhe Institute of Technology), and Dan Wallach (Rice University)

With support from: Daniel P. Lopresti (Lehigh University), Brian LaMacchia (Microsoft Research), Khari Douglas (Computing Community Consortium) and Maddy Hunter (Computing Community Consortium).

1. The Need for Evidence-Based Elections 2

2. Risk-Limiting Audits 5
- 2.1 SHANGRLA Framework 7
- 2.2 Open Questions 9

3. End-to-End Verifiability 11
- 3.1 Open Questions 13
  - 3.1.1. Improving the Properties 13
  - 3.1.2 Expanding the Working Scenarios 14
  - 3.1.3 Human Factors in End-to-End Verifiability 15

4. The Role of Internet Technologies in Verifiable Elections 16
- 4.1 Limitations of Today’s Technologies 16
- 4.2 Opportunities for Internet Technologies to Help with Elections 19
  - 4.2.1 Open Questions and Research Challenges 21

5. Other Research Questions in Voting Systems 23

6. References and Related Work 24

[1] This material is based upon work supported by the National Science Foundation under Grant No. 1734706. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

1. The Need for Evidence-Based Elections

There is a crisis of confidence today in U.S. elections. Millions of Americans do not believe the announced 2020 presidential election results, and many other recent elections have had vocal doubters. Regardless of the basis of these concerns, our current electoral system fails to provide convincing evidence that the reported results reflect the will of the eligible voters who participated in the election. Electoral processes need to produce evidence that the election outcome is correct. Here, we discuss several emergent technologies that can enable *evidence-based elections*—elections that provide convincing public evidence that the reported outcomes are indeed correct (Stark & Wagner, 2012, Appel & Stark, 2020). The primary technologies which can be brought forward are (1) well-curated hand-marked paper ballots with *risk-limiting audits* (discussed in Section 2) and (2) cryptographic *end-to-end verifiability* (Section 3). The former creates public evidence that the reported outcome agrees with what an accurate tabulation of the votes would yield, and the latter allows voters to verify for themselves that their votes have been correctly included and that all publicly posted votes were tabulated correctly. These technologies have not yet been widely deployed and there are many opportunities for the research community to improve their form and function in order to enable evidence-based elections; for instance, (1) how to effectively deploy these technologies, (2) how to apply these technologies in particular to complex voting scenarios, (3) how to make these technologies usable and understandable for all, (4) and how to translate the trustworthiness created by these technologies into public trust in election results.

Unfortunately, there is not necessarily a connection between the *trustworthiness* of an election and the public *trust* it receives. But we can provide evidence that can be checked by candidates, media outlets, interest groups, and voters themselves rather than just asking voters to trust election officials, processes, and equipment. In many modern democracies, elections for representative assemblies are conducted with paper ballots placed into a public urn or ballot box. The voting process is intuitive and based on simple technologies, but some trust must be vested in the observers of the count; therefore, paper ballot custody is critically important, but few, if any voters can check whether the chain of custody was broken or watch the entire count, let alone visit every poll site across the country.

Electronic electoral processes can introduce additional components that may or may not be trustworthy. Thus, elections must ensure *software independence* (Rivest and Wack 2008), which means that an undetected error in the software cannot produce an undetectable change in the election outcome. Strong software independence requires *recovery* (resilience in the face of tampering) as well as detection. Without software independence there is no hope of producing evidence that the outcome is right. Software independence is a more subtle notion than it might seem at first glance, partly because of the difficulty of defining "detectable," but suitable definitions can be generalized to include *hardware independence* and independence from election personnel—whether inadvertent or malicious (Jamroga et al., 2022). Who is in a position to "detect" problems and whether the evidence of problems is public are also key, leading to the notion of *contestability* of an election system (Appel et al., 2020)?

There are several ways to expand the notion of evidence-based elections. *Accountability* ensures that when an error is detected it is possible to learn who or what caused it. *Dispute resolution* ensures that false alarms can be distinguished from genuine problems and that observed problems are taken seriously.

Electoral processes should also provide *tamper evidence*: the means for voters and observers to detect deliberate or accidental errors or fraud (at least if such problems altered the electoral outcome). However, *detectability* of fraud and tampering does not ensure *detection.* Actual detection requires action and may be probabilistic and/or depend upon assumptions e.g., about voter behavior.

Open-source software, while often desirable, is insufficient to ensure trustworthiness. Source disclosure and licensing do not ensure that software is free from defects, and it is difficult to provide voters and observers convincing evidence that the software actually running on a voting device matches the claimed source. The same holds for evaluating and certifying the software, which is often desirable but insufficient to ensure trustworthiness.

The various available election technologies—paper ballots, physical security, cryptographic verification, statistical audits—are all able to produce evidence about votes and election outcomes, but under different assumptions about who or what needs to be trusted and for what purpose.

A well-designed system does not require the public to choose one trust model or the other: it can implement both. See, for instance STAR-Vote and ElectionGuard in Bell et al. (2013), Burt (2019), and Chaum et. al (2008). The effective combination of different forms of evidence is an important direction for research.

Additionally, vote anonymity, or the secret ballot, is a precondition for free public elections. It defends against coercion and vote-buying. Both paper-based evidence trails (e.g., for recounts and risk-limiting audits) and cryptographic evidence trails (i.e., end-to-end verifiability) need to be implemented carefully to ensure ballot secrecy. In the case of cryptography, extra care must also be taken to ensure that voters cannot produce evidence of how they voted, even when the system does not expose it. (This property is called *receipt freeness*. Read more in Tuinstra & Benaloh [1994]). Evidence and privacy are in tension in elections, but it is not impossible to provide both (Bernhard et al., 2017).

Elections also vary greatly from state to state in the U.S. and from country to country. A solution that works well for a few simple first-past-the-post races may not work at all for instant runoff voting, for a jurisdiction with dozens of contests on one ballot, or for a country with a parliamentary electoral system. In Canadian parliamentary elections, votes are cast on paper and counted by hand—this simple transparent process would not be feasible in much of the U.S. because a single US election typically involves many contests and referenda. Criteria and priorities vary too: in the US, voting is optional; in Australia, it is compulsory; and in Switzerland and Germany, not only is it important to maintain privacy of votes, but the identities of those who voted must also be kept private, making it much more complicated to produce evidence that eligibility determinations were accurate.

Research therefore needs to expand knowledge along several axes:

- Different people accept different types of evidence or are willing to trust different people. Therefore, it is useful to diversify the types of evidence and distribute the trust base. In the U.S., it is currently very difficult (or impossible) to refute false claims that vote counts are inaccurate.
- Different countries (and states) run different types of elections using different social choice functions. Therefore, it is useful to expand the types of elections and social choice functions for which good evidence can be produced.
- In different contexts, different criteria are prioritized, such as verifiability, usability, accountability, convenience, and ballot secrecy. Therefore, it is useful to expand the sets of options that are possible, though no system is likely to meet all desiderata.

This white paper offers an overview of technologies that can address the lack of sufficient evidence in current elections and describes numerous research questions related to these technologies. It is not intended to provide detailed, self-contained descriptions of these technologies; instead, it gives just enough of an explanation to provide the context required to describe the many opportunities for new research to make evidence-based elections more available and viable.

2. Risk-Limiting Audits

A risk-limiting audit (RLA) is any procedure with a known minimum chance of correcting the reported electoral outcome if the reported electoral outcome is wrong—that is, if the reported winner(s) did not really win—and zero chance of altering a correctly reported outcome. No procedure can offer such guarantees unless there is a trustworthy record of the vote the procedure can rely on. (To establish that the record is trustworthy generally requires the election to have been conducted well in terms of ballot accounting, eligibility determinations, etc., as well as passing an additional step of scrutiny called a *compliance audit*. See Appel & Stark [2020]).

RLAs frame audits as statistical hypothesis tests. The "null" hypothesis is that the reported outcome is incorrect, i.e., that one or more reported winners did not really win. An RLA terminates either by finding strong statistical evidence that the null hypothesis is false—that every reported winner did in fact win—or by conducting a full manual tabulation of the votes, which reveals the true winners if the paper trail is trustworthy. The chance a RLA stops without a full manual tabulation when one or more outcomes is wrong is at most the *risk limit.* Typical risk limits are 1%, 5%, and 10%.

RLAs can use samples drawn either at the ballot level (sampling individual ballots) or the batch level (sampling physically identifiable groups of ballots, such as the ballots cast in a particular precinct or tabulated using a particular machine). Samples can be drawn with or without replacement. They can be drawn without stratification, or the population of ballots can be divided into strata, with independent samples drawn from different strata. Samples can be drawn with equal weight or using a method such as probability proportional to a measure of "size." Using smaller batches in RLAs typically decreases the workload when contest outcomes are correct. Of course, when contest outcomes are incorrect, RLAs, regardless of their sampling strategy, are supposed to lead to a full hand tabulation with high probability.

There are two general strategies for RLAs: *polling* and *comparison*. *Polling* RLAs involve manually reading the votes from randomly selected ballots or batches of ballots. *Comparison* audits additionally involve comparing that human interpretation to how the voting system tabulated the same ballots. There are ways of combining the two approaches in the same audit (Ottoboni et al., 2018 [SUITE]).

It has become standard to use *sequential* hypothesis tests[2] in RLAs, that is, methods that control the significance level but allow the sample to be expanded at will as long as the null hypothesis has not been rejected.

Using sequentially valid tests allows auditors to start with a sample size that is expected to suffice if the reported tabulation is accurate, then expand the sample size incrementally (potentially to a full hand tabulation) if the initial sample does not provide sufficiently strong evidence that the reported winner(s) really won.

Risk-limiting audits and recounts provide trustworthy evidence about electoral outcomes only if the paper trail is a trustworthy record of voters' expressed votes. That requires the paper records to have been properly produced and protected—and evidence that they were. Aspects of recounts and audits that must be observed for them to merit trust vary, but at a minimum, there must be evidence that every eligible voter had the opportunity to vote and that the paper trail includes every validly cast vote and no others. In turn, that requires evidence about eligibility determinations, chain of custody, ballot accounting, pollbook reconciliation, and so on. No audit procedure can limit the risk that an incorrect result will become final unless there is a trustworthy record of voters' expressed votes (see Appel et al. [2020] for a discussion of *expressed* votes). Otherwise, even a full manual tabulation might not show the true outcome.

To draw a random sample of ballots, RLAs generally require a trustworthy *ballot manifest*, a detailed description of how the ballots are stored. For instance, a ballot manifest might say,

[2] Sequential tests were first developed in the 1940s by Wald (1945). More recent sequential tests generally follow from Ville's Inequality (Ville, 1939), which states that the chance a nonnegative martingale *ever* exceeds a given multiple of its mean is at most the reciprocal of that multiple.

"there are N ballots in all, stored in M containers, labeled 1 to M. Container 1 has N_1 ballots, container 2 has N_2 ballots, …" The ballot manifest should be derived by physical accounting, without relying on the voting system. There are RLA methods that do not require ballot manifests for sampling (e.g., methods that use Bernoulli sampling [Ottoboni et al., 2018]), but absent control on the number of ballots and how ballots are stored, little can be said about whether a reported result is correct. Audit methods have to accommodate real-world issues, such as missing ballots and mismatches between the number of ballots according to the voting system and according to the ballot manifest. See, e.g., SHANGRLA by Stark (2020).

The most efficient approach to RLAs is a comparison audit done by selecting individual ballots at random and comparing a human interpretation of the votes to the machine interpretation of each corresponding ballot, aka *cast vote records*. However, many voting systems do not create cast vote records or cannot export cast vote records in a way that makes it possible to determine which cast vote record corresponds to which physical ballot. Additional large efficiency gains are possible when ballots consist of more than one piece of paper and when not every contest is on every ballot card (Glazer et al., 2021)

The evidence that must be published for an RLA to justify public trust in election outcomes depends on the RLA method. For instance, comparison audits require evidence about how the voting system interpreted individual ballots or groups of ballots, while polling audits do not. Revealing those interpretations may compromise the anonymity of votes. This has been addressed by schemes that provide a cryptographic commitment to the interpretations without publishing plaintext cast vote records (Benaloh et al., 2011, Benaloh et al., 2019), enabling the public to verify that the audit used the same cast vote records the tabulation relied on.

RLA methods have been developed for a broad range of sampling schemes, including sampling individual ballots with or without replacement; sampling clusters of ballots with or without replacement, with equal probabilities or with probability proportional to a measure of size; stratified sampling; and Bernoulli sampling (Stark, 2008, 2009, 2020 [SHANGRLA]); Higgins et al. 2011; Lindeman & Stark, 2012; Ottoboni et al., 2019 [Bernoulli Ballot Polling]). Methods have also been developed that allow different strategies to be used in different strata, for instance, using polling for some subsets of the ballots and ballot-level comparison or batch-level comparison for other subsets (Ottoboni et al., 2018 [SUITE]; Stark, 2020 [SHANGRLA]).

2.1 SHANGRLA Framework

The SHANGRLA framework (Stark, 2020) provides a unified treatment of all these sampling schemes, for a broad range of social choice functions, including all scoring rules (e.g., plurality, multi-winner plurality, Borda, and STAR-Voting); super-majority; instant-runoff voting (IRV, a form of ranked-choice voting); and proportional representation schemes such as D'Hondt and Hamilton elections. SHANGRLA, which stands for "sets of half-average nulls generate risk-limiting audits," reduces auditing election outcomes to multiple instances of a simple statistical question: is the mean of a list of nonnegative, bounded numbers greater than ½? That reduction then allows advances in statistical methodology for testing hypotheses about the means of finite populations to be applied immediately to risk-limiting audits.

Consider a single-winner plurality contest with three candidates, Alice, Bob, and Carol. Suppose Alice is the reported winner. Alice really won if she received more valid votes than Bob and than Carol. We will express this as the *assertion* that the means of two lists of nonnegative, bounded numbers are greater than ½, as follows.

Each list has as many elements as there are ballots: the lists are constructed by assigning a number to each ballot based on what the ballot shows. The first list assigns a ballot the value 1 if it shows a valid vote for Alice, the value 0 if it shows a valid vote for Bob, and the value ½ if it shows a valid vote for Carol or does not contain a valid vote in the contest. The second list assigns a ballot the value 1 if it shows a valid vote for Alice, the value 0 if it shows a valid vote for Carol, and the value ½ if it shows a valid vote for Bob or does not contain a valid vote in the contest. Both of these lists are nonnegative and bounded above by 1.

If the mean of the first list is greater than ½, Alice really got more votes than Bob. If the mean of the second list is greater than ½, Alice really got more votes than Carol. If both lists have means greater than ½, Alice really won. In this example, the *canonical assertions* that the two means are greater than ½ are necessary and sufficient for Alice to be the true winner.

A RLA can check the correctness of the outcome—the truth of the two assertions—by testing the two *complementary null hypotheses* that the means are less than or equal to ½. If both complementary null hypotheses are rejected, that is evidence that both assertions are true, and thus that Alice really won.

The sketch above is not the only way the correctness of the reported outcome could be transformed into assertions that a set of lists have means greater than ½. For instance, suppose it was reported that Bob got more votes than Carol. Consider a third list that assigns a ballot the value 1 if it shows a valid vote for Bob, the value 0 if it shows a valid vote for Carol, and the value ½ if it shows a valid vote for Alice or does not contain a valid vote in the contest. If the mean of this list is greater than ½, Bob really got more votes than Carol. Thus, Alice really won if the means of the first and third lists are both greater than ½. In this encoding, however, the two assertions are sufficient to guarantee that Alice really won, but not necessary, since Alice really won even if Carol got more votes than Bob, provided Alice got more votes than Carol.

Currently, necessary and sufficient canonical assertions have been developed for all scoring rules, D'Hondt, and Hamilton elections. Assertions that are sufficient but not necessary have also been developed for single-winner IRV/RCV (instant-runoff voting, ranked-choice voting). Assertions guaranteeing the correctness of the outcome of single-transferable vote (STV) have not yet been constructed.

2.2 Open Questions

Open questions about risk-limiting audits include theoretical issues, efficiency and logistical issues, and issues related to public trust.

Theoretical issues: SHANGRLA (Stark, 2020) provides a unifying framework for risk-limiting audits that encompasses all social choice functions for which there is currently a known RLA method, including plurality, multi-winner plurality, supermajority, STAR-Voting, Borda count, RCV, and every social choice function that is a *scoring rule*. SHANGRLA does not apply to social choice functions that depend on the order in which ballots were cast or tabulated.[3] But it is not known whether SHANGRLA applies to *every* social choice function that depends only on the set of votes and not their order. For what social choice functions is there a sufficient set of assertions in the SHANGRLA framework? For what social choice functions is there a necessary and sufficient set of assertions in the

[3] Some preferential voting systems used in political elections, such as the version of single transferable vote used in Ireland, depend on the order in which ballots are tabulated.

SHANGRLA framework? Are all sets of necessary and sufficient conditions equally expensive to audit, given the set of cast votes? Are any social choice functions intrinsically harder to audit than others, e.g., because computing the margin is NP-hard, or because the margins are typically small?

Stratification: theory and practice: There are legal and practical reasons for using stratified sampling in RLAs. For instance, some states leave it to individual jurisdictions to draw their own audit samples. Stratification can also increase efficiency when tabulation equipment is heterogeneous. Current methods for stratified audits rely on combining p-values across strata, for instance using Fisher's combining function (Ottoboni et al., 2019; Stark, 2020). Are there sharper audit methods for stratified samples? How should the sample be allocated across strata in stratified audits? When strata correspond to different jurisdictions, it is not clear that minimizing the total sample size is an appropriate objective. For instance, if some jurisdictions have voting systems that permit more efficient auditing methods (such as ballot-level comparison audits versus ballot-polling audits), should those jurisdictions get the benefit of their investments, or should they help shoulder the burden of jurisdictions with less-efficiently auditable systems? If the audit needs to expand, how should the sample be augmented from stratum to stratum? If discrepancies are discovered in some jurisdictions but not others in the audit of a cross-jurisdictional contest, should all jurisdictions with votes in the contest increase their sample sizes proportionately, or should the burden fall disproportionately on the jurisdictions with discrepancies? These questions have technical, economic, legal, and ethical aspects.

Sequential tests, batch-sequential tests, and audit escalation schedules: The sequentially valid tests at the heart of many RLA methods are based on Ville's inequality for nonnegative martingales. Many different nonnegative martingales can be developed for a given assertion; see, e.g., Waudby-Smith et al. (2021). Are there optimal (most powerful) martingales for testing SHANGRLA assertions, for stratified and unstratified sampling, for ballot-polling, comparison audits, and for "hybrid" approaches? The earliest RLA methods relied on a "schedule" of increasing sample sizes (Stark, 2008, 2010), but most extant RLA methods have focused on methods that are sequentially valid, no matter how the sample is expanded. If the sample will be expanded a prespecified number of times following prescribed rules ("batch-sequential audits"), how much could sample sizes be reduced? Zagorski et al. (2021) provide a partial answer for unstratified ballot-polling audits of two-candidate plurality contests with no invalid votes and no ballots that do not contain the contest, using sampling with replacement. In that case, the distribution of the data is in a known parametric family (each draw is an independent Bernoulli trial). What gains are possible if the contest has additional candidates, invalid votes, or the contest is not on every ballot? What if the sample is drawn without replacement or if the sample is stratified? The answers involve efficient inference in the presence of nuisance parameters, an active area of statistical research. What gains are possible for comparison audits, which are generally more efficient than ballot-polling audits? The answer involves longstanding questions in nonparametric statistics regarding testing hypotheses about bounded finite populations.

Public trust, transparency, legislation, and outreach: While some “ingredients” needed to justify public trust in RLAs are known (e.g., evidence about chain of custody and eligibility determinations, code disclosure, public ceremonies to generate the seed for random sampling, commitments to the voting system’s interpretation of ballots for comparison audits), can RLAs be modeled as an end-to-end verifiable process? What else does the public need to know for RLAs to justify public trust? Does the length of time required to conduct an audit affect public trust? How can trustworthy risk-limiting audits and their prerequisites (such as *compliance audits*) best be expressed in legislative and regulatory language? This will require interdisciplinary research. How can risk-limiting audits be explained to a lay audience in a way that will lead to understanding and trust? This will also require interdisciplinary research.

3. End-to-End Verifiability

End-to-end (E2E) verifiable voting systems are made up of a set of technologies which together allow voters to check for themselves that their votes have been accurately counted. Although the name is relatively new, the roots of E2E-verifiability go back to the early 1980s (Chaum, 1981). An election is end-to-end verifiable if two properties are achieved:

1. Voters are able to confirm that their intended selections have been accurately recorded (also called voter-verification), and
2. Anyone can confirm that all recorded ballots have been accurately tallied.

Some authors also add a third property, *eligibility verifiability,* meaning the opportunity to verify that each vote came from a distinct eligible voter. In some situations, it suffices to simply associate a voter with each encrypted ballot or publish a list of all voters who cast ballots. In more privacy-sensitive settings, eligibility verifiability requires more complex protocols.

When combined, these properties give voters the opportunity to confirm the correct counting of their votes. E2E-verifiability provides evidence of an accurate election outcome under the assumption that enough voters verify, but it does not guarantee that this assumption holds. It also says nothing about the secrecy of the ballots, and it is very easy to achieve E2E-verifiability in an open-ballot election. The greater challenge is to achieve these properties in a secret-ballot election, with easy-enough verification to be widely and successfully used. This is almost always done by encrypting and then publishing encrypted ballots. (Some systems work by publishing plaintext ballots with anonymized identifiers.) There are many effective—but challenging to implement—methods that enable voters to verify that their encrypted ballots reflect their intended selections—without having to perform computations or understand anything about encryption. The second property—showing that a set of encrypted ballots corresponds to an announced tally—is where more sophisticated cryptographic techniques are typically used. The two most common techniques are MixNets and homomorphic tallying—the former dissociates voters from the encrypted ballots before decrypting the ballots, while the latter produces verifiable tallies of encrypted ballots without ever decrypting them.

In practice, voters may engage in some process that allows them to verify the accurate recording of their intended selections while casting their ballots. Ideally, this process is optional so as not to unnecessarily encumber those voters who want to cast their ballots with a minimum of effort. Once this process is complete, voters generally have access to a copy of their encrypted ballots—which enables them to confirm that these encrypted ballots have not been altered—but there is typically no mechanism for voters to review the contents of their ballots once they have been cast. This is to prevent voters from showing their selections to others and thereby prevents vote-selling and coercion.

Skilled programmers can write tools to perform the computations necessary to verify the correct tally of a set of encrypted ballots. Individual voters and observers, such as candidates, news media, and interest groups, can use one or more of these verification tools written themselves or provided by independent programmers. Thus, there is a chain of trust that is enabled by E2E-verifiability:

- Cryptographers can review a detailed specification of a design and confirm that, if it is followed, E2E-verifiability is achieved;

- Independent programmers can implement verification tools according to the specification; and
- ordinary voters and observers can use these tools to ensure that their votes have been correctly tallied.

The principal benefit of E2E-verifiable voting system is that voters can now choose to proxy their trust instead of being compelled to trust their local election officials and the equipment and vendors they choose to utilize. Through E2E-verifiability voters are able to believe their preferred sources, download and run verification tools from one or more preferred sources, or build and run their own verification tools.

For voter-verification, human factors become crucially important. It is not enough for a mechanism to exist in theory. The verification step must be usable in practice by ordinary voters, i.e., they should be able to verify and to notice if their vote has been manipulated. This is in particular important for cast-as-intended verification. This step can currently only be performed by voters as vote secrecy would be broken if conducted by others.

Human factors in security (and privacy) are a relatively new research field. Past security research was focused on the technological aspects and proposed a variety of security mechanisms that provide useful security properties. However, in many cases these properties only hold in theory because they are not aligned with users' mental models and capabilities. In order to practically build secure systems, we must consider the human factors, which include, but are not limited to, (1) supporting users in better protecting their devices and data against cyberthreats with more usable security mechanisms, (2) raising awareness for security risks and how they can be mitigated, and (3) supporting users in making more informed decisions e.g., which tools to use. To reach this goal, a human-centered security by design approach (Sasse and Rashid, 2019) must be applied, i.e., future users should be involved in the entire design process—including the requirement phase—of the to-be-developed mechanisms (or even cryptographic protocol). E2E-verifiable electronic voting schemes are not different from other security mechanisms. In order to increase the level of security more human centered security research is required.

We have outlined the potential benefits associated with E2E-verifiability, but there are many interesting research problems to be overcome in order to deploy these systems more widely. These challenges are discussed in more detail in the following subsections.

3.1 Open Questions

3.1.1. Improving the Properties

Dispute resolution: How can a system ensure that malicious behaviors are not only discoverable but also provide tangible evidence that can be used to convince third parties of such behaviors? Most current technologies allow voters to confirm the correct recording of their selections, but they are less effective at allowing voters to prove to third parties any transgressions that they discover. (Though this is not universally true: systems with pre-printed ballots, such as Scantegrity II [Chaum et al., 2008] and Prêt à Voter [Ryan, 2009], may allow a voter to prove that a ballot is malformed, or that the system will not open it.) A proof of malfeasance is critical for two reasons: first, it means that public detection is possible even if only a few voters actually verify (whereas a small number of claimed failures might be ignored). Second, it helps to defend the system against false accusations of malfeasance, on the assumption that it is not possible to forge a proof. Without evidence to prove malfeasance, the integrity of an election depends statistically upon a sampling of voters checking the integrity of their ballots.

Iterative verification in the case of close margins: Sequentially valid statistical audits of paper ballots can escalate—increase the sample—after the outcomes are announced, allowing election officials to dig deeper if an election is close to make sure the reported outcome is correct. Existing cryptographic methods lack this feature: they generally only allow voters to verify that their votes are cast as they intend before they cast their votes, and hence before the results are announced. (They can, of course, check that their votes are present on a public list of votes.) Hence cast-as-intended verification cannot be escalated if the election margin turns out to be very close. Is there any way to achieve RLA-style escalation of verification for cryptographic cast-as-intended verification?

Long-term vote secrecy: Cryptography is a dynamic science—what is secure today may not be tomorrow. Votes whose confidentiality are protected by today's encryption methods could be revealed by new discoveries and advances. This is a particular concern with recent advances in quantum computation. A technique called everlasting privacy has the potential to permanently protect the confidentiality of ballots at the cost of adding cryptographic assumptions to the integrity of an election. This can be an appealing trade-off because the assumptions need hold only until an election is complete, and confidentiality will then be maintained in perpetuity. Further research on this technology may allow it to become practical for use within election systems.

Verifiable privacy: Traditional paper-based voting systems generally allow voters an intuitive way to verify that their votes are private, such as seeing them deposited in a physical urn in which all ballots are either hidden or indistinguishable, and are later publicly shuffled. No current cryptographic system has this property—cryptographic systems generally rely on (threshold) trust assumptions for the secret ballot. This is a much weaker property for two reasons: firstly, it relies on trusting some authorities to behave properly); secondly, individual good behavior is impossible to verify - there may be a bug in each trustee's key generation algorithm that makes it possible for vote decryption to occur without the participation of a threshold of trusted parties. So, the research question is, is there a design in which voters could see for themselves that their votes were private?

Combining different forms of evidence: Recent good designs combine RLAs with cryptographic-style evidence. The combination is probably the right way forward, but meaningfully combining different kinds of reasoning is not easy. The aim is that the verification process should be valid if *either* (or any) of the sets of assumptions are valid.

Public trust: How does trustworthiness translate to public trust? We can provide a system with great mathematical properties for verifiability. But if the public does not understand the system, it may not believe the claims; and a highly trustworthy system may not achieve its full potential. It would be valuable to better understand the relationship between mathematical verifiability and public trust, how to increase it, and to better understand the relation between understanding and public trust, as well as between communicating trust assumptions and public trust.

Verification by other voters: Voters are able to verify the correct recording of their own votes and the correct counting of all recorded votes, but is this sufficient? If very few voters bother to verify the correct recording of their votes, how much value is there in the ability to verify? With assumptions about randomly distributed verification, we can quantify the probability of detection of incorrect recording of votes, but can we do better?

3.1.2 Expanding the Working Scenarios

Vote-by-mail: Most E2E-verifiable techniques require an interactive process for voters to confirm the correct recording of their selections. The most promising systems also have a plain paper-ballot that can be audited with an RLA to produce evidence under a complementary set of trust assumptions. Relatively little research has gone into adapting

E2E-verifiability into a remote voting scenario with a paper backup—this could be regarded as adding verifiability to postal voting, or using the Internet to deliver the candidate information to the voter, with some cryptographic verification added to paper ballot return. While techniques exist that enable E2E-verifiable voting by mail, challenges remain in making them simple and usable enough to not encumber the voting process. Options and open questions are examined in Section 5.2.

Complex election methods: Various ranked-choice voting systems are becoming popular in many jurisdictions. While there are benefits to these more expressive voting methods, they present challenges for some of the E2E-verifiable technologies. There is an opportunity here for new research to better accommodate these more expressive voting methods.

3.1.3 Human Factors in End-to-End Verifiability

Usable security & accessibility of tools to verify: The initial development of E2E-verifiable techniques focused on the cryptography and attempted to minimize the underlying assumptions. As a consequence, the proposed tools to verify the recording of votes as intended were not very usable and did not reach an adequate level of accessibility for those with special needs. Subsequent focus was given to improving the interfaces but not changing the underlying techniques. This approach has its limits (see e.g., Kulyk et al [2019]). We recommend that future research takes a more human-centered approach in order to create more usable and accessible E2E-verifiable techniques (e.g., verification steps closer to what is known from other security contexts such as e-banking) that allow every voter (including those with special needs) to detect a manipulation of their vote. These techniques would need to be evaluated in corresponding user studies. Equities are particularly important for public elections—all voters should not only have the chance to cast their votes but also to verify their votes. Genuine E2E-verifiability for voters with special needs requires more research.

Risk awareness: E2E-verifiability will only be able to address public distrust of election outcomes if people understand the risks of the election processes in place, how E2E-verifiability reduces these risks, and which risks remain. Therefore, it is important to better understand people's mental models on security risks (for an overview see Volkamer & Renaud [2013]) in general, but, particularly in the context of voting and transparency and privacy in elections. Research is needed to understand how to best educate people about how to use these tools as well as to motivate them to use these tools. It is important to study how to raise this risk awareness without making people afraid of their vote privacy (at least not more than the actual risks are). Therefore, it might be necessary to study more broadly how people evaluate risks in the (e-)world, as well as to give them in general a better understanding of the concept of assumptions.

Decision support systems for E2E-verifiable schemes: There are already many different E2E-verifiability designs which have different properties with respect to the underlying security assumptions, performance, usability, accessibility, and costs. Some may even have pre-conditions that do not hold for all election settings. As the complexity of these cryptographic E2E-verifiability techniques may be too great for many election management boards to effectively evaluate, a decision support system is needed. As a pre-condition, the protocol descriptions and analyses need to be changed to become more easily comparable.

4. The Role of Internet Technologies in Verifiable Elections

On its surface, Internet Voting seems like a logical extension on the techniques of end-to-end cryptographic voting systems. After all, if an in-precinct voting system can produce an encrypted ballot and a voter can take home a “receipt” of some sort, then it would seem to be a simple extension to do the very same work in a web browser or smartphone application, transmitting the ballot to an online server, or perhaps publishing the ballot on some sort of distributed “blockchain” storage system.

This section describes a variety of open research challenges for Internet voting systems (challenges in addition to those mentioned before), which make these systems completely infeasible using present-day technologies and the current Internet architecture, although we point to a variety of adjacent opportunities where these kinds of technologies can improve voting systems and voter experiences.

4.1 Limitations of Today’s Technologies

In 2018, The National Academies of Sciences, Engineering, and Medicine produced a consensus report titled *Securing the Vote: Protecting American Democracy*, authored by our nation’s top experts in computer security, cryptography, election administration, and related fields. They offer many detailed recommendations, including (1) the need for physical ballots (whether machine or hand-marked), (2) the segregation of computer systems, (3) the need for new security standards for electronic poll books and related voter registration systems, (4) the use of pre- and post-election audits concerning all aspects of elections, (5) and the need to pilot end-to-end-verifiable election systems with paper ballots. The report also includes several recommendations against the use of Internet Voting:

> "At the present time, the Internet (or any network connected to the Internet) should not be used for the return of marked ballots. [*Footnotes 13 and 14 below.*] Further, Internet voting should not be used in the future until and unless very robust guarantees of security and verifiability are developed and in place, as no known technology guarantees the secrecy, security, and verifiability of a marked ballot transmitted over the Internet. [*Footnote 15 below.*]
>
> Footnote 13: Inclusive of transmission via email or fax or via phone lines.
>
> Footnote 14: The Internet is an acceptable medium for the transmission of unmarked ballots to voters so long as voter privacy is maintained and the integrity of the received ballot is protected.
>
> Footnote 15: If secure Internet voting becomes feasible and is adopted, alternative ballot casting options should be made available to those individuals who do not have sufficient access to the Internet" (p. 9).

These recommendations are built on a detailed understanding of how today's computers are vulnerable to attack: e.g., both personal computers and smartphones are often not updated with the latest security patches, reducing the level of sophistication necessary for an attacker. Particularly in "battleground states" with close margins of victory, an attacker might only need to compromise a modest number of voters in order to have a dispositive impact on an election.

Such automated attacks, compromising millions of personal devices, could lie dormant on the computer, waiting for a vote to be cast, and only then intervening to tamper with the resulting vote. No voter would have the ability to identify whether they were or were not subject to such an attack, nor would they be able to simultaneously verify that their vote was correctly recorded and tabulated without also being able to compromise their privacy and prove to a third-party who they voted for (thus creating opportunities for bribery and coercion).

In the current environment, end-to-end verifiability can mitigate—but not completely solve—this problem. A voter using an end-to-end verifiable system can use an independent device to confirm the correct recording of a vote. If the independent device reveals a problem, the voter could have the opportunity to correct the problem—either by removing the detected malware or by using another device. However, there is an important difference between this scenario and the one in which an end-to-end verifiable system is used for in-person voting.

The expectation is that only a fraction of voters will take the steps necessary to confirm the correct recording of their ballots. This is sufficient for in-person voting where detection of vote-tampering implicates a public voting device and calls into question the entire election; but detection of tampering on a personal device would be unlikely to do anything other than indicate malware on the voter's device. While a voter discovering an irregularity on a personal device could correct the vote, other voters who did not bother to confirm their votes would have had their votes successfully altered. This subtle but important difference in the likely response to detection of tampering between the in-person scenario and the Internet scenario means that the assurance of integrity of election tallies is substantially higher when end-to-end verifiability is used for in-person voting.

So why is Internet voting—sometimes also called "mobile voting," when it is done from a smartphone rather than a desktop computer—considered so attractive? In part, Internet voting addresses concerns that the postal system may be slow and unreliable, particularly for voters who are at a great distance from home, such as military and overseas voters. Internet voting also potentially allows for the use of bespoke accessibility solutions, customized to each voter's needs on personal computers, giving those voters the ability to cast their votes privately, without human assistance, and without needing to travel to a polling place.

Nonetheless, it is indisputable that our nation-state adversaries have a recent history of using the Internet in a variety of ways with the goal of manipulating our elections. This includes misinformation/disinformation on social networks as well as what seems to have been an exploration of their ability to directly manipulate our election systems. In the face of a skilled adversary like this, we cannot simply approve of Internet voting for its convenience, in the face of its relative ease of manipulation by a skilled actor.

Lastly, we note that some small companies have designed voting systems that include "blockchain" technologies, commonly used as distributed ledgers for monetary transactions, to instead serve as ledgers for encrypted ballots, protecting them from tampering. However, blockchains do not solve any of the challenges of Internet voting (e.g., client malware), and they introduce many new challenges (e.g., the ability of a majority of miners/stakeholders to control the contents of a blockchain). Blockchains also do not inherently address issues like voter privacy and accurate tallying. While it is possible to add features to blockchains to provide suitable privacy and verifiability, once these features are in place, the blockchains themselves become superfluous.

4.2 Opportunities for Internet Technologies to Help with Elections

Even though fully Internet-based voting systems are not feasible with today's computers and today's Internet, there are several opportunities to use the Internet and today's computers to improve elections. Some of these opportunities will require additional research effort, while others are feasible with today's products.

Vote centers and remote kiosks: In some states voters have as long as two weeks prior to the date of the election to visit an "early voting" location. Unlike traditional in-precinct voting, such voters can often go to any early voting location in their respective counties. For large counties this could include hundreds of locations. Some states also do this for voting on the day of the election itself, allowing for smaller individual precincts to be merged into larger "election centers" or "vote centers." Because voters have the potential to visit multiple vote centers, these elections must have mechanisms to defeat individual voters wishing to cast multiple ballots. The common solution is an online pollbook infrastructure to ensure that each voter is checked off in a central database.

Online pollbook infrastructure could well be subject to attacks from nation-state adversaries. An adversary might consider it a win if they can simply knock the pollbook infrastructure offline, especially if they can target such attacks to achieve a partisan bias in the election outcome. How can we build robust online pollbook infrastructure? This is an outstanding research challenge where we could potentially leverage concepts from the distributed systems community. For example, each local pollbook might retain a full copy of the voter registration database, allowing for degraded offline operation even if the central database is temporarily unavailable, with eventual synchronization that would, at a minimum, detect voters casting multiple ballots even if it cannot prevent them.

Another intriguing possibility would be to extend the “vote center” model beyond the limits of a specific election jurisdiction. Could a voter from Texas cast a vote while physically at a California polling location? This creates a variety of logistical challenges, including ironing out contradictory election regulations. For example, California requires that ballots **not** have unique numbers on them, while Texas requires that they do. Assuming these regulations could be harmonized, there might be an opportunity for short-term electronic transmission and efficient tabulation, followed up by bulk transfer of the paper ballots through the postal mail for auditing and related processes. This could even extend all the way to embassies, consulates, and overseas military bases where a “voting kiosk” could be established where voters could present themselves in-person to cast their votes, but would not need to worry about whether their votes make it back before any sort of deadline. By voting in-person at a remote vote center or kiosk, voters could more easily avail themselves of the benefits of an E2E-verifiable system.

Postal voting: Today, most vote-by-mail ballots in the U.S. require two trips through the postal system. A blank ballot is delivered to the voter, and a completed ballot is returned to the election administrator. The envelopes are typically coded to allow the Postal Service to do detailed tracking, which can then be used to notify the voters when their ballots were received. For some voters, they can access a blank ballot as a simple PDF file from their election administrators, allowing them to avoid the outbound delivery of the blank ballot, while still using the postal mail for the marked ballot return.

A relatively new technology, *assistive vote-by-mail* (AVBM), allows for the ballot to be marked on a screen via a web browser, then printed and returned. By running inside a browser, voters with assistive devices will be able to fill out their ballot as they would interact with any other form on the web. This creates an interesting research challenge to add E2E-verifiable technologies to AVBM, giving the AVBM voter access to many of the same verification features as an in-precinct E2E-verifiable voter might use.

Similar questions arise for any vote-by-mail system. There is surprisingly little research around this high-impact area, though there is a vast design space that trades off accessibility, convenience, access, privacy, and simplicity of verification.

Some example scenarios are:

- Voters download blank ballots, print them, mark them by hand, and return them by mail;
- Voters download blank ballots, fill them in electronically, print them, and return them by mail;
- Either of the above scenarios, with some enhancements to produce cryptographic evidence that the votes have been properly included and counted;
- Mailouts of preprinted ballots from an E2E-verifiable voting system, with some method of allowing voters to challenge some and cast one;
- Combinations of E2E-verifiable Internet voting systems with a paper evidence trail to be mailed in for an audit.

There are probably other scenarios worth considering.

4.2.1 Open Questions and Research Challenges

Ballot marking devices and/or AVBM printout verification: Any system that produces a human-readable ballot, where that paper record is the primary record of the voter's intent, faces the challenge that the human might not notice if what they did on the screen is different from what is on the paper. Even though they *can* verify it, they might not bother. This creates a challenge, from a usability perspective, to get the users to perform this additional step, which would then mitigate against the risk of malware trying to change the voter's intent.

Client-side platform security: In the AVBM case, or more generally, in any case where there is a computer between the voter and the record of their intent, we have to be concerned that the computer might have malware or even just software bugs. Such malware might try to compromise a voter's privacy, change the voter's recorded intent, or might simply fail to work. Any of these issues could be engineered by malware to achieve a partisan shift in the outcome of the election. This leads to a very general, and still very much open, challenge in computer security: verifying that a given computer is free of malware and/or that it is free of vulnerabilities that might allow for the introduction of malware.

There is a specific solution that is proposed for voting systems called *code voting*, where the voter receives a series of *codes*, ahead of the election, and only enters the codes for their chosen candidates. This makes it impossible for a computer to change the voter's intent because it never knows the codes for any but the chosen candidates. Code voting schemes, of course, require an independent channel to transmit these codes, and they appear to have weaker usability properties. However, the broad idea remains that we can design voting-specific solutions that avoid some risks of client malware without needing to solve the general-purpose problem. This area needs further research exploration.

Better E2E cryptographic technologies: Current E2E-verifiable voting technologies, such as Microsoft's ElectionGuard (Burt, 2019), are built on well-understood public-key cryptographic primitives that were established in the academic literature decades ago and are widely used in real systems today. Nonetheless, there are several interesting research challenges in making them better. For example, every modern E2E system relies on the selection of truly random numbers as part of the encryption process. If an adversary can predict these random values, then they can decrypt the ciphertexts without needing to know the cryptographic keys. This leads to a challenge of how we might have *verifiable* randomness, or otherwise leverage some specific property of voting systems to ensure the strength of how random numbers are selected.

Similarly, we expect that remote voting technologies will inevitably rely on remote authentication mechanisms commonly used on the Internet, like OpenID and OAuth. In a remote vote center or kiosk, it might be possible to conduct the user authentication on a separate computer from the voting machine. While it is natural to imagine an air gap to ensure no data about user identification flowing to the voting machine, there are interesting opportunities for allowing information across in controlled ways, such as to allow for *remote provisional voting*, where an election official would be able to include or exclude a vote based on the identity of the voter, while not being able to see that voter's preferences.

Formal verification and correctness issues: When we ponder the design of a cryptographically verifiable election technology with ballots flowing through the postal mail and cryptographic receipts, we are fundamentally designing a *distributed system protocol*. There's a long history of these kinds of protocols having subtle issues discovered, sometimes decades after they were first proposed. In recent years, a variety of *formal verification tools* have become widely adopted to automatically discover these kinds of vulnerabilities. There are important research challenges to apply formal verification tools and techniques toward understanding the kinds of voting systems and networked pollbook systems that we're considering. Of particular interest, voting systems inevitably rely on a mix of different security guarantees at different stages of the tabulation process. Formal modeling and verification of the entire system, including steps performed by humans and computers, will help identify unstated assumptions or even unnecessary steps in these processes, allowing for both more efficient and more secure procedures.

Equity issues: It's entirely possible that we might reach a security conclusion that requires minimum platform levels for security purposes. This effectively excludes voters using older technologies. For example, if you are poor, you probably do not have the newest smartphone with the latest security features. This creates a difficult challenge where researchers need to identify mechanisms that have the desired security properties without simultaneously creating a barrier to voting.

5. Other Research Questions in Voting Systems

In the world of voting security, we could easily raise important concerns from adjacent fields that need to be addressed. This section briefly discusses some of these adjacent concerns and points the reader to additional information about them.

Once every decade states redraw their Congressional and legislative districts based on updated census information. This creates opportunities to draw these maps for partisan gains ("gerrymandering"). Understanding this process, and trying to define legal standards for "fair" maps is an active research area. (See Election Science: A Proposed NSF Convergence Accelerator, Alvarez et al., 2021.) Another important topic is voter registration, the process of determining who is and is not allowed to vote, and voter authentication, the process of determining that a person who presents to vote corresponds to a specific record in a voter registration database.

To the extent that voting systems are computer systems, they are subject to all the same issues with any computer systems that might have bugs, and they might be vulnerable to security exploits. Much of the research in electronic voting is focused on techniques that *mitigate* against bugs and vulnerabilities by providing external evidence that an election has the correct outcome. However, ongoing research into creating "high assurance" computer systems, both hardware and software, is certainly relevant to computerized voting systems. If we had techniques to provably write software without bugs or vulnerabilities and run it on hardware without bugs or vulnerabilities, and if these properties could be externally verified on running systems, then we would have correspondingly better and more reliable voting systems. Please see the related CCC quadrennial paper *A Research Ecosystem for Secure Computing* (Bliss et al. 2020a) for more information on high-level research challenges in computer security.

Similarly, misinformation and disinformation play an unfortunate and growing role in modern elections, among other topics in modern society. Please see the related CCC quadrennial paper *An Agenda for Disinformation Research* (Bliss et al. 2020b), which discusses the research needs in this area and possible interventions to limit the spread of disinformation.

Lastly, we note that there is an active and engaging research area called “social choice theory,” which considers the design of exactly how voters might specify their preferences, the mathematics of how those votes are aggregated to determine winners, and the corresponding incentives that candidates might face to adjust their policies to attract more voters. It is certainly the case that some of these systems would be more complex to audit and verify than traditional systems, but if other voting designs become popular then researchers can and will explore methods to ensure that they can be audited and verified.

6. References and Related Work

Alvarez, R. M., Duchin, M., Macht, G., & Stewart, C., III. (2021). Election Science: A Proposed NSF Convergence Accelerator (No. 12). CalTech/MIT. https://vote.caltech.edu/reports/12

Appel, A. W., DeMilo, R. A., & Stark, P. B. (2020). Ballot-Marking Devices Cannot Ensure the Will of the Voters. Election Law Journal: Rules, Politics, and Policy, 19(3), 432–450.

Appel, A. W., & Stark, P. B. (2020). Evidence-Based Elections: Create a Meaningful Paper Trail, Then Audit. Georgetown Law Technology Review, 523(4), 523–541.

Bell, S., Benaloh, J., Byrne, M. D., DeBeauvoir, D., Eakin, B., Fisher, G., Kortum, P., McBurnett, N., Montoya, J., Parker, M., Pereira, O., Stark, P. B., Wallach, D. S., & Winn, M. (2013). STAR-Vote: A Secure, Transparent, Auditable, and Reliable Voting System. USENIX Journal of Election Technology and Systems, 1(1), 18–37.

Benaloh, J., Jones, D., Lazarus, E. L., Lindeman, M., & Stark, P. B. (2011). SOBA: Secrecy-preserving Observable Ballot-level Audit. Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE ’11). https://static.usenix.org/events/evtwote11/tech/final_files/Benaloh.pdf

Benaloh, J., Stark, P. B., & Teague, V. (2019). VAULT: Verifiable Audits Using Limited Transparency. https://www.stat.berkeley.edu/~stark/Preprints/vault19.pdf

Bernhard, M., Benaloh, J., Halderman, J. A., Rivest, R. L., Ryan, P. Y. A., Stark, P. B., Teague, V., Poorvi, V. L., & Wallach, D. S. (2017). Public Evidence from Secret Ballots. Electronic Voting, 84–109.

Bliss N., Gordon L., Lopresti D., Schneider F., & Venkatasubramanian S. (2020a). A Research Ecosystem for Secure Computing. https://cra.org/ccc/resources/ccc-led-whitepapers/#2020-quadrennial-papers

Bliss N., Bradley E., Garland J., Menczer F., Ruston S., Starbird K., & Wiggins C. (2020b). An Agenda for Disinformation Research. https://cra.org/ccc/resources/ccc-led-whitepapers/#2020-quadrennial-papers

Burt, T. (2019, May 6). Protecting democratic elections through secure, verifiable voting. https://blogs.microsoft.com/on-the-issues/2019/05/06/protecting-democratic-elections-through-secure-verifiable-voting/

Chaum, D. (1981). Untraceable Electronic Mail, Return Addresses and Digital Pseudonyms. In Advances in Information Security (pp. 211–219). https://doi.org/10.1007/978-1-4615-0239-5_14

Chaum, D., Carback, R., Clark, J., Essex, A., Popoveniuc, S., Rivest, R. L., Ryan, P. Y. A., Shen, E., & Sherman, A. T. (2008). Scantegrity II: End-to-End Verifiability for Optical Scan Election Systems using Invisible Ink Confirmation Codes. EVT, 8, 1–13.

Glazer, A. K., Spertus, J. V., & Stark, P. B. (2021). More Style, Less Work: Card-style Data Decrease Risk-limiting Audit Sample Sizes. Digital Threats: Research and Practice, 2(4). https://dl.acm.org/doi/pdf/10.1145/3457907

Higgins, M. J., Rivest, R. L., & Stark, P. B. (2011). Sharper p-Values for Stratified Election Audits. In Statistics, Politics, and Policy (Vol. 2, Issue 1). https://doi.org/10.2202/2151-7509.1031

Jamroga, W., Ryan, P.Y.A.,Schneider, S., Schürmann, C., & Stark, P.B. (2022). A Declaration of Software Independence, in Festschrift for Joshua Guttman, Springer-Nature, to appear.

Kulyk, O., Henzel, J., Renaud, K., & Volkamer, M. (2019). Comparing “Challenge-Based” and “Code-Based” Internet Voting Verification Implementations. In Human-Computer Interaction – INTERACT 2019 (pp. 519–538). https://doi.org/10.1007/978-3-030-29381-9_32

Lindeman, M., & Stark, P. B. (2012). A Gentle Introduction to Risk-Limiting Audits. IEEE Security & Privacy (Vol. 10, Issue 5, pp. 42–49). https://doi.org/10.1109/msp.2012.56

National Academies of Sciences, Engineering, and Medicine. (2018). Securing the Vote Protecting American Democracy. The National Academies Press. https://www.nationalacademies.org/our-work/the-future-of-voting-accessible-reliable-verifiable-technology

Ottoboni, K., Bernhard, M., Halderman, A., Rivest, R.L., and Stark, P.B. (2019). Bernoulli Ballot Polling: A Manifest Improvement for Risk Limiting Audits, Proceedings of the 4th Annual Workshop on Advances in Secure Electronic Voting (Voting’19).

Ottoboni, K., Stark, P. B., Lindeman, M., & McBurnett, N. (2018). Risk-Limiting Audits by Stratified Union-Intersection Tests of Elections (SUITE). In Electronic Voting (pp. 174–188). https://doi.org/10.1007/978-3-030-00419-4_12

Rivest, R. L., & Wack, J. P. (2008). On the notion of “software independence” in voting systems. Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences, 366(1881), 3759–3767.

P. Y. A. Ryan, D. Bismark, J. Heather, S. Schneider and Z. Xia. (2009). Prêt à Voter: a Voter-Verifiable Voting System. In IEEE Transactions on Information Forensics and Security (Vol. 4, No. 4, pp. 662-673). doi: 10.1109/TIFS.2009.2033233.

Sasse, M. A., & Rashid, A. (2019). Human Factors Knowledge Area Issue 1.0. https://www.cybok.org/media/downloads/Human_Factors_issue_1.0.pdf

Stark, P. B. (2008). Conservative statistical post-election audits. In The Annals of Applied Statistics (Vol. 2, Issue 2). https://doi.org/10.1214/08-aoas161

Stark, P. B. (2009). CAST: Canvass Audits by Sampling and Testing. In IEEE Transactions on Information Forensics and Security (Vol. 4, Issue 4, pp. 708–717). https://doi.org/10.1109/tifs.2009.2034210

Stark, P. B. (2020). Sets of Half-Average Nulls Generate Risk-Limiting Audits: SHANGRLA. In Financial Cryptography and Data Security (pp. 319–336). https://doi.org/10.1007/978-3-030-54455-3_23

Stark, P. B., & Wagner, D. (2012). Evidence-Based Elections. In IEEE Security & Privacy (Vol. 10, Issue 5, pp. 33–41). https://doi.org/10.1109/msp.2012.62

Tuinstra, D., & Benaloh, J. (1994). Receipt-Free Secret-Ballot Elections. STOC '94 Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 544–553.

Ville, J. (1939). Étude critique de la notion de collectif, Monographies des Probabilités, 3, Gauthier-Villars, Paris.

Volkamer, M., & Renaud, K. (2013). Mental Models – General Introduction and Review of Their Application to Human-Centred Security. In Number Theory and Cryptography (pp. 255–280). Springer, Berlin, Heidelberg.

Wald, A. (1945). Sequential Tests of Statistical Hypotheses. The Annals of Mathematical Statistics, 16(2), 117–186.

Waudby-Smith, I., Stark, P. B., & Ramdas, A. (2021). RiLACS: Risk Limiting Audits via Confidence Sequences. In R. Krimmer, M. Volkamer, D. Duenas-Cid, O. Kulyk, P. Rønne, M. Solvak, & M. Germann (Eds.), 6th International Joint Conference, E-Vote-ID 2021, Virtual Event, October 5–8, 2021, Proceedings. Springer International Publishing. Retrieved December 13, 2021, from https://www.springerprofessional.de/en/rilacs-risk-limiting-audits-via-confidence-sequences/19702902

Zagórski, F., McClearn, G., Morin, S., McBurnett, N., & Vora, P. L. (2021). Minerva– An Efficient Risk-Limiting Ballot Polling Audit. Proceedings of the 30th USENIX Security Symposium, 3059–3076.

For citation use: Benaloh J., Stark P.B., Teague V., Volkamer M., and Wallach D. (2022). *Research Opportunities in Evidence-Based Elections*. A white paper prepared for the Computing Community Consortium committee of the Computing Research Association. http://cra.org/ccc/resources/ccc-led-whitepapers/