

# Checking election outcome accuracy

## Post-election audit sampling

Kathy Dopp

kathy.dopp@gmail.com

Comments: 37 pages + 19 pages of appendices & references, 3 figures, 10 tables

Subj-class: stat.AP

July 18, 2009

### Abstract

This article advances and improves existing post-election audit sampling methodology. Methods for determining post-election audit sampling have been the subject of extensive recent research. This article

- provides an overview of post-election audit sampling research and compares various approaches to calculating post-election audit sample sizes, focusing on risk-limiting audits,
- discusses fundamental concepts common to all risk-limiting post-election audits, presenting new margin error bounds, sampling weights and sampling probabilities that improve upon existing approaches and work for any size audit unit and for single or multi-winner election contests,
- provides two new simple formulas for estimating post-election audit sample sizes in cases when detailed data, expertise, or tools are not available,
- summarizes four improved methods for calculating risk-limiting election audit sample sizes, showing how to apply precise margin error bounds to improve the accuracy and efficacy of existing methods, and
- discusses sampling mistakes that reduce post-election audit effectiveness.

Adequate post-election audit sampling is crucial because analyzing discrepancies found in too-small samples can determine little except that the sample size is inadequate. This article is one of three articles in a series *Checking Election Outcome Accuracy*. The other two articles discuss post-election auditing procedures and an algorithm for deciding whether to increase the sample or to certify the election outcome in response to any discrepancies found during a post-election audit.

## 1 Introduction

In any field the primary purpose of auditing is to detect incorrect results due to unintended innocent or deliberate acts by insiders such as administrators or computer programmers.

This article defines “post-election audit” as a check of the accuracy of reported election results done by manually counting all the voter-verifiable paper ballots associated with

randomly sampled reported initial vote counts, and checking such additional records as necessary to ensure the integrity of the electoral process.

“Risk-limiting” post-election audits are election audits that are designed to provide a minimum high probability that incorrect initial election outcomes are detected and corrected before the final certification of election results.

Election winners control budgets and contracts worth millions to trillions of dollars, so this article assumes that election rigging could occur by miscounting the minimum number of initial reported vote counts that could cause an incorrect election outcome (an incorrect winner).

## **Definitions**

An “audit unit” or “auditable vote count” is defined in this article as a tally of votes that is publicly reported for an election contest. Each tally is obtained from a group of one or more ballots that are:

- counted at one place and time or
- counted by one voting device, or
- cast by voters who live in the same voting precincts or districts, or
- are obtained from publicly reported votes on single ballots that sum to the reported results.

Audit units can be precinct vote counts, electronic voting device counts, or batches or decks of paper ballots. Audit units can be counted by hand or by automatic tabulating equipment where each tally is associated with a number of ballots maintained as a group. An audit unit or auditable vote count may be an individual ballot only if the voting system produces a public report of vote counts on each ballot with humanly readable identifiers for individual ballots and yet preserves ballot privacy.

An “audit sample size” is the number of audit units that are randomly drawn for manually counting and comparing with the initial reported audit units.

Under and over-votes are cast ballots eligible to vote in the contest that have no valid vote counted on them for any candidate.

## Background

In 1975, ahead of his time, Roy Saltman proposed conducting post-election audits using sample sizes that would detect an amount of miscount that could cause an incorrect election outcome, and suggested a formula for estimating the minimum number of miscounted auditable vote counts that could cause an incorrect election outcome (Saltman, 1975, appendix B).<sup>1</sup>

Saltman’s work and the topic of election auditing was largely neglected until more recently when political scientists, mathematicians, and computer scientists began to recommend that “rather than relying on ad hoc detection and litigation of electoral problems” that we should “systematically monitor and audit elections in a preventive fashion” (Mebane, Jr. *et al.*, 2003; Jones, 2004; Dill, 2005; Dopp & Baiman, 2005–2006; Democracy & Management, 2005; US General Accounting Office, 2005).

By 2006, the US League of Women Voters membership voted to recommended post-election auditing and the National Institute of Standards and Technology and The US Election Assistance Commission’s Technical Guidelines Development Committee recommended a variety of safeguards for voting systems for the 2007 Voluntary Voting Systems Guidelines (Project ACCURATE, 2005, pp. 18–20, 40–41), (Rivest & Wack, 2006; Burr *et al.*, 2006; NIST staff, 2007; Burr, 2007).

In July 2006 Saltman’s formula was re-discovered in a modified form that considers the possibility of miscounted over and undervotes<sup>2</sup> and a numerical computer algorithm was provided for doing these calculations when vote count size varies (Dopp & Baiman, 2005–2006; Dopp & Stenger, 2006).<sup>3</sup> These methods rely on uniform random sampling.

Neff and Wand (Neff, 2003; Wand, 2004) had showed that the smaller the size of reported vote counts (in number of ballots), the fewer the total number of ballots that need to be audited to achieve any probability for detecting the same level of vote miscount and computer scientists proposed sampling individual ballots to make risk-limiting audits more efficient (Walmsley, 2005; Calandrino *et al.*, 2007b).<sup>4</sup> However, most current voting system tabulators are designed to produce reports only of precinct vote counts, thus making it

---

<sup>1</sup>Saltman calculated the minimum number of minimum miscounted audit units that could cause an incorrect election outcome by using the margin as a percentage of votes cast divided by two times a “maximum level of vote switching that would be undetectable by observation”.

<sup>2</sup>The newer formula handled over and under-votes by calculating the margin as a percentage out of the total number of ballots cast, rather than out of votes counted

<sup>3</sup>However some authors continued to recommend using Saltman’s original method based on the number of votes counted (McCarthy *et al.*, 2007; Norden *et al.*, 2007b), (Hall, 2008b, Appendix D)

<sup>4</sup>In other words, risk-limiting election audits require less work for the same benefit when a larger number of smaller-sized vote counts (audit units) are initially publicly reported and sampled.

difficult to report and to sample smaller-sized audit units (Dopp, 2009).

Two groups of computer scientists developed weighted sampling methods for post-election auditing in order to be able to sample fewer ballots yet achieve the same probability for detecting incorrect outcomes, by targeting ballots having more potential for producing margin error (Calandrino *et al.*, 2007b; Aslam *et al.*, 2008).

In December 2007, a more precise calculation method for post-election audit sample sizes and sampling weights was developed by using upper margin error bounds for the just-winning and just-losing candidate pair (Dopp, 2007–2008c; Aslam *et al.*, 2008; Dopp, 2008), (Stark, 2008c, p. 13).<sup>5</sup>

Once sampling weights are determined, fair and efficient methods for making random selections for audit units have been developed (Cordero *et al.*, 2006; Calandrino *et al.*, 2007a; Aslam *et al.*, 2008; Hall, 2008a; Rivest, 2008) although there is some debate among election integrity advocates as to which selection methods are preferred, the more understandable methods such as rolling ten-sided dice, or computer methods such as pseudo-random number generators that are more efficient and may be more verifiably fair.

## Assumptions

*For the purposes of discussing post-election audit sampling methods, it is assumed that:*

- Voter-marked paper ballots are used to provide a durable record of voter intent;
- Effective chain of custody and security procedures are used to prevent and detect any illicit addition, subtraction, substitution, or tampering with ballots and other audit records; and
- Effective procedures are used during a post-election audit so that the manual counts are therefore accurate, and that when differences are found, at a minimum, recounts are performed until two counts agree — either the machine and a manual count or two manual counts.<sup>6</sup> See (Deputy Secretary of State Anthony Stevens, 2007; Dopp & Straight, 2006–2008; Dopp, 2009).

---

<sup>5</sup>Failure to use accurate within audit unit upper margin error bounds translates to a failure to meet the assumptions that are used to determine the sample. Within audit unit margin error bounds are used in all risk-limiting post-election auditing methods for calculating audit sample sizes and sampling weights, and for analyzing the discrepancies found in the audit.

<sup>6</sup>Luther Weeks, in *Results of Post-Election Audit of the May 4th Municipal Election* <http://www.ctvoterscount.org/?p=2077> suggests that recounts are performed until two counts agree.

Table 1: Post-Election Auditing Method Comparison

Method	Purpose	Sample Size	Effectiveness
<b>Fixed Rate Audit</b>	To ensure that voting machines are accurate to within a specified tolerance	A fixed percentage of publicly reported audit units	Effectiveness at detecting inaccurate initial outcomes ranges widely
<b>Risk-limiting Audit</b>	To ensure that election outcomes are accurate	Varies from one (1) to all audit units, as needed to detect incorrect election outcomes to a desired probability	Provides roughly equal probability (e.g., at least 95%) that any incorrect election outcomes are detected
<b>Manual Recount</b>	To ensure that all votes are accurately counted	100% of audit units	Provides 100% assurance of detecting incorrect election outcomes

## 2 Post-Election Auditing Approaches

### Three approaches to post-election auditing

The appropriate sample size for conducting post-election audits depends on the audit’s purpose. Table 1 compares three approaches to checking the accuracy of reported election results.

*If the purpose of an election audit is to ensure that:*

- election outcomes are accurately decided, then *risk-limiting* audits achieve that purpose effectively and efficiently. Risk-limiting post-election audits provide a desired minimum probability that the audit sample will contain one or more miscounted audit units whenever the minimum amount of miscount occurs that could cause an initial incorrect election outcome.
- voting machines have counted election results accurately to within a certain tolerance level, then a *fixed rate* audit is the solution. Commonly used in manufacturing, fixed rate audits work best when any errors are widely distributed and typically use a larger sample in wide-margin election contests, and a smaller sample size in close margin contests than do risk-limiting audits.
- every eligible vote is accurately counted, then a *100% manual* audit or recount is the best solution.

Figures 1 and 2 compare the efficiency and effectiveness of “fixed rate” versus “risk-limiting” post-election audits using a 500 precinct election contest with 150,000 total ballots

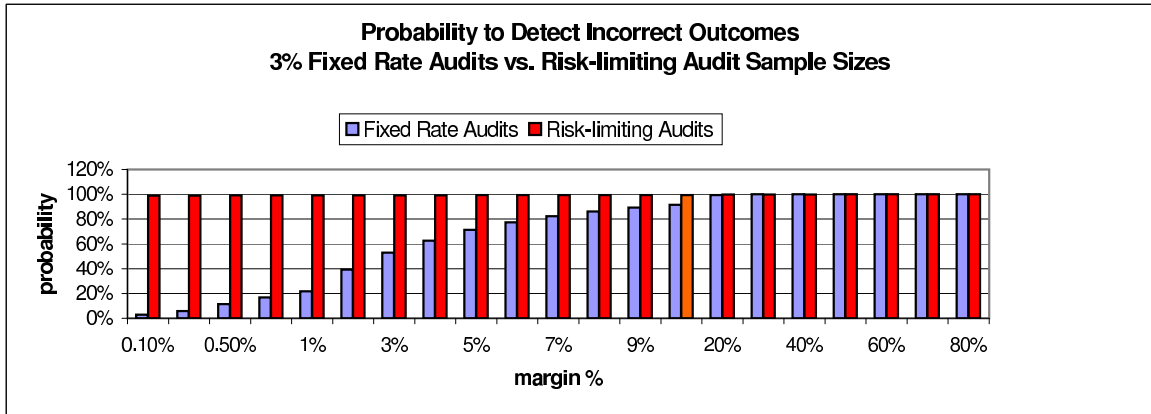


Figure 1: Probability for Detecting Incorrect Election Outcomes

cast and various initial margins.<sup>7</sup>

Figure 1 shows that 3% fixed rate audits (blue bars) provide unequal probabilities for detecting the smallest amount of miscount that could cause incorrect outcomes. A 3% flat rate audit provides very low chance to detect inaccurate outcomes in close contests (less than 10% probability in some cases), but provides very high minimum probabilities (virtually 100% in most cases) for detecting miscount that could cause incorrect outcomes in wide-margin contests. On the other hand, risk-limiting audits (red bars in Figure 1) provide roughly equal assurance to all candidates and voters that any outcome-changing vote miscount is detected and corrected regardless of the winning margins or the number of precincts.

Figure 2 shows how sample sizes for risk-limiting audits increase as winning margins decrease. Thus risk-limiting audits could eliminate the need for automatic recounts because all sufficiently close-margin election contest would automatically receive a 100% manual count whenever necessary for ensuring that the election outcome were correct. A fixed 3% audit (blue bars) samples more than is necessary for ensuring the correctness of wide-margin US House outcomes. In fact, the total overall amount of vote counts that would be audited nationwide for US House contests would be roughly equal for 3% nationwide fixed rate audits and for 99% risk-limiting audits.

There are different ways to categorize post-election auditing approaches based on the sampling approach. Some authors categorize risk-limiting audits that are designed to detect

<sup>7</sup>The audit sample sizes for risk-limiting audits in Figures 1 and 2 are calculated using the uniform estimate method presented later in this article.

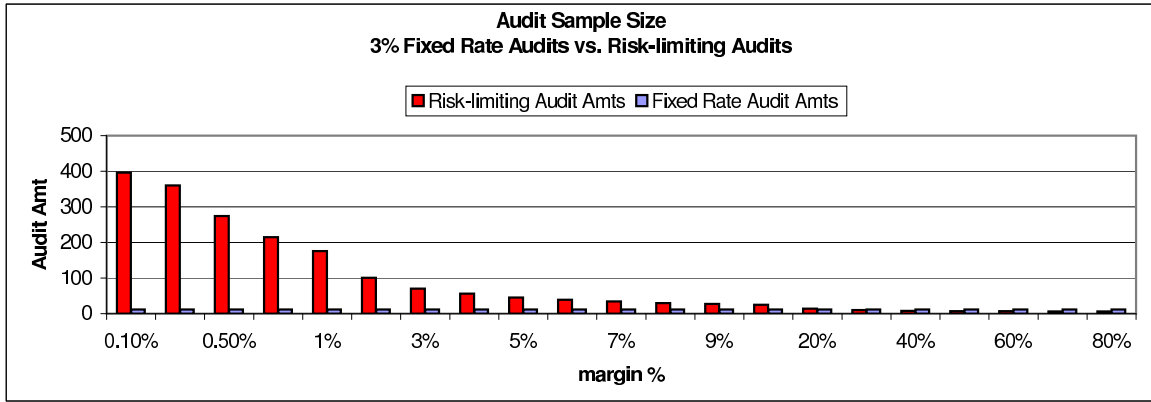


Figure 2: Sample Size Comparison

incorrect election outcomes in a category called “Variable (or Adjustable) Rate Audits”, along with tiered flat rate audits<sup>8</sup> that do not provide any minimum probability for detecting incorrect election outcomes (Norden *et al.*, 2007b), (Norden *et al.*, 2007a, p. 19), (Hall, 2008b, p. 73). This article categorizes risk-limiting election audits that *are* designed to limit the risk of certifying inaccurate election outcomes in a separate category from election audits that do not limit the risk of certifying incorrect election outcomes to a desired low probability.

Despite the recent development of risk-limiting post-election auditing methods, most States have adopted, and some authors continue to recommend fixed rate audits designed to detect at most certain levels of error (Appel, 1997; Norden *et al.*, 2007b; Norden *et al.*, 2007a; Atkeson *et al.*, 2008).

### 3 Risk-limiting Election Audits

The first risk-limiting post-election audit in the US was conducted in Cuyahoga County, Ohio in 2006<sup>9</sup> (The Collaborative Audit Committee, 2007; Dopp, 2007a). Since that time citizen groups in various States are having limited success convincing election officials and legislators to implement risk-limiting post-election audits, and some have been conducted in States such as Colorado and California (Stark, 2008b; McBurnett, 2008; Hall *et al.*, 2009).

<sup>8</sup>An example of a “tiered” flat rate audit is the audit proposed by Larry Norden and Representative Rush Holt, D-NJ that audits 3%, 5%, or 10% of precincts depending on the margin percentage between the just-winning and just-losing candidate pair. This particular proposal gives as low as a 10% chance of detecting incorrect outcomes when measured against 2004 US House contest election results.

<sup>9</sup>However, the Cuyahoga County auditors did not correctly analyze the discrepancies found by the audit based on their sample size design.

In January 2009 The League of Women Voters of the United States endorsed risk-limiting audits, stating that, “The number of audit units to audit should be chosen so as to ensure there is only a small predetermined chance of confirming an incorrect outcome.” (The Election Audits Task Force, 2009)

All post-election audit sample sizes and sampling weights are estimates because *all* risk-limiting post-election audit sample sizes and selection weights depend on inputs that are estimates (such as estimates for the number of miscounted audit units that could cause an incorrect election outcome or estimates for the amount of maximum within audit unit margin error that would not be immediately seen as suspicious.) The methods proposed in this article improve upon the accuracy and conservatism of these estimates.

There is more than one method for calculating risk-limiting post-election audit sample sizes. Which method is appropriate depends on the answers to questions such as:

- Are initial detailed audit unit and ballot data available for all audit units?
- Is a computer program or spreadsheet available to do the detailed precise calculations?
- Will the random sample be drawn using a uniform probability distribution or using a weighted sampling method?
- Do we need a quick estimate for planning purposes or do we need the more precise audit amount that achieves at least the desired minimum probability of detecting an incorrect outcome?

### 3.1 Methods common to risk-limiting post-election audits

Concepts common to *all* risk-limiting post-election audits include “maximum level of undetectability”, and two types of upper margin error bounds – bounds specific to one particular winning-losing candidate pair and margin error bounds for *all* winning-losing candidate pairs.

#### Maximum level of undetectability

To reduce chances of detection by a post-election audit, a perpetrator could possibly miscount the smallest number of total audit units possible to cause an incorrect election outcome (Saltman, 1975; Dopp & Stenger, 2006).<sup>13</sup> However a perpetrator cannot miscount

---

<sup>13</sup>For instance several popular voting systems allow vote counts to be easily manipulated in under a minute with access to the central tabulator, leaving no log record.

Table 2: This table contains all the variables used in this article to calculate risk-limiting election audit sample sizes and sampling weights.

TABLE OF VARIABLES			
Variable Name	Variable Letter	Formula	Description
Ballots cast	for each audit unit $b_i$ and in total $b$	$b = \sum_i b_i$	the total number of ballots cast that are eligible to vote in the election contest
Votes counted for a winning candidate	for each audit unit $w_i$ and in total $w$	$w = \sum_i w_i$	the total number of votes counted for the winning candidate
Votes counted for the just-losing candidate	for each audit unit $r_i$ and in total $r$	$r = \sum_i r_i$	the total number of votes counted for the losing candidate who has the most initial votes
Votes Counted for Losing Candidates & Under and Over-votes	for each audit unit $l_i$ and in total $l$	$l = \sum_i l_i$	the total number of votes counted for any losing candidate plus the total number of ballots with no votes recorded
Margin	$M$ between a winning-losing candidate pair	$w - r$ $= \sum_{i=1}^N w_i - \sum_{i=1}^N r_i$	the difference between the number of votes counted for a winning candidate and the number of votes counted for a losing candidate <sup>10</sup>
Percentage margin	$m$	$M/b$	the margin divided by the total number of ballots cast eligible to vote in the election contest
Margin error	$e_i$	$e_i = (w_i - r_i) - (w_a - r_a)$	the difference between the reported initial margins and the audit margins
Margin error upper bound	for each audit unit $u_i$ or <code>error_bound(i)</code> and for all audit units $E = \sum_{j=1}^N \text{error\_bound}(j)$	Formula varies, depending on methods and purpose. Total error bound is $E = \sum_i u_i$	the maximum amount of margin error in #ballots, within audit units or in total that could reverse an an election outcome
Total number of audit units	$N$		the total number of reported audit units <sup>11</sup> in the contest
Maximum level of Undetectability for margin error	$k$ or $MLU$	a constant $k$ : $0 < k < 1$	an assumed maximum rate of margin error that could occur without raising enough suspicion to be detected without an audit <sup>12</sup>
Confidence probability	$P$	Suggest $0.95 \leq P \leq 1$ See appendix C	the desired minimum probability that the audit sample will detect one or more miscounted audit units if an initial election outcome is incorrect
Audit unit random selection probability	$p_i$	$0 \leq p_i \leq 1$	the probability that an audit unit will be randomly selected
The number of miscounted audit units	$C$	methods vary See Table 7	the minimum number of miscounted audit units that could cause an incorrect election outcome
Post-election audit sample size	$S$	methods vary See Table 7	the election audit sample size or number of audit units to manually count and compare with reported results

all the votes within any one audit unit because if all available votes were switched to count for a perpetrator’s favorite candidate then anyone who had voted for another candidate would immediately know that the election results were incorrect. Hence, to avoid detection a perpetrator would miscount at most some maximum rate  $k : 0 < k < 1$  of the available margin error.

Thus we assume a maximum level of undetectability  $k$ , a maximum rate of margin error, such that if more margin error than  $k$  times the upper margin error bound occurs, it would look suspicious and cause immediate action by election officials or by candidates and their supporters (Saltman, 1975, appendix B).

Risk-limiting post-election audit design assumes that a maximum rate  $k \leq 1$  of the total available margin error within each audit unit<sup>14</sup> could be miscounted in favor of an initial winner without raising immediate suspicion. In uniform sampling methods, this maximum level of undetectability assumption is used to estimate the minimum number of miscounted audit units that could cause an incorrect election outcome.

For multi-ballot audit units, the larger the maximum level of undetectability as  $k \rightarrow 1$ , the fewer miscounted audit units it takes to cause an incorrect election outcome; and the larger the audit sample size  $S$  must be in order to ensure that one or more miscounted audit units are sampled whenever an initial election outcome is incorrect.<sup>15</sup> Even for one-ballot audit units, we still need to use a maximum level of undetectability,  $k : 0 < k \leq 1$ , because we know that a perpetrator cannot target 100% of all ballots having a particular vote pattern without being immediately noticed.

A crucial consequence of using a “maximum level of undetectability” when calculating risk-limiting post-election audit sample sizes is the necessity of allowing candidates or their representatives to select one or more suspicious-looking additional audit units for auditing in addition to the randomly selected sample.

The “maximum level of undetectability” is always multiplied times the maximum margin error available within each audit unit to get the most margin error that it is believed could reasonably occur without immediately being detected as suspicious.

One proposed approach is to use 1 or 100% for the maximum level of undetectability (MLU) in sample size calculations (Stark, 2008c; Stark, 2008d; Stark, 2008a; Hall *et al.*, 2009). Using an MLU of one would result in very conservative (large) sample sizes if the

---

<sup>14</sup>or that a maximum rate of the total number of each type of voted ballots could be miscounted when individual ballots are used as audit units

<sup>15</sup>As real-life post-election audits are conducted, more may be learned about what rates of miscount are not seen as suspicious and what values are appropriate for the maximum level of undetectability  $k$ .

methods suggested in this article were used. However, those who recommend using an MLU of one cancel by division any MLU multiplier (Hall *et al.*, 2009, p. 15), or use tests that cancel any MLU from both sides of an inequality involving ratios of their upper margin error bound measures (Stark, 2009b, p. 6–10) (Stark, 2009a, p. 4), or otherwise negate any conservative effects of using an MLU of one. This cancellation is similar to how the maximum level of undetectability cancels when calculating the sampling weights for the probability proportional to margin error bound with replacement (PPMEBWR) method described below.

Assuming a 100% maximum level of undetectability is of course unreasonable because some voters and candidates would immediately notice the fact that there were zero votes cast in their precincts for their candidates. A perpetrator can not expect to steal 100% of available target votes for a candidate and not be noticed.

An important consequence of assuming a “maximum level of undetectability” of less than one is the necessity of allowing losing candidates to select discretionary audit units to be manually audited at the same time as are the randomly selected audit units. The necessity of auditing discretionary audit units is discussed further in section 6 in this article.

## Upper margin error bounds

*Why are within audit unit upper margin error bounds important?*

Within audit unit upper margin error bounds are a crucial input to all calculation methods for determining post-election audit sample sizes and sampling weights.

Aslam, Popa, and Rivest derived precise audit unit upper margin error bounds for particular winning-losing candidate pairs in an intermediate calculation, but nonetheless recommended using Saltman’s earlier method of multiplying two (2) times a maximum level of undetectability of 20%, times the number of votes cast,  $v$ , to approximate the maximum undetectable margin error for their sample size calculations. Dopp derived the exact same within audit unit upper margin error bounds,  $b + w - r$  (the number of ballots plus the margin in votes), and recommended using them to improve the accuracy of post-election audit sample size calculations. Later Stark also recommended using within audit unit upper margin error bounds in his unique calculations. This paper extends the increased precision of using error bounds by recommending the use of the same margin error bound for calculating sample sizes, but suggests a more appropriate margin error bound for weighting random selections.

For risk-limiting audits, when audit units are larger than one ballot, most authors

continue to use or recommend using the less precise expression  $2sv$  for approximating maximum undetectable within audit unit margin error (Calandrino *et al.*, 2007b, p. 5), (Aslam *et al.*, 2008, p. 16), (Saltman, 1975, p. 5 of Appendix B), (McCarthy *et al.*, 2007, p. 6 and Appendix B), (Stanislevic, 2006, pp. 6–10, 15), (Norden *et al.*, 2007a, Appendix B), (Hall, 2008b, p. 153, Appendix D), (The Election Audits Task Force, 2009, pp. 11, 28–30), (McBurnett, 2008).

Using less precise margin error bounds can produce insufficient sample sizes and, if used for sampling weights, can cause the audits to unfairly favor some candidates over others. For example, using the expression  $2sv$  to approximate the maximum amount of undetectable margin error:

1. ignores the partisanship of within audit unit vote counts. The partisanship of the vote counts affect the amount of maximum margin error that can exist between specific winning-losing candidate pairs and between all winning-losing pairs,
2. calculates an impossible amount of error when certain vote shares occur — more than the possible margin error that is available to contribute to causing an incorrect outcome — in cases when just-losing candidate vote share is high,
3. does not account for the unequal amounts of margin error that results from different causes such as shifting votes from a winner to a loser or vice-versa, shifting votes between two different losing candidates or between different winning candidates, and not counting votes for a losing or winning two candidates. Each miscounted vote may produce a vote margin error of -2, -1, 0, 1, or 2 in the initial margin for a particular winning-losing candidate pair. There is no way to accurately express these variations precisely in terms of votes or ballots cast by using an expressions like  $2v$  or  $2b$  without making awkward, unnecessary assumptions about the relative proportions of each type of error (such as vote-switching between two winners, a winner and a loser, or two losers or simply not counting votes);
4. by using votes cast,  $v$ , fails to consider miscounted under or over-votes (although that particular problem could be corrected by using ballots cast,  $b$ );
5. significantly understates the possible margin error in most cases because actual upper margin error can be as high as 200% of the number of ballots (Eg. 40% times  $0 \leq s \leq 1$  times the number of cast votes significantly under-states 40% times the amount of possible within audit unit margin error); and

Thus using  $2sv$  or  $2sb$  where  $0 \leq s \leq 1$  to estimate the maximum margin error that is likely not to raise suspicion underestimates the sample size in close margin contests, or equivalently over-states the minimum probability for detecting vote miscount that could cause an incorrect election outcome.<sup>16</sup>

This article presents two types of precise within audit unit upper margin error bounds that are should be used when calculating risk-limiting post-election audit sample sizes and sampling weights:

1. upper margin error bounds for a specific winning-losing candidate pair, and
2. upper margin error bounds for the error that can occur between any candidate pair

Figure 7 in Appendix D graphically compares the  $2sv$  margin error measure with the actual upper error bounds.

### Upper margin error bounds for specific winning-losing candidate pairs

The margin error between any winning-losing candidate pair is defined as the signed difference between their initial reported margin and their margin found during a 100% manual audit. Within each initial audit unit  $i$ , let

- $b_i$  = the number of ballots cast,
- $w_{ij}$  = the number of initial votes for winning candidate,
- $l_{ij}$  = the number of initial votes for losing candidate,
- $w_{aj}$  = the number of audit votes for winning candidate,
- $l_{aj}$  = the number of audit votes for losing candidate.

Then the margin error within each audit unit  $i$  found during the audit between winning-losing candidate pair  $j$  is the difference between the initial margin and the audit margin:

$$e_{ij} = (w_{ij} - l_{ij}) - (w_{aj} - l_{aj})$$

and in all cases

$$(w_{ij} - l_{ij}) - (w_{aj} - l_{aj}) \leq b_i + w_{ij} - l_{ij}$$

because  $-(w_{aj} - l_{aj}) \leq b_i$

Therefore, authors agree (Dopp, 2007–2008b; Dopp, 2007–2008c; Aslam *et al.*, 2008; Stark,

---

<sup>16</sup>Dopp compares and contrasts the use of votes cast versus using the actual within audit unit upper margin error bounds to calculate audit sample sizes in (Dopp, 2007a).

2008c) that the maximum possible within audit unit initial margin error that could occur between any winning-losing candidate pair  $j$  for each audit unit  $1 \leq i \leq n$  is

$$u_{ij} = b_i + w_{ij} - l_{ij} \tag{1}$$

Note that, in the case of a multi-winner contest and individual ballot audit units, the expression for the upper margin error that any vote on the ballot could contribute to a winning-losing candidate pair ( $w_{ij}$  and  $l_{ij}$ ) reduces to:

$$\begin{aligned} b_i + w_{ij} - l_{ij} &= 2 \text{ if a vote is for winning candidate } j, \\ &= 1 \text{ if a vote is for another winning candidate,} \\ &= 0 \text{ if a vote is for losing candidate } j, \\ &= 1 \text{ if a vote is for another losing candidate,} \\ &= 1 \text{ if a vote is an over or under-vote} \end{aligned}$$

The upper margin error bound for that ballot is also zero (0) when all losing candidates receive votes on the ballot. (Calandrino *et al.*, 2007b, p. 7)

The audit unit upper margin error bound for winning-losing candidate pair  $j$  can be written as a percentage by dividing by the number of ballots cast, resulting in the expression

$$(1 + (w_j - l_j)/b_i)$$

For a simple example of the overall upper margin error, if the initial election results shows the initial winner has 100 votes and the initial runner-up has 0 votes with no under-votes or other candidates, but a full manual recount shows that the initial runner-up really had 100 votes and the initial winner had 0 votes, then the total margin error is  $200 = 100 + 100 - 0$  votes or 200%.

**Example** If the winner has 51% of the reported ballots cast and the runner-up has 48%, then the reported margin is 3%. For the reported winner to be incorrect there must be at least 3% margin error plus one vote. What is the minimum number of corrupt vote counts that could cause 3% or more margin error and thus result in an incorrectly reported election winner? The total possible percentage margin error in this example contest is 103% if all votes not counted for the runner-up should actually have been counted for the runner-up, so that the vote share of the runner-up should have been

Per Vote Miscount-Caused Margin Error for Candidate Pair A & B				
Vote initially reported & counted for	Vote actually cast for			
	candidate A	candidate B	candidate C	under or overvote
Initial winning candidate A	x	2	1	1
Initial losing candidate B	-2	x	-1	-1
Initial losing candidate C	-1	1	x	0
Initial under or over-vote	-1	1	0	x

Table 3: This table shows all possible error values that one miscounted vote could cause for the margin between a winning candidate A and a losing candidate B due to vote misallocation in a one-winner (one-seat) contest with three candidates A, B, and C. The margin error is 2 when a vote is initially counted incorrectly for the winning candidate A that should have been counted for candidate B; is 1 when a vote is initially reported for winning candidate A that should have gone to another losing candidate or should have been an under or over-vote; and is 1 when a vote is initially reported for another losing candidate or as an under or over-vote that should have been counted for candidate B.

100% with 0% for all other candidates. The upper margin error bound is thus found by taking the actual margin minus the reported margin between the winner and the runner-up or  $100\% - (-3\%) = 103\%$ .

### Upper margin error bounds for *all* winning-losing candidate pairs

Notice that one initial incorrectly recorded vote can contribute margin errors of  $-2, -1, 0, 1$  or 2 votes (See figure 3.1), so that the maximum margin error that any individual vote for a winning candidate can contribute is 2 votes. One initial incorrect vote for an initial losing candidate or an under or over-vote can contribute margin errors of  $-2, -1, 0$ , or  $+1$  votes, so that the maximum margin error that an initial vote for a losing candidate can contribute is 1 vote. Therefore the upper margin error bound for all winning-losing candidate pairs within each audit unit  $i$  is

$$u_i = 2 \sum_i w_i + \sum_i l_i \tag{2}$$

where  $l_i$  is the number of total votes for any losing candidates plus the number of total under or over-votes (cast ballots eligible to vote in the contest that have no vote counted

for any candidate) and  $w_i$  is the number of total votes for any winning candidate within audit unit  $i$ .

Note that, in the case of a single-winner contest, when using individual ballot audit units, the expression for the maximum margin error bound that could contribute to a losing candidate becoming the actual winner for each winning-losing candidate pair reduces to:

$$\begin{aligned} 2 \sum_i w_i + \sum_i l_i &= 2 \text{ if the vote is for the winning candidate,} \\ &= 1 \text{ if the vote is for a losing candidate,} \\ &= 1 \text{ if the vote is an over or under-vote} \end{aligned}$$

Note that an equivalent, perhaps simpler, expression for the upper margin error bound for *all* winning-losing candidate pairs is simply

$$u_i = b_i + \sum_i w_i \tag{3}$$

where  $b_i$  is the number of total ballots cast in audit unit  $i$ . The formula is equivalent because the number of ballots includes votes for all winning and losing candidates and under and over-votes, and  $\sum_i w_i$  doubles the number of votes for the winning candidates.

Calandrino, Halderman, and Felten (Calandrino *et al.*, 2007b) point out that there is an exception to the above rule. There must be at least one losing candidate not on the ballot in order for a ballot to contribute any margin error that could cause an incorrect election outcome. Therefore the upper margin error bound is zero for a ballot in the case that all losing candidates have votes on a particular ballot because such a ballot can only contribute negative margin error towards causing an incorrect outcome.

### **Use the just-winning and just-losing candidate pair to calculate sample sizes**

Using accurate sampling weights, any sample size that is sufficient to detect an incorrect election outcome between the just-winning and just-losing candidate pair will have at least the same minimum probability for detecting incorrect election outcomes that may have reversed other winning-losing candidate pairs.<sup>17</sup> Therefore post-election audit sample size calculations only need consider the just-winning and just-losing candidate pair error bounds and margin.

The reason that the most conservative (largest) overall post-election audit sample size is calculated by using the winning-losing candidate pair margin error bounds is because

---

<sup>17</sup>Thanks to Calandrino, Halderman, and Felten for coining the phrases “just-winning” and “just-losing” candidates.

Example Upper Margin Error Bounds in a Wide Margin Contest						
2004 Utah State Rep, Dist 3 Vote Counts					Upper Margin Error Bnds	
Precinct	#Ballots	Buttars	Hurtson	Elwell	Just-winning/just-losing candidate margin error bnds	All candidate pair margin error bnds
Totals	13,495	9,614	2,930	236	20,179	23,109
SMI1	871	638	185	11	1,324	1,509
HYD2	996	599	340	16	1,255	1,595
SMI2	819	599	165	15	1,253	1,418
NL04	864	595	222	20	1,237	1,459
NL03	754	553	144	17	1,163	1,307
SMI5	736	543	152	14	1,127	1,279
NLO1	787	512	208	17	1,091	1,299
SMI4	700	510	144	10	1,066	1,210
LO17	699	488	132	13	1,055	1,187
NL02	633	465	108	19	990	1,098
LO25	600	439	93	10	946	1,039
LO04	582	389	138	11	833	971
RCH1	507	390	89	13	808	897
RCH2	503	389	87	9	805	892
LO31	533	368	100	5	801	901
LO30	504	372	84	11	792	876
HYD1	594	381	182	9	793	975
LEW1	490	363	110	4	743	853
SMI3	475	358	94	7	739	833
LEW2	282	227	48	1	461	509
TREN	241	182	48	2	375	423
COVE	202	153	38	2	317	355
CORN	123	101	19	0	205	224

Table 4: The election data and margin error bounds shown above are used to demonstrate and compare the four methods for calculating risk-limiting post-election audit sample sizes that are discussed in this paper. The upper margin error bounds for the just-winning-losing candidate pair and the upper margin error bounds for *all* candidate pairs are shown here for the Utah State Representative District # 3 wide-margin contest in the 2004 general election. Another example of applying all four methods to determine risk-limiting sample size for a close-margin election contest is shown in an appendix.

that approach produces the smallest ratio of overall vote margin to the sum of within audit unit margin error bounds

$$\frac{M}{E}$$

where  $E_j = \sum_i u_i$  is the sum of all upper margin error bounds for winning-losing candidate pair  $j$  over all audit units  $i$  and  $M = w - r$  is the overall margin. In other words, the winning-losing candidate pair with the smallest ratio ( $M/E$ ) is always the just-winning and just-losing candidate pair.

$$\frac{M}{E} = \frac{w - r}{\sum_i b_i + w_i - r_i}$$

where  $w$  is the number of total votes for the winner with the least number of votes and  $r$  is the total number of votes for the losing candidate with the most number of votes (the runner-up).

A proof of this fact is provided in Appendix E.<sup>18</sup>

### **When weighting random selections or analyzing discrepancies consider *all* winning-losing candidate pairs**

When weighting random selections of audit units or when analyzing discrepancies found during an audit, focus should probably not be limited to particular initial winning-losing candidate pairs.

The only type of error that can overcome the winning margin of any winning-losing candidate pair is margin error that affects that particular candidate pair. The overall total margin error found during the audit within all audit units separately for each winning-losing pair, and the maximum within audit unit margin error for all audit units separately for each winning-losing candidate pair may need consideration when analyzing discrepancies. Discrepancy error for each winning-losing pair may need to be considered, but should be considered separately for each winning-losing pair over all audit units when analyzing discrepancies found during an audit.

Table 3.1 shows the election results data and the spreadsheet calculation of within audit unit upper margin error bounds for a specific election contest, the Utah State Representative District # 3 wide-margin contest in the 2004 general election. Both the upper margin error bounds for the just-winning/just-losing candidate pair and for all winning-losing candidate pairs are shown.

---

<sup>18</sup>This finding is contrary to the recommendations of some authors who recommend comparing within audit unit margin error bounds for all winning-losing candidate pairs at once for each audit unit when calculating sample sizes and doing discrepancy analysis (Stark, 2008a).

## 4 Uniform sampling methods

Given the minimum number of miscounted audit units  $C$  that could cause an incorrect outcome, then we can calculate the minimum sample size that is required to give a desired probability of having at least one such miscounted audit unit by solving the probability equation for  $S$ :

$$P = 1 - \frac{\binom{C}{0} \binom{N-C}{S}}{\binom{N}{S}}$$

Therefore, the first step to determine sample sizes for risk-limiting audits that employ uniform random sampling is to determine the smallest number of miscounted audit units  $C$ , that could cause an incorrect initial election outcome. Then, assuming the presence of at least  $C$  miscounted audit units, the sample is sized to have at least the desired minimum probability (say 95%+) of including at least one of these  $C$  miscounted audit units.

The larger the minimum number of vote counts  $C$  that it would take to cause an incorrect election, the smaller the sample size  $S$  is. Thus if  $C$  is overestimated, then the audit sample size may be too small to achieve the desired probability  $P$  for detecting incorrect election outcomes.

### **Calculate $C$ , the number of miscounted audit units that could cause an incorrect outcome**

Calculations that employ the detailed individual initial vote counts and ballots cast within each audit unit provide the most accurate sample size by providing the most accurate estimate for  $C$ , the minimum number of miscounted audit units that could cause an incorrect outcome. Using the detailed data takes into account the variation upper margin error bounds that exist within different audit units.

The most accurate estimate for estimating the minimum number of miscounted initial audit units,  $C$ , that could cause an incorrect election outcome is obtained by ordering all initial audit units in descending order of their upper margin error bounds and then, beginning with the audit unit with the largest error bound for the just-winning and just-losing candidate pair, add a fixed proportion (the maximum level of undetectability) times their upper margin error bounds until there is sufficient error to negate the entire margin between the just-winning and the just-losing candidate pair. The value of  $C$  is the number of audit units it takes to reach this level of margin error.

A program or a spreadsheet can do the calculations as follows (in pseudo-code):

- Create an array `error_bound(i) = bi + wi - ri` for each auditable vote count  $i = 0$  to  $N - 1$
- Sort the `error_bound()` array in descending order (from largest to smallest)
- Find the cumulative maximum possible margin error for the vote count with the largest error bound, then the two vote counts with the two largest error bounds, etc. For  $j = 0$  to  $N - 1$ ,
 
$$\text{CumulativeError}(j) = \text{error\_bound}(0),$$

$$\sum_{i=0}^1 \text{error\_bound}(i),$$

$$\sum_{i=0}^2 \text{error\_bound}(i)$$
- At each step compare the cumulative maximum error to the Margin and if  $M < k \times \text{Cumulative\_Error}(j)$  then  $C \approx (j + 1)$  is the minimum number of corrupt vote counts that could cause an incorrect election outcome and this value of  $C$  is used to use to calculate the audit sample size,  $S$ .

This calculation can be performed in a spreadsheet by listing the number of ballots  $b$  and vote counts for the just-winning  $w$  and just-losing  $r$  candidate pair within each audit unit, then calculating

$b_i + w_i - r_i$  for  $i = 0$  to  $N - 1$  and ordering the results from largest to smallest.

This method should be used to calculate the sample size prior to making the random selections.

Once  $C$  is known, we can use our desired probability  $P$ , and the number of total audit units  $N$ , we can calculate  $S$  the sample size by using an accurate algebraic estimate derived by Aslam, Popa, and Rivest (Aslam *et al.*, 2007)

$$S = (N - (C - 1)/2) \left( 1 - (1 - P)^{1/C} \right)$$

or we can find  $S$  exactly by using a numerical algorithm as shown by Dopp and Stenger (Dopp & Stenger, 2006).

See Table 7 for a summary of the uniform method of calculating post-election audit sample sizes.

### **Estimate $C$ the number of miscounted audit units that could cause an incorrect outcome**

Alternatively, rough estimates for sample sizes may be obtained for planning purposes such as for estimating the funds to allocate or the number of vote count auditors that should be

hired, etc.

Estimating the sample size by using the overall initial results, rather than using the initial detailed vote count (audit unit) data, in effect assumes that all vote counts contain a uniform amount of margin error. This assumption could over or under-estimate  $C$  and thus over or under-estimate the sample size needed to detect an incorrect initial outcome.

Example of the Improved Uniform Method & New Uniform Estimation Method			
Sample Size	Uniform Sampling Method		Uniform Estimation Method
	4		$5 \approx S \leq 8$
min #corrupt AUs	17		$12 \leq C \approx 20$
Precinct	$ku_i$	cumulative margin error	
SMI1	529.6	529.6	
HYD2	502.0	1,031.6	
SMI2	501.2	1,532.8	
NL04	494.8	2,027.6	
NL03	465.2	2,492	
SMI5	450.8	2943.6	
NLO1	436.4	3,380.0	
SMI4	426.4	3,806.4	
LO17	422.0	4,228.4	
NL02	396.0	4,624.4	
LO25	378.4	5,002.8	
LO04	333.2	5,336.0	
RCH1	323.2	5,659.2	
RCH2	322.0	5,981.2	
LO31	320.4	6,301.6	
LO30	316.8	6,618.4	
HYD1	317.2	6,935.6	
LEW1	297.2		
SMI3	295.6		
LEW2	184.4		
TREN	150.0		
COVE	126.8		
CORN	82.0		

Table 5: This table shows the steps for using the improved and new uniform sampling methods for calculating risk-limiting post-election audit sample sizes as applied to the 2004 Utah State Representative District # 3 wide-margin contest (Data shown in Figure 3.1). The confidence probability is  $P = 0.99$  and the maximum level of undetectability is  $k = 0.4$ . The prior unimproved uniform method using the old  $2sv_i$  error bounds calculated a sample size of only one (1) audit unit for this same contest because it under-estimates the maximum margin error that could occur within each audit unit.

Table 7 summarizes three methods and mechanisms for calculating risk-limiting election audit sample sizes.

Hence to obtain a rough estimate for the minimum number of audit units that could cause an incorrect outcome,  $C$ , using the overall total results, a more conservative formula

for estimating both  $C$  and  $S$  is suggested here.

When detailed data or tools and expertise are unavailable that are needed to obtain more precise estimates, these quick estimates can be made using the overall vote totals and the number of ballots cast in the largest audit unit in the contest (this can be estimated from a prior election if it is unavailable):

$$C_{avg} = \frac{(w - r)N}{k(b + w - r)} \quad (4)$$

where  $C_{avg}$  is the minimum number of corrupt audit units that could cause an incorrect election outcome if all audit units had the average upper margin error bound as all audit units, and

$$C_0 = \frac{b}{Nb_0} C_{avg} = \frac{b(w - r)}{kb_0(b + w - r)} \quad (5)$$

where  $C_0$  is the minimum number of corrupt audit units that could cause an incorrect outcome if all audit units have the maximum upper margin error bound estimated using  $b_0$ , the number of ballots cast in the largest audit unit in the contest, where  $b$  is the number of total ballots cast in the contest and  $N$  is the total number of audit units in the contest. See Appendix A for the derivation of equation 4. Equation 5 can simply be derived by estimating the upper margin error bound of the largest precinct using the number of ballots cast, assuming its margin is the same as the overall margin, and then solving for the ratio of  $C_{avg} : C_0$  which always nicely reduces to  $\frac{Nb_0}{b}$

This gives us the relationship  $C_0 \leq C \approx C_{avg}$ . If  $C_0 \leq 1$  then  $C = 1$ . If  $C_0 > 1$  then some further analysis may have to be done to estimate whether  $C$  will be closer to the value of  $C_0$  or  $C_{avg}$ .

Using a value of  $k = 0.40$  for estimating the sample size needed to obtain probability at least  $P$  for detecting  $C$  corrupt objects out of  $N$  objects and then using an estimation formula for  $S$  suggested by Aslam, Popa, and Rivest that sometimes slightly over-estimates the sample size needed to detect at least one miscounted audit unit if there are  $C$  miscounted units,

$$S \approx N \left( 1 - (1 - P)^{1/C_{avg}} \right) \quad (6)$$

$$S_{C_{avg}} \approx S \leq S_{C_0} \quad (7)$$

Substituting the expression for  $C_{avg}$  into the formula for  $S$ , and combining we get

$$S \approx S_{C_{avg}} = N \left( 1 - (1 - P)^{k(b+w-r)/(N(w-r))} \right) \quad (8)$$

and substituting the expression for  $C_0$  into the formula for  $S$ , and combining we get

$$S \leq S_{C_0} = N \left( 1 - (1 - P)^{kb_0(b+w-r)/(b(w-r))} \right) \quad (9)$$

See Appendix A for the derivation of Equation 8 and Appendix B for the derivation of Equation 6.

Other methods to estimate risk-based post-election audit sample sizes from overall initial election results can be developed using the pattern of upper margin error bounds found in prior elections' audit units. Such methods are specific to patterns found in specific election jurisdictions.

Using Formula 8 and 9 to estimate post-election audit sample sizes is simply done by using the following steps

1. Select a desired probability,  $P$ , for detecting the minimum level of miscount that could cause an incorrect election outcome. (A value  $P \geq 0.95$  is suggested).
2. Select an assumed maximum rate of undetectability that it is believed would not be immediately noticed if it occurs. (An initial value  $k \geq 0.40$  is suggested.)

3. Estimate the minimum number of miscounted audit units,  $C_{avg}$ , that could cause an incorrect election outcome

$$C_{avg} = (N * (w - r)) / ((k * (b + w - r)))$$

4. Calculate  $S \approx N \left( 1 - (1 - P)^{1/C_{avg}} \right)$  using a spreadsheet or a calculator.

It only takes one row and five columns in a spreadsheet to estimate the risk-limiting post-election audit sample size for various election contests using this new formula.<sup>19</sup>

## 5 Weighted sampling methods

When random selections are weighted by margin error bounds, the probability proportional to margin error bound (PPMEB) method can be used to determine the risk-limiting election

---

<sup>19</sup>A spreadsheet formula for calculating  $C$ , the number of corrupt counts that could cause an incorrect election outcome is

‘‘=(N\*(w-r))/(k\*(b+w-r))’’. A spreadsheet formula for calculating  $S$ , the audit sample size needed to provide  $P$ , probability for detecting one or more corrupt vote counts if there are  $C$  corrupt vote counts is ‘‘=N\*(1-(1-P)^(1/C))’’.

audit sample size (Aslam *et al.*, 2008; Dopp & Straight, 2006–2008; Dopp, 2007–2008a; Calandrino *et al.*, 2007b; Dopp, 2007–2008c).

Probability proportional to margin error bound methods for post-election auditing reduce the post-election audit sample size necessary to achieve a desired detection probability by weighting the selections of audit units using the audit unit margin error bounds (Aslam *et al.*, 2008, p.10 and 16), (Dopp, 2007–2008a; Dopp, 2007–2008c; Calandrino *et al.*, 2007b).

This article proposes new weights and probabilities for randomly selecting audit units that will work well for both individual ballot audit units proposed by Calandrino, Halderman, and Felten and for variable sized audit units proposed by Aslam, Popa, and Rivest (Aslam *et al.*, 2008; Calandrino *et al.*, 2007b).

In calculating the random sampling weights and probabilities, it is probably best to avoid making assumptions about which initial winning-losing candidate pair could be incorrect. Thus the upper margin error bounds that each ballot could contribute to margin error for *any* winning-losing candidate pairs should be considered developing sampling weights.

To ensure the most conservative approach (to ensure an adequate sample size), the margin in total ballots to be overcome is still the margin between the initial just-winning and the just-losing candidate pair.

This method applies equally well to any size audit unit and to election contests with any number of seats being elected. The sample size is calculated from the probabilities that each audit unit is randomly selected.

Appendix D describes some flaws with the upper margin error bounds and the sampling weights proposed in some other authors' recommendations.

## Calculate the random selection weights

Note that each initial losing candidate  $j$  would certainly prefer that the sampling weights used to select audit units are the losing candidate's own upper margin error bounds with an initial winning candidate as follows:

$$u_i = b_i + w_{ij} - l_{ij}$$

because then initial votes cast for that particular losing candidate would tend to escape scrutiny, while votes of other candidates would receive more scrutiny.

However, in order to provide equal treatment to all losing candidates when weighting random selections, we avoid using upper margin error bounds for particular winning-losing

candidate pairs and thus avoid any assumption about which initial winning-losing candidate pair(s) are incorrectly reported.

If a perpetrator with good insider access knows that the just-winning and just-losing upper margin error bounds are used to weight selections, that could provide a strategy to hide miscount by repositioning the relative order of candidate vote totals. In other words, a perpetrator could try to cause the initial results to show a different ordering of the candidates to reduce the chances of scrutiny of certain candidates' votes.

Hence we need a within audit unit upper margin error bound for *all* winning-losing candidate pairs for weighting random samples.

Not that there may be exceptions to this rule when auditors are increasing an audit sample in response to detecting errors in certain candidates' votes and believe that certain candidates' votes need the most scrutiny.

The selection weights for audit units suggested here allow for the possibility that any initial losing candidate might be a rightful winner, and any initial winning candidate could be a rightful losing candidate by using the upper margin error bounds for each audit unit  $i$  of:

$$u_i = 2 \sum_i w_i + \sum_i l_i \tag{10}$$

where

$l_i$  = the number of votes for any losing candidates

& under-over-votes,

$w_i$  = the number of votes for any winning candidate,

$w - r$  = the overall initial margin of just-winning and  
just-losing candidate pair.

See Appendix F for an example showing how this maximum amount of margin error could occur within an audit unit.

Another benefit of this method is that if there is more than one winner in a multi-seat election, the sampling weights suggested here do not assume which winner may be an incorrect winner, and do not assume which initial losing candidate may be a rightful winning candidate.

This weighted sampling proposal is consistent with the amount of upper margin error that each ballot can contribute to causing an incorrect outcome, given that we do not know

which initial winning-losing candidate pair may be incorrect before auditing.

## 5.1 The Improved PPMEB Method

Probability proportional to margin error bound (PPMEB) methods use margin error bounds for sampling weights.

One PPMEB method for sampling individual ballots looks at how many of each particular type of audit unit with its own particular voting patterns could be used to cause an incorrect election outcome (Calandrino *et al.*, 2007b). This section suggests using new more accurate sampling weights for the approach first developed by Calandrino et al. and generalizes the method to any size audit unit.<sup>20</sup>

Using this improved method, the fewest number of such miscounted audit units that could possibly cause an incorrect election outcome for each type of audit unit  $1 \leq i \leq n$  is:

$$c_i \approx \frac{(w - r)}{k(2 \sum_i w + \sum_i l)} \quad (11)$$

$c_i$  is an estimate for the number of similar audit units it would take to alter the election outcome by taking the maximum reasonable proportion of the over margin between the just-winning and just-losing candidate pair that could be eaten up by miscounted ballots with this vote pattern without raising immediate suspicion.

Therefore, the probability that each audit unit (individual ballot)  $1 \leq i \leq n$  should be selected for auditing is:

$$p_i = 1 - (1 - P)^{1/c_i} \text{ and substituting for } 1/c_i \text{ we get} \quad (12)$$

$$p_i = 1 - (1 - P)^{k(2 \sum_i w_i + \sum_i l_i) / ((w - r))} \quad (13)$$

where within each audit unit,

$l_i$  = the number of votes for losing candidates

& under & over-votes,

$w_i$  = the number of votes for any winning candidate,

$w - r$  = the overall initial margin of just-winning,

just-losing candidate pair, and

$k$  = the assumed maximum level of undetectability.

---

<sup>20</sup>Appendix D explains why the original method proposed by Calandrino et al. for random selection weights do not work as well as the improved method presented in this paper.

## The sample size

The expected value for the overall post-election audit sample size is equal to the sum of the probabilities that each audit unit is selected for auditing:

$$\sum_i p_i = \sum_i (1 - (1 - P)^{k(2\sum_i w_i + \sum_i l_i)/(w-r)}) \quad (14)$$

## 5.2 The Improved PPMEBWR Method

Another PPMEB approach looks at how much each audit unit could contribute overall to causing an incorrect election outcome. The probability proportional to margin error bound with replacement (PPMEBWR) approach was first developed by Aslam, Popa, and Rivest. (Aslam *et al.*, 2008, p. 16)

This section improves upon the method recommended by Aslam et al. by using the more precise just-winning/just-losing candidate pair upper margin error bounds to calculate the number of random draws and by using the new all-candidate-pair sampling weights presented here to calculate the sampling weights for the PPMEBWR method.<sup>21</sup>

Overall the PPMEBWR method is simply described as follows:

The number of random draws  $t$  with replacement of audit units is

$$t = \ln(1 - P) / \ln(1 - M/E_{wr})$$

where  $M$  is the overall vote margin for the just-winning and just-losing candidate pair, and  $E_{wr} = k \sum u_i$ , the maximum level of undetectability  $k$  times the sum of the within audit unit upper margin error bounds for the just-winning and just-losing candidate pair. In other words,  $E_{wr} = k \sum_i (b_i + w_i - r_i)$  where  $b_i$  is the total number of ballots cast,  $w_i$  is the number of initial votes counted for the just-winning candidate, and  $r_i$  is the number of initial votes reported for the just-losing candidate in audit unit  $i$ , and where  $P$  is the desired probability that there is at least one miscounted audit unit in the sample if the initial reported election outcome is incorrect.  $E_{wr}$  simply reduces to  $E_{wr} = k(b + w - r)$  or  $k$  times the total number of ballots + total votes for the just-winning candidate minus the total votes for the just-losing candidate in the contest because  $k \sum_i (b_i + w_i - r_i) = k(\sum b_i + \sum w_i - \sum r_i)$ .

So, as not to bias the sample in favor of a particular initial reported losing candidate

---

<sup>21</sup>Appendix D explains why the original Aslam et al. upper margin error bounds and random selection weights do not obtain the stated probability for detecting incorrect election outcomes.

over another, the probability  $p_i$  for sampling each audit unit  $i = 1$  to  $i = N$  is

$$p_i = \frac{2 \sum_{ij} w_{ij} + \sum_{ij} l_{ij}}{E_a}$$

$$p_i = \frac{b_i + \sum_{ij} w_{ij}}{E_a}$$

where the sum of margin error bounds is  $E_a = \sum_i (2 \sum_{ij} w_{ij} + \sum_{ij} l_{ij})$  the sum of the upper margin error bounds for *all* winning-losing candidate pairs, and where the number of votes for winning candidates  $j$  in audit unit  $i$  is  $w_{ij}$  and the number of votes for losing candidates  $j$  and under-votes in audit unit  $i$  is  $l_{ij}$ .

### First step: Sum the total error bounds

The first step is to calculate  $E_{wr} = k \sum_i (b_i + w_i - r_i)$  the sum of the error bounds for the  $i = 1$  to  $i = N$  audit units for the just-winning and just-losing candidate pair. We multiply the upper error bounds times  $k$ , the maximum level of undetectability where  $0 \leq k \leq 1$  as discussed above because not all ballots can be miscounted or it would be immediately evident without an audit. This sum can be easily calculated by adding the total number of ballots cast in the election contest, plus the total number of initial votes counted for the just-winning candidate, minus the total number of initial votes counted for the just-losing candidate.

Also calculate  $E_a = \sum_i (2 \sum_{ij} w_{ij} + \sum_{ij} l_{ij})$  the sum of the upper margin error bounds for *all* winning-losing candidate pairs for the  $i = 1$  to  $i = N$  audit units by summing two times the total initial votes counted for any winning candidate plus the total initial votes counted for any losing candidate.

### Second step: Calculate the number of draws

Calculate  $t$  using the margin in ballots between the just-winning and just-losing candidate pair,  $M = w - r$ , and  $E$ , and  $P$  the desired confidence probability that there will be at least one miscounted audit unit in our sample if the initial election outcome is incorrect.

$$t = \ln(1 - P) / \ln(1 - M/E_{wr}) \tag{15}$$

The derivation for the number of draws for the weighted sampling method is described in previous literature.(Aslam *et al.*, 2008, p. 10), (Dopp, 2007–2008a).

### Third step: Calculate the selection weights

Now calculate the sampling weights for each of the  $i = 1$  to  $i = N$  audit units.

$$p_i = \frac{2 \sum_{ij} w_{ij} + \sum_{ij} l_{ij}}{E_a} \text{ or } p_i = \frac{b_i + \sum_{ij} w_{ij}}{E_a} \text{ or}$$

using the upper margin error bounds for *all* candidates in audit unit  $i$  as shown in equation 10.

The expected value for the sample size  $S$  will be:

$$S = \sum_i (1 - (1 - p_i)^t)$$

## 6 Other Sample Size Considerations

### Losing candidates select additional audit units

To achieve the desired minimum probability for detecting incorrect election outcomes, due to the use of the assumption of a maximum level of undetectability,  $k$ , losing candidates or their representatives must be allowed to select at least one additional discretionary audit unit to supplement the randomly-selected sample. Discretionary audit units are necessary due to basing random sample sizes on the assumption that suspicious-looking audit units with more than a fixed level of margin error  $k$ , say 40% or 50%, would be investigated without the necessity for an audit (Dopp & Stenger, 2006, p. 14), (Dopp & Straight, 2006–2008), (Dopp, 2007–2008b), (Stark, 2008d, p. 7), (Appel, 1997; Hall, 2008b).

This crucial procedure would be thwarted if the losing candidates were required to pay the costs of auditing any discretionary units or if these discretionary audits were separately administered. Risk-limiting election audits that do not allow for auditing additional discretionary audit units over-state the confidence probability for detecting incorrect outcomes because they put no upper limit on the margin error that could occur within audit units, thus negating the assumptions used to calculate sample sizes.

Discretionary audit units should be included during the initial manual audit without cost to candidates, or a risk-limiting audit could fail to achieve its stated minimum probability to detect erroneous initial outcomes.

### Select one additional audit unit from each “missed” jurisdiction

Unless at least one audit unit is sampled from each separately administered election jurisdiction where an election contest occurs, innocent ballot programming errors, voting system

Example applying the Improved Weighted Sampling Methods				
	Improved PPMEB approach		Improved PPMEBWR approach	
Expected Sample Size	6		3	
			#Draws	3
Precinct	$c_i$	$p_i$	$p_i$	$1 - (1 - p)^t$
SMI1	11.07	0.34	0.07	0.18
HYD2	10.48	0.36	0.07	0.19
SMI2	11.78	0.32	0.06	0.17
NL04	11.45	0.33	0.06	0.18
NL03	12.79	0.30	0.06	0.16
SMI5	13.06	0.30	0.06	0.16
NLO1	12.86	0.30	0.06	0.16
SMI4	13.81	0.28	0.05	0.15
LO17	14.08	0.28	0.05	0.15
NL02	15.22	0.26	0.05	0.14
LO25	16.08	0.25	0.04	0.13
LO04	17.21	0.23	0.04	0.12
RCH1	18.63	0.22	0.04	0.11
RCH2	18.73	0.22	0.04	0.11
LO31	18.55	0.22	0.04	0.11
LO30	19.08	0.21	0.04	0.11
HYD1	17.14	0.24	0.04	0.12
LEW1	19.59	0.21	0.04	0.11
SMI3	20.06	0.21	0.04	0.10
LEW2	32.83	0.13	0.02	0.06
TREN	39.50	0.11	0.02	0.05
COVE	47.07	0.09	0.01	0.05
CORN	74.6	0.06	0.0	0.03

Table 6: This table shows the steps for using the improved weighted sampling methods for risk-limiting post-election audits applied to the 2004 Utah State Representative District # 3 wide-margin contest (data shown in Table 3.1). This example uses a confidence probability of  $P = 0.99$  and a maximum level of undetectability of  $k = 0.4$ . In this case, the improved PPMEBWR method shows an expected sample size of 3 audit units versus the old PPMEBR method that used the  $2sv$  error bound that would have calculated a sample size of only 1 audit unit. The improved PPMEB method is more conservative than the improved PPMEBWR method with an expected sample size of 6 audit units.

Table 7: Summary of Improved Methods for Risk-Reducing Post-Election Audit Sampling

Data & Tools Available?	Random Selection Method	
	Uniform Probability Distribution	Probability Proportional to Margin Error Bound (PP-MEB)
Detailed initial audit unit (vote count) data and a spreadsheet or computer program are available	<p>For each vote count <math>i</math></p> $\text{error\_bound}_{wr}(i) = k(w_i - r_i + b_i)$ <p>where <math>0.4 \leq k \leq 1</math> for the just-winning/just-losing candidate pair. Order <math>\text{error\_bound}</math> in descending size order. Then for each <math>j</math> from 0 to <math>N - 1</math>, if <math>w - r &lt; \sum_{i=0}^j \text{error\_bound}(i)</math> then stop and <math>C = j + i</math>. Then the sample size <math>S</math> can be found for detecting <math>C</math> corrupt audit units using a precise numerical method (Dopp &amp; Stenger, 2006) or by using an estimate (Aslam <i>et al.</i>, 2007)</p> $S = (N - (C - 1)/2) \left(1 - (1 - P)^{1/C}\right)$	<p><i>PPMEB METHOD</i></p> <p>If miscounted, each audit unit <math>i</math> with a particular vote pattern would take approximately at most <math>c_i \approx \frac{(w-r)}{k(b_i + \sum_i w)}</math> of such units to overcome the closest contest margin. So select each audit unit with probability</p> $p_i = 1 - (1 - P)^{1/c_i}$ <p>The expected value for the sample size is</p> $S = \sum_i p_i$ <p><i>PPMEBWR METHOD</i></p> <p><math>E_{wr} = k(b + w - r)</math> the sum of <math>k</math> times the upper margin error bounds for the just-winning/just-losing candidate pair.</p> <p>For each audit unit <math>i</math></p> $\text{error\_bound}_a(i) = b_i + \sum_{ij} w_{ij}$ <p>where</p> $E_a = \sum_{i=0}^{N-1} \text{error\_bound}_a(i)$ <p>is the sum of bounds for <i>all</i> winning-losing candidate pairs. The selection probability for each audit unit is</p> $p_i = \frac{\text{error\_bound}_a(i)}{E_a}$ <p>The number of selection rounds is</p> $t = \ln(1 - P) / \ln(1 - M/E_{wr})$ <p>and the expected sample size is</p> $S = \sum_i (1 - (1 - p_i)^t)$
Overall, but not detailed, election results data are available	<p>Estimate the number of corrupt vote counts to cause an incorrect election outcome</p> $C_{avg} = ((w - r)N) / (k(b + w - r))$ $C_0 = \frac{b}{Nb_0} C_{avg}$ <p>then the sample size is between these two values</p> $N \left(1 - (1 - P)^{1/C_{avg}}\right) \leq S \leq N \left(1 - (1 - P)^{1/C_0}\right)$ <p>where <math>0.5 \leq k \leq 1</math>. See Appendices A &amp; B for more information.</p>	Not possible

problems, or fraud that is peculiar to one jurisdiction could be missed. It is important to make these additional random selections only *after* the initial random selections are made from any missed jurisdictions because otherwise audits would insufficiently sample high-population areas, thus providing a map to potential perpetrators for what areas to target in order to increase the chance that audits would not detect miscount.<sup>22</sup>

### **Size audit units as uniformly as possible**

It is important to keep the size variation of auditable vote counts as small as possible for two reasons:

1. Wide variation in the number of ballots cast (and the margin error available) in different audit units can result in sufficient margin error to cause an incorrect election outcome existing in just a few of the largest audit units. Especially if uniform sampling methods are used (as unfortunately is required by some state's auditing statutes) outcome-altering error may be missed and wide variation in audit unit sizes increases the need for manually counting more ballots.
2. Risk-limiting audits can be conducted more efficiently if the total number of ballots in all audit units is roughly uniform by evening out within audit unit margin error potential somewhat. To achieve the same risk level, it requires manually counting fewer ballots overall when the audit unit sizes are more uniform.<sup>23</sup>

### **Small-sized audit units are more efficient**

The smaller the number of total ballots within reported audit units, the fewer the number of ballots overall will need to be manually audited to provide a desired probability for detecting incorrect outcomes ((Wand, 2004; Walmsley, 2005; Atkeson *et al.*, 2008)). In other words, a larger number of smaller audit units are more efficient to audit.

Voting system design and procedural obstacles exist to conducting effective, efficient audits today. For example almost all commercial voting system tabulators are designed to only report precinct vote counts and not to report which machines tallied those votes and not to report tallies for ballots that are counted and stored together. These design flaws make precincts or polling places the only audit unit that election officials can conveniently

---

<sup>22</sup>Attorney Paul Lehto pointed this out in emails.

<sup>23</sup>With large size variation of vote counts the Uniform method for estimating sample sizes without detailed data will underestimate the sample size needed to achieve the desired probability of detecting incorrect outcomes, and thus under-states the maximum risk level for a calculated sample size.

use for auditing election results accuracy without taking time-consuming extra measures that delay the tallying and public release of initial election results just at the time when candidates and press are anxiously waiting for results (Dopp, 2009).

Election officials may wait to begin a post-election audit after all provisional and mail-in ballots have been counted and publicly reported, or alternatively officials can use somewhat inconvenient, time-consuming ways to count and to publicly report mail-in and provisional ballots in batches that are roughly equal in size to the median or average-sized audit units. See the upcoming article in this same series *Checking Election Outcome Accuracy — Post-Election Audit Procedures*.

### Some errors to avoid

Some authors state that risk-limiting post-election audits can be performed using any initial sample size (Stark, 2008c, p. 18), (Stark, 2009a; Hall *et al.*, 2009). However, audits that use insufficient initial sample sizes are either ineffective or inefficient — Ineffective because insufficient initial samples are not likely to detect well-hidden vote fraud in cases when a minimum number of audit units are miscounted to cause an incorrect outcome and yet avoid a state-mandated recount; and inefficient and administratively burdensome because even if no discrepancies are found in a too-small sample, limiting the risk requires manually auditing another round of randomly selected audit units.

Because keeping the overall winning margin  $w - r$  constant and decreasing the sum of the maximum level of undetectability times the upper margin error bounds,  $E$ , increases the quantity  $M/E$  and decreases calculated post-election audit sample sizes:

1. taking minimums of within audit unit upper margin error bounds out of all the winning-losing pairs produce an insufficient sample size,
2. taking minimums of values that are less than the margin error bounds for the just-winning and just-losing candidate pair when calculating the total upper margin error bound  $E$ , produce an insufficient sample size,
3. using proportions of total votes or of total ballots to approximate the total upper margin error bound  $E$  (such as  $\sum 2v_i$ ) that are less than the sum of actual margin error bounds produce insufficient sample sizes and poor sampling weights. (See Appendix D for a discussion.)

Methods that produce insufficient sample sizes will not achieve their stated minimum

probability for detecting incorrect outcomes, unless the audit sample is expanded even when no discrepancies are found in the initial sample.

Failure to use, or to understand the logical implications of using, a maximum level of undetectability of less than one (1) may cause:

- a test result that says to expand an audit sample unnecessarily, in some cases even after a 100% manual count has already been performed; and
- an unmerited expansion of the sample size when there is only a one ballot discrepancy. (Stark, 2009b, p. 6–9)

When calculating sample sizes, sampling weights, or analyzing discrepancies, methods that use a different winning-losing candidate pair’s margin error bound for each different audit unit or even mix up the data for different election contests into one calculation method (Stark, 2008c; Stark, 2008d; Stark, 2009a; Hall *et al.*, 2009) are less precise. Such methodology does not save significant computation time or resources because the data and resources used to do more precise calculations are about the same. The sum or maximum of within audit unit upper margin error bounds for each winning-losing candidate pair separately could just as easily and more accurately be calculated for one election contest at a time, and then that sum or maximum compared with those of other winning-losing candidate pairs for each separate election contest. The method of mixing up multiple election contests into one calculation will initially over-audit some contests and under-audit others, making the discrepancy analysis conclusions less efficient for some contests and less precise, possibly causing unnecessary false positives or failures to detect some incorrect election outcomes. Also trying to manually count multiple contests on the same ballot at the same time during an audit, is likely to negate the efficiencies of using the sort and stack method for counting paper ballots (Deputy Secretary of State Anthony Stevens, 2007).

Sampling methods that assume that vote miscount is uniformly distributed throughout the population of all audit units cannot reliably detect well-hidden vote fraud where the minimum number of audit units are miscounted that could cause an incorrect election outcome.

All election contests require a sample size larger than zero (0). In fact the formula given to show that in some cases a post-election audit is not required to confirm an outcome (Stark, 2008c, p. 10), (Stark, 2009b, p. 6) can be instead be used to prove that in any contest with more than one candidate, the sample size must be greater than zero to confirm the outcome because winning candidates have more initial reported votes than losing candidates.

## 7 Summary & Recommendations

Risk-limiting post-election audits limit the risk that an incorrect election outcome, the wrong winner, is incorrectly certified to any desired small maximum probability.<sup>24</sup> Table 7 summarizes the improved methods presented in this paper for calculating post-election audit sample sizes and for weighting random selections that ensure that post-election audits achieve the desired minimum probability of detecting and correcting any incorrect initial outcomes.

The new upper margin error bounds for all winning-losing candidate pairs that are presented in this article improve the sample size calculations and sampling weights of existing approaches and work for any size (1-ballot or many-ballot) audit units, for single or multi-winner election contests and for approaches treating all losing candidates equally. Using precisely correct upper margin error bounds ensures the adequacy of post-election audit sample sizes and allows random selection weights focusing on all winning-losing candidate pairs or on a particular winning-losing candidate pair.

When weighting random selections, this author recommends using the more conservative PPMEB method that uses a larger post-election audit sample size. Both the PPMEB and the PPMEBWR weighted sampling methods can be generalized for any sized audit units (from 1 ballot to many ballots) using the improved methods suggested in this paper.

Methods and materials for auditing elections and training auditors still need development. Election officials, vote count auditors, election integrity advocates, and voting system vendors, would benefit from

- Better voting system design specifications and technical features to make voting machines and tabulators more audit-able and accountable, provide more convenient methods to check vote count accuracy and to determine how, when, and where errors occur and the cause of errors.
- An accurate, understandable post-election auditing manual and an easy-to-use toolkit that includes a clear explanation of and a program for calculating risk-limiting election audit sample sizes, including procedures for conducting effective and efficient post-election audits. This manual should provide pictures, forms and toolkits; and

---

<sup>24</sup>Risk-limiting audits should limit risk of incorrect outcomes even in cases where there are a large number of under-votes. In the 2006 Sarasota County, Florida Congressional District 13 race, there were 18,000 missing votes (undervotes) in a Democratic-leaning county recorded on paperless ES&S DREs in a tight election. Statistics show that these undervotes probably altered the outcome, causing Christine Jennings to lose the US House race.

explain election auditing in simple-to-follow terms for lay persons<sup>25</sup>.

- Use of open publicly defined computer data recording format standards uniformly adopted by all election districts to provide consistent access to all electronic ballot records and making voting system components, including auditing devices, interoperable<sup>26</sup>.
- Conferences that bring together experts in election auditing methods together with State and county election officials and State and Federal legislators.
- Clear, easy-to-follow instructions and computer programs for making verifiable fair, weighted or uniform random selections of audit units.
- Methods to generate audit vote fraud and discrepancy test data to test the ability of various audit methods' to detect various vote fraud strategies.
- A complete set of methods, tools, and decision-making algorithms for analyzing post-election audit discrepancies, including algorithms for deciding whether to certify an election outcome or to expand the audit sample.<sup>27</sup>
- The development of precise methods to assist losing candidates in selecting discretionary “suspicious-looking” audit units to add to the randomly selected sample.
- A college textbook or textbook chapter to explain post-election auditing methodology and principles.

The knowledge, resources and skills needed to implement routine post-election risk-limiting audits that provide high confidence in the accuracy of final election outcomes need on-going development and dissemination.

There are two other articles in this series *Checking Election Outcome Accuracy — Post-Election Audit Discrepancy Analysis*, and *Checking Election Outcome Accuracy — Post-Election Audit Procedures*.

---

<sup>25</sup>In order to be successful, so that election jurisdictions do not have to hire statisticians to plan every post-election risk-limiting audit, such a project would require professional manual writers, and open source computer program developers, to create an easy-to-use manual and tool-kit in collaboration with election officials, security experts, and election auditing experts.

<sup>26</sup>To date only the Secretary of State office in California has reported precinct level election results using international recording standards.

<sup>27</sup>Most current discrepancy analysis methods expand post-election audit sample sizes whenever audit discrepancy exceeds an arbitrary fixed amount or rate of error and otherwise certify the election outcome. These approaches could incorrectly certify very close margin contests or unnecessarily expand audits in wide-margin contests. The third article in this series *Checking Election Outcome Accuracy — Post-Election Audit Discrepancy Analysis* covers algorithms and methods for analyzing post-election audit discrepancies.

## Acknowledgments

Thanks to David Webber of Open Voting Solutions<sup>28</sup> for reviewing this article and making helpful suggestions and to Political Science Professor Andrew Gellman for suggesting the addition of an introduction describing the work of other authors in the field of post-election auditing methods. Thanks to the contributions of other authors who formerly developed weighted sampling approaches for calculating risk-limiting post-election audit sample sizes or who derived formulas for estimating uniform sample sizes that this paper advances and improves upon. Profuse thanks to Nelson Beebe for taking many patient hours teaching me how to use LaTeX and to improve this article's formatting. Thanks to Mathematics and Computer Science Professor Frank Stenger for making a crucial stylistic suggestion and to Mathematics Professor Stewart Ethier for valuable formatting and LaTeX suggestions, and to other dedicated election integrity advocates who motivated me to keep working on and improving this paper.

---

<sup>28</sup><http://openvotingsolutions.com>

## Appendix A: Derivation of Estimate for the Minimum Number of Corrupt Vote Counts to Alter an Outcome

The smallest amount of margin error that could change the reported outcome is the actual margin in ballots between the winning candidate with the smallest number of votes (the just-winning candidate) and the losing candidate with the most votes (the just-losing candidate).

### The upper bound for margin error

The upper bound in number of ballots for the total margin error that could contribute to an incorrect election outcome is given by the expression  $b + w - r$ , or as a percentage of ballots, is  $1 + \frac{w-r}{b}$ , where  $b$  is the total number of ballots cast in the election contest,  $w$  is the reported number of votes counted for the winner, and  $r$  is the number of reported votes counted for the closest runner-up.

### Estimating the minimum number of corrupt vote counts, $C$ that could cause an incorrect outcome

*C<sub>avg</sub>*, the estimate if margin error bounds were all average sized

One way to estimate the minimum number of corrupt audit units,  $C$ , that could cause an incorrect winner, is to divide the margin in ballots between the just-winning and just-losing candidates by the average upper margin error bound for all the reported vote counts. This gives a measure for how many corrupt audit units could cause an incorrect election outcome if all audit units have the average upper margin error bounds.

$$\frac{\text{TotalMarginError2ChangeOutcome}}{\text{AverageMarginErrorPerAuditUnit}} = \#AuditUnits2ChangeOutcome$$

This estimates the number of audit units with sufficient possible margin error to cause an incorrect election outcome.

Thus, this estimate for  $C_{avg}$  is

$$C_{avg} \approx \frac{w - r}{\frac{k(b+w-r)}{N}}$$

which reduces to

$$C_{avg} \approx \frac{N(w - r)}{k(b + w - r)}$$

However, this method usually underestimates  $C$  in real elections because the average amount of margin error in all audit units is always less than the amount of possible margin error

in the largest audit units. Hence using a larger  $k$  value such as  $k \geq 0.50$  will help to compensate by finding a more conservative estimate (larger sample size) for  $S$ .

### $C_0$ , the most (overly) conservative estimate

From the above formula for  $C_{avg}$  and some estimates for the reported margin in the largest audit unit, a formula can be derived and simplified that provides a more conservative (smaller) estimate for the number of corrupt audit units  $C_0$  and thus provides a more conservative (larger) sample size estimate.

Simply use the relationship that

$$C_0 \approx \frac{bC_{avg}}{Nb_0} \quad \text{where } b_0 \text{ is the number of ballots in the largest audit unit with the most ballots cast}$$

This estimate assumes that the margin in the largest audit unit is the same as the overall margin in the election contest. That is, the margin error bound in the largest audit unit is  $b_0(1 + (w - r)/b)$  so that the ratio of the margin error in the largest audit unit to the audit unit with average margin error algebraically reduces to  $Nb_0/b$ .

## Appendix B: Derivation of a Uniform Election Audit Sample Size Estimate

This derivation has been previously described elsewhere (Rivest, 2007; Dopp, 2007b). Two well-known mathematical facts are used in the derivation:

1. For values of  $0 < c < 1$ ,  $(1 - c) \approx e^{-c}$  so that  $(1 - c)^S \approx e^{-cS}$  and therefore taking the natural log of both sides  $\ln(1 - c)^S \approx -cS$ ;
2. The formula for estimating the number of distinct elements,  $S$ , in a sample of size  $t$  drawn (with replacement) from a set of size  $N$  is  $S \approx N(1 - e)^{-t/N}$ .

We begin with  $N$  total vote counts, of which  $C$  are corrupt (miscounted) and ask what randomly selected sample size  $S$  will give us at least probability  $P$  for having one or more corrupt vote counts. Estimating the desired probability from sampling with replacement (an easier equation to solve than sampling without replacement), we find the probability of not detecting any miscount. Because,  $P(0) + P(1) + \dots + P(S) = 1$  and  $S \leq C$ , then the probability of drawing one or more corrupt vote counts is

$$P(1) + P(2) + \dots + P(S) = 1 - P(0).$$

If we randomly draw one vote count, then the chance of drawing a corrupt vote count is  $C/N$  and the chance of not drawing a corrupt precinct,

$$P(0) = \frac{N - C}{N}$$

so that

$$P(0) = 1 - \frac{C}{N}.$$

If we sample with replacement (each draw is an independent event) then the probability of drawing no corrupt precincts in  $S$  draws is

$$\left(1 - \frac{C}{N}\right)^S$$

and thus the probability of drawing one or more corrupt vote counts is

$$P = 1 - \left(1 - \frac{C}{N}\right)^S.$$

If the rate of corrupt vote counts  $c$  is  $\frac{C}{N}$  where  $0 < c < 1$ , then the chance of selecting zero corrupt counts in  $S$  draws with replacement is  $(1 - c)^S$  for selecting zero corrupt vote counts. Therefore, the estimated probability for selecting one or more corrupt vote counts in  $S$  draws with replacement is  $P = 1 - (1 - c)^S$ .

Beginning with the probability for not detecting any miscount

$$1 - P \approx (1 - c)^S$$

taking the log of both sides

$$\ln(1 - P) \approx \ln(1 - c)^S$$

and

$$\ln(1 - P) \approx -cS$$

and solving for  $S$  gives

$$S \approx \frac{\ln(1 - P)}{-c}$$

To further improve the estimate for  $S$ , we use the formula for estimating the number of distinct elements,  $S$ , in a sample of size  $t$  drawn (with replacement) from a set of size  $N$ :  $S \approx N(1 - e)^{\frac{t}{N}}$  and replace  $t$  by our estimate for  $S$  above, resulting in

$$S \approx N \left(1 - e^{\frac{\ln(1 - P)}{Nc}}\right) \tag{16}$$

(Rivest, 2007; Dopp, 2007b).

Equivalently, because  $Nc = C$ , we get

$$\begin{aligned} S &\approx N \left( 1 - e^{-\frac{\ln(1-P)}{C}} \right) \\ &\approx N \left( 1 - (e^{\ln(1-P)})^{\frac{1}{C}} \right) \\ &\approx N \left( 1 - (1-P)^{\frac{1}{C}} \right), \end{aligned}$$

a simple formula for estimating post-election audit sample sizes to provide at least  $P$  probability for detecting one or more corrupt vote counts in a sample of size  $S$  if there are  $N$  total vote counts and  $C$  miscounted vote counts.

Then, as seen in Appendix A, we can estimate

$$\begin{aligned} C &\approx \frac{N(w-r)}{k(b+w-r)}, \\ S &\approx N \left( 1 - (1-P)^{\frac{k(b+w-r)}{N(w-r)}} \right). \end{aligned}$$

A slightly more exact numerical method for calculating the risk-limiting election audit sample size  $S$  is found by solving the sampling-without-replacement formula, by employing the detailed estimate for the minimum number of corrupt vote counts,  $C$ , using the gamma function for evaluating

$$\ln(1-P) - \ln\left[\frac{(N-C)!(N-S)!}{N!(N-C-S)!}\right],$$

$$\begin{aligned} y &= \ln(1-P) + \text{gamma}\ln(N-C-S+1) - \text{gamma}\ln(N-S+1) \\ &\quad + \text{gamma}\ln(N+1) - \text{gamma}\ln(N-C+1). \end{aligned}$$

via the numerical method of bisections (Dopp & Stenger, 2006).

## Appendix C: Double-Checking the Audit Sample Size

Any post-election audit sample size  $S$  can be checked to see what minimum probability the sample size provides for detecting the minimum amount of vote miscount necessary to cause an incorrect election outcome by using the formula for the probability for drawing one or more miscounted audit units in a randomly drawn sample of audit units of size  $S$ , drawn without replacement, when there are  $C$  corrupt vote counts out of a total of  $N$  vote counts.

$$P = 1 - \frac{\binom{C}{0} \binom{N-C}{S}}{\binom{N}{S}}$$

can be calculated in a spreadsheet using the formula  $1 - \text{HYPGEOMDIST}(0, S, C, N)$  (See (Dopp & Baiman, 2005–2006)) This formula for checking the probability that the audit sample would detect outcome-altering vote miscount may be applied to both fixed rate audits and to risk-limiting audits.

## Appendix D: Some Weighted Random Sampling & Error Bound Methods That Do Not Work Well

This appendix points out details of the work of authors who have contributed new approaches to post-election auditing mathematics and methods that need some improvements.

For instance, the sampling weights recommended for post-election auditing in *Machine-Assisted Election Auditing* seem to be incorrect (Calandrino *et al.*, 2007b, pp. 7–8).

Auditors desire a confidence level  $c$  that no fraud significant enough to change the election’s outcome occurred. First the authors’ define their variables:

“... let  $v_1, \dots, v_n$  be the electronically reported vote totals for the candidates in decreasing order. Therefore  $v_1, \dots, v_k$  correspond to winning candidates. Because a single ballot may contain votes for up to  $k$  candidates, we need to consider the combination of votes on each ballot. Given a ballot, let  $C_s$ , where  $1 \leq s \leq k$ , be the winning candidate with the lowest vote total that received a vote on the ballot. (Let  $C_s$  be null if the ballot does not contain votes for any winning candidate.) Let  $C_t$ , where  $k + 1 \leq t \leq n$ , be the losing candidate with the highest vote total that did not receive a vote on the ballot. (Let  $hC_t$  be null if the ballot contains votes for all of the losing candidates.)”

Calandrino et al. continue:

If  $C_s$  is non-null, then we need to audit this ballot with probability at least  $1 - (1 - c)^{1/b_1}$ , where  $b_1 = v_s - v_{k+1}$ . Intuitively, one possible result-changing scenario involving an error in this ballot would be to add  $v_s - v_{k+1}$  incorrect votes for candidate  $s$ .

The last sentence of the preceding paragraph is incorrect because the most number of votes that any single ballot can possibly contribute is  $k$  votes, where  $k$  is the number of seats being elected in the contest. Thus the most number of incorrect votes that any one ballot can contribute is  $k$  votes.

On the other hand, the difference between the initial reported number of votes for any winning and losing candidate could be in the thousands, a number much larger than the maximum number of incorrect votes  $k$  that any ballot can produce in the initial reports.

In addition, the formula above cannot be generalized to audit units with more than one ballot and does not seem to reliably produce sensible sampling probabilities.<sup>29</sup>

Similarly, the audit unit sampling weights recommended by Aslam, Popa, and Rivest in *On Auditing Elections When Precincts Have Different Sizes* (Aslam *et al.*, 2008, p. 16) described by the expression

$$e_i = \min(2sv_i; M; v_i + r_{ij_1} - \min_j r_{ij})$$

(It is OK just to use the first term, so that  $e_i = 2sv_i$ .) . . .

Also compute the total error bound:

$$E = \sum_{1 \leq i \leq n} e_i.$$

most often result in using the quantity  $2sv_i$  where  $s = 0.20$  and  $v_i$  is the number of votes cast within each audit unit. The quantity  $2sv_i$  is then used for both the sampling weights and to calculate the overall error  $E$ . Using the quantity  $2sv_i$  for weighting random selections is less desirable than weighting random selections by the number of ballots cast within each audit unit because  $2sv_i$  does not account for under and over-votes since it uses the quantity “votes cast” rather than “ballots cast”. Using  $2sv_i$  rather than the actual upper margin error bounds will often result in an inadequate sample size and puts too much focus on auditing ballots that contain votes for losing candidates where not as much margin error could contribute to causing an incorrect initial outcome.

Figure 7 below shows that the quantity  $2sv_i$  is most often less than the actual within audit unit upper margin error bound, thus under-estimating the maximum possible margin error, increasing the quantity  $M/E$  and producing an inadequate sample size. Figure 7 also shows how  $2sv$  is sometimes impossible because there are not enough votes for the just-winning candidate in some audit units to contribute 40% margin error. Note that the upper margin error bounds “ $b + w - r$ ” and “ $2w + r + \text{other}$ ” can be as high as 200% of the total number of within audit unit ballots while  $2sv$  is always 40% of the total number of ballots. Note that the actual upper margin error bounds increase as the winning candidate share increases, and that since the winning candidate normally has more within audit unit vote share than the losing candidate the  $2sv$  error bound under-estimates the possible margin error, thus over-estimating the number of audit units needed to cause an incorrect outcome

---

<sup>29</sup> $p_i = 1 - (1 - c)^{1/b_1}$ .

and under-estimating the post-election audit sample sizes necessary to detect incorrect election outcomes.

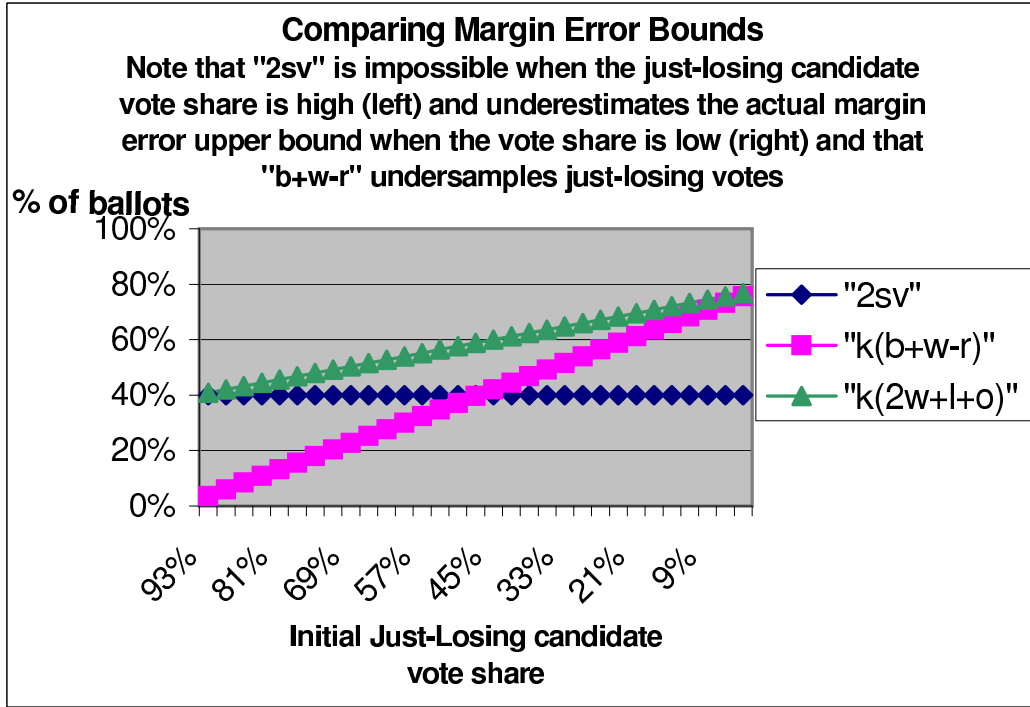


Figure 3: The  $2sv$  error bounds compared with actual upper margin error bounds for the “just winning-losing” candidate pair,  $k(b + w - r)$ , and with the upper margin error bounds for all “winning-losing” candidate pairs,  $k(2w + r + \text{other})$  where  $k = 2s$  — plotted by initial vote share of the just-losing candidate

Despite being an excellent upper error bound for calculating sample sizes, the just-winning-losing candidate pair bounds,  $b_i + w_i - r_i$ , may be incorrect sampling weights. As a sampling weight the just-winning-losing candidate pair bounds could increase the chance for certain election rigging strategies to prevail under some circumstances.

## Appendix E: The just-winning-losing candidate pair bounds produce the largest sample size

PROOF that using the winning-losing candidate pair with the smallest margin produces the most conservative (largest) post-election audit sample size:

Clearly  $(w - r) \leq (w - l)$  where  $l$  is the number of votes for any initial losing candidate. So take any losing candidate with a greater margin than the runner-up and his margin can

be expressed as  $w - r + y$  where  $y > 0$ . Now compare the ratio ( $M_r/E_r$ ) of the runner-up and any other initial losing candidate.

We want to compare

$$\frac{M_r}{E_r} = \frac{w - r}{b + w - r} \text{ and}$$

$$\frac{M_l}{E_l} = \frac{w - (r - y)}{b + w - (r - y)}$$

to see which is bigger.

For simplicity, let  $w - r = x$  so we compare

$$\frac{M_r}{E_r} = \frac{x}{b + x} \text{ and}$$

$$\frac{M_l}{E_l} = \frac{x + y}{b + x + y}$$

Multiplying both fractions top and bottom by the necessary factors to get a common denominator and comparing numerators we get

$$xb + x_2 + xy \leq xb + x_2 + xy + by$$

so that  $\frac{M_r}{E_r} \leq \frac{M_l}{E_l}$

Q.E.D.

## Appendix F: Example of maximum *all* candidate-pair margin error bound

If 200 ballots are cast in a precinct, and all four candidates — two apparent winners and two apparent losers — initially receive 50 votes apiece, the within audit unit upper margin error bound for *all* winning-losing pairs is  $2w + l = b + w = 2(100) + 100 = 300$  votes where  $w$  is the number of votes counted for initial winners and  $l$  is the number of all other ballots, and  $b$  is the number of total ballots cast.

This appendix shows one way out of the twelve possible ways that the maximum 300 vote margin error can occur in this particular situation.

One way that the maximum 300 vote margin error can occur is to assume that candidates A & B are the initially, but incorrectly, reported winners, and that candidates C & D are the initially incorrectly reported losers. Given this scenario there are two possible scenarios that result in a 300 vote margin error.

Eg. Miscount the initial votes as follows in this precinct.

1. The 50 votes for candidate A should really have been counted for candidate C, causing a 100 vote margin error for the pair A & C.
2. The 50 votes for candidate D should really have gone to candidate C, causing a 50 vote margin error for the pair A & C.
3. The 50 votes counted for candidate B should really have been counted for candidate D, causing a 100 vote margin error for the pair B & D.
4. The 50 votes counted for candidate C should really have been counted for candidate D, causing a 50 vote margin error for candidate pair B & D.

These errors add up to a 300 vote margin error for this scenario.

For each of the 6 possible pairs of winners and losers possible for this reviewer's scenario, there are 2 such examples of a 300 vote margin error, for 12 total examples of how a 300 vote margin error can occur.

## **Appendix G: Example of risk-limiting audit sampling methods in a close-margin election contest**

The tables below shows an example applying the margin error bounds and improved methods presented in this article to the election data for the close-margin 2004 Utah State Senate District # 1 election contest.

Example Upper Margin Error Bounds in a Narrow Margin Contest							
2004 Utah State Senate, Dist # 1 Vote Counts					Upper Margin Error Bnds		
Precinct	#Ballots	Fife	Evans	Jenkins		Just-winning/just-losing candidate margin error bnds	All pair candidate margin error bnds
Totals	16,987	7,981	7,553	463	338	16,859	24,412
SL2004	574	296	233	17	7	637	870
SL2052	518	262	191	12	5	589	780
SL1308	536	266	226	22	11	576	802
SL2034	572	269	275	8	4	566	841
SL2226	486	260	184	9	16	562	746
SL2056	518	252	212	23	9	558	770
SL2006	597	247	293	12	12	551	844
SL2030	545	254	260	8	13	539	799
SL2214	407	238	138	9	6	507	645
SL2008	453	228	186	16	10	495	681
SL2224	377	237	119	9	2	495	614
SL1216	499	225	232	13	12	492	724
SL2049	521	224	258	19	5	487	745
SL1214	562	220	300	12	8	482	782
SL2242	440	219	189	12	4	470	659
SL2036	517	208	272	9	10	453	725
SL1306	458	199	215	9	18	442	657
SL2206	333	204	102	9	13	435	537
SL2014	392	206	165	12	3	433	598
SL2252	346	200	116	9	5	430	546
SL1302	466	193	231	14	10	428	659
SL1327	455	190	227	14	10	418	645
SL2003	378	198	159	8	3	417	576
SL2222	383	192	159	7	10	416	575
SL2246	372	194	150	6	11	416	566
SL1218	374	177	147	12	21	404	551
SL1328	378	182	168	9	8	392	560
SL2038	290	170	86	14	1	374	460
SL2254	315	167	117	12	0	365	482
SL2007	454	162	267	7	8	349	616
SL2002	332	160	153	7	4	339	492
SL1320	411	148	229	13	7	330	559
SL2204	256	148	86	6	8	318	404
SL1322	321	137	142	15	11	316	458
SL1350	341	129	172	17	12	298	470
SL1346	347	133	190	8	8	290	480
SL1210	256	116	114	8	3	258	372
SL1303	343	105	199	7	9	249	448
SL2050	287	104	153	15	7	238	391
SL2053	196	108	67	6	3	237	304
SL1351	253	85	132	9	8	206	338
SL2001	73	42	25	0	1	90	115
SL2216	44	27	14	0	2	57	71

Table 8: The election data and margin error bounds for the 2004 Utah State Senate District # 1 narrow-margin contest shown above are used in Tables 7 and 7 to demonstrate the four improved methods for calculating risk-limiting post-election audit sample sizes presented in this paper.

Example # 2 of the Improved Uniform Method & New Uniform Estimation Method			
Sample Size	Uniform Sampling Method		Uniform Estimation Method
	40		—c—
min #corrupt AUs	2		$35 \approx S \leq 40$
			$2 < C \approx 3$
Precinct	$ku_i$	cumulative margin error	
SL			
SL2004	254.8	254.8	
SL2052	235.6	490.4	
SL1308	230.4		
SL2034	226.4		
SL2226	224.8		
SL2056	223.2		
SL2006	220.4		
SL2030	215.6		
SL2214	202.8		
SL2008	198		
SL2224	198		
SL1216	196.8		
SL2049	194.8		
SL1214	192.8		
SL2242	188		
SL2036	181.2		
SL1306	176.8		
SL2206	174		
SL2014	173.2		
SL2252	172		
SL1302	171.2		
SL1327	167.2		
SL2003	166.8		
SL2222	166.4		
SL2246	166.4		
SL1218	161.6		
SL1328	156.8		
SL2038	149.6		
SL2254	146		
SL2007	139.6		
SL2002	135.6		
SL1320	132		
SL2204	127.2		
SL1322	126.4		
SL1350	119.2		
SL1346	116		
SL1210	103.2		
SL1303	99.6		
SL2050	95.2		
SL2053	94.8		
SL1351	82.4		
SL2001	36		
SL2216	22.8		
SL2005	0		

Table 9: This table shows the improved and new uniform sampling calculations for the 2004 Utah State Senate District # 1 narrow-margin contest (Data shown in Table 7). The confidence probability is  $P = 0.99$  and the maximum level of undetectability is  $k = 0.4$ . The prior unimproved uniform method using  $2sv_i$  error bounds calculates a significantly smaller sample size of 34 audit units for this contest because it under-estimates the within audit unit maximum margin error that could occur within each audit unit.

Example # 2 applying the Improved Weighted Sampling Methods					
Expected Sample Size	Improved PPMEB approach			Improved PPMEBWR approach	
	40			34	
				#Draws	3
Precinct	$c_i$	$p_i$		$p_i$	$1 - (1 - p)^t$
SL2004	0.98	0.99	-	0.03	0.92
SL2052	1.10	0.98	-	0.03	0.90
SL1308	1.07	0.99	-	0.03	0.91
SL2034	1.02	0.99	-	0.03	0.92
SL2226	1.15	0.98	-	0.03	0.89
SL2056	1.11	0.98	-	0.03	0.90
SL2006	1.01	0.99	-	0.03	0.92
SL2030	1.07	0.99	-	0.03	0.91
SL2214	1.33	0.97	-	0.03	0.85
SL2008	1.26	0.97	-	0.03	0.87
SL2224	1.39	0.96	-	0.02	0.84
SL1216	1.18	0.98	-	0.03	0.88
SL2049	1.15	0.98	-	0.03	0.89
SL1214	1.09	0.99	-	0.03	0.90
SL2242	1.30	0.97	-	0.03	0.86
SL2036	1.18	0.98	-	0.03	0.88
SL1306	1.30	0.97	-	0.03	0.86
SL2206	1.59	0.94	-	0.02	0.80
SL2014	1.43	0.96	-	0.02	0.83
SL2252	1.57	0.95	-	0.02	0.80
SL1302	1.30	0.97	-	0.03	0.86
SL1327	1.33	0.97	-	0.03	0.85
SL2003	1.49	0.95	-	0.02	0.82
SL2222	1.49	0.95	-	0.02	0.82
SL2246	1.51	0.95	-	0.02	0.81
SL1218	1.55	0.95	-	0.02	0.80
SL1328	1.53	0.95	-	0.02	0.81
SL2038	1.86	0.92	-	0.02	0.74
SL2254	1.78	0.93	-	0.02	0.76
SL2007	1.39	0.96	-	0.02	0.84
SL2002	1.74	0.93	-	0.02	0.77
SL1320	1.53	0.95	-	0.02	0.81
SL2204	2.12	0.89	-	0.02	0.70
SL1322	1.87	0.91	-	0.02	0.74
SL1350	1.82	0.92	-	0.02	0.75
SL1346	1.78	0.92	-	0.02	0.76
SL1210	2.30	0.86	-	0.01	0.67
SL1303	1.91	0.91	-	0.02	0.73
SL2050	2.19	0.88	-	0.02	0.68
SL2053	2.82	0.81	-	0.01	0.59
SL1351	2.53	0.84	-	0.01	0.63
SL2001	7.44	0.46	-	0.00	0.29
SL2216	12.06	0.32	-	0.00	0.19

Table 10: This table shows two improved weighted sampling methods for risk-limiting post-election audits applied to the 2004 Utah State Senate District # 1 narrow-margin contest (data shown in Table 7). This example uses a confidence probability of  $P = 0.99$  and a maximum level of undetectability of  $k = 0.4$ . In this case, the improved PPMEBWR method shows an expected sample size of 34 audit units versus the old PPMEBR method using the  $2sv$  error bound that would calculate a slightly smaller sample size of 33 audit units. The improved PPMEB method is more conservative than the improved PPMEBWR method with an expected sample size of 40 audit units.

## References

- Appel, Andrew W. 1997 (Mar.). *Effective audit policy for voter-verified paper ballots in New Jersey*. Princeton Computer Science Department Web site. <http://www.cs.princeton.edu/~appel/papers/appel-nj-audits.pdf>.
- Aslam, Javed A., Popa, Raluca A., & Rivest, Ronald L. 2007. On estimating the size and confidence of a statistical audit. *In: Proceedings of the Electronic Voting Technology Workshop (EVT'07), August 6, 2007, Boston, MA*. Berkeley, CA, USA: USENIX Association. <http://www.usenix.org/events/evt07/tech/>, <http://people.csail.mit.edu/rivest/AslamPopaRivest-OnEstimatingTheSizeAndConfidenceOfAStatisticalAudit.pdf>.
- Aslam, Javed A., Popa, Raluca A., & Rivest, Ronald L. 2008 (Jan.). *On auditing elections when precincts have different sizes*. Massachusetts Institute of Technology Web site. <http://people.csail.mit.edu/rivest/AslamPopaRivest-OnAuditingElectionsWhenPrecinctsHaveDifferentSizes.pdf>.
- Atkeson, Lonna Rae, Alvarez, R. Michael, & Hall, Thad E. 2008 (Sept.). *The New Mexico 2006 Post Election Audit Report*. Pew Center on the States Web site. [http://www.pewcenteronthestates.org/uploadedFiles/NM\\_Audit\\_Report1.pdf](http://www.pewcenteronthestates.org/uploadedFiles/NM_Audit_Report1.pdf).
- Burr, Bill. 2007 (May). Introduction & VVPR. *In: Technical Guidelines Development Committee, May 21-22, 2007, Plenary Meeting, Computer Security Division of NIST*. <http://vote.nist.gov/meeting-05212007/Burr-VVPR.pdf>.
- Burr, W., Kelsey, J., Peralta, R., & Wack, J. 2006 (Dec.). *Requiring software independence in VVSG 2007: STS Recommendations for the TGDC Technical Guidelines Development Committee for the EAC*. National Institute of Standards and Technology. <http://vote.nist.gov/DraftWhitePaperOnSIinVVSG2007-20061120.pdf>.
- Calandrino, Joseph A., Halderman, J. Alex, & Felten, Edward W. 2007a (Aug.). *In defense of pseudorandom sample selection*. Center for Information Technology Policy and Dept. of Computer Science, Princeton University Woodrow Wilson School of Public and International Affairs. [http://www.usenix.org/event/evt08/tech/full\\_papers/calandrino/calandrino\\_html/](http://www.usenix.org/event/evt08/tech/full_papers/calandrino/calandrino_html/).
- Calandrino, Joseph A., Halderman, J. Alex, & Felten, Edward W. 2007b (Aug.). *Machine-assisted election auditing*. Center for Information Technology Policy and Dept. of

- Computer Science, Princeton University Woodrow Wilson School of Public and International Affairs. <http://itpolicy.princeton.edu/voting/audit07full.pdf>.
- Cordero, Arel, Wagner, David, & Dill, David. 2006 (June). *The role of dice in election audits*. University of California at Berkeley Web site. <http://www.cs.berkeley.edu/~daw/papers/dice-wote06.pdf>.
- Democracy, Center For, & Management, Election. 2005 (Sept.). *Building confidence in US elections — Report of the Commission on Federal Election Reform*. American University. [http://www.american.edu/ia/cfer/report/full\\_report.pdf](http://www.american.edu/ia/cfer/report/full_report.pdf).
- Deputy Secretary of State Anthony Stevens. 2007. *Hand counting using on-going verification*. Office of the Secretary of State of New Hampshire. <http://utahcountvotes.org/legislature/NH-Manual-Counting/Hand-Counting-and-Ongoing-Verification.ppt>.
- Dill, David. 2005 (Apr.). *David dill's testimony before the Commission on Federal Election Reform (The Carter-Baker Commission)*, American University, Washington, DC. Verified Voting Foundation Web site. <http://www.verifiedvotingfoundation.org/article.php?id=5987>.
- Dopp, Kathy. 2007a (May). *First US scientific election audit reveals voting system flaws but questions remain unanswered — critique of the collaborative public audit of Cuyahoga County Ohio, November 2006 election*. US Count Votes Web site. <http://electionarchive.org/ucvAnalysis/OH/CuyahogaElectionAudit.pdf>.
- Dopp, Kathy. 2007b (Jan.). *How big should an election audit be? Fixed rate audits do not work for elections*. US Count Votes Web site. <http://electionmathematics.org/ucvAnalysis/US/paper-audits/ElectionAuditEstimator.pdf>.
- Dopp, Kathy. 2007–2008a (Dec.). *Derivation of the formula for the number of selection rounds for the probability proportional to margin error bound (PPMEB) method for determining samples for vote count audits*. US Count Votes Web site. <http://electionarchive.org/ucvAnalysis/US/paper-audits/PPMEB-Auditing.pdf>.
- Dopp, Kathy. 2007–2008b (Oct.). *The history of confidence election auditing development (1975 to 2008) & overview of election auditing fundamentals*. US Count Votes Web site. <http://electionarchive.org/ucvAnalysis/US/paper-audits/History-of-Election-Auditing-Development.pdf>.

- Dopp, Kathy. 2007–2008c (Feb.). *Post-election vote count audits — probability proportional to margin error bound (PPMEB) method*. US Count Votes Web site. <http://electionarchive.org/ucvAnalysis/US/paper-audits/VoteCountAudits-PPMEB.pdf>.
- Dopp, Kathy. 2008 (Feb.). *Critique of the post-election auditing recommendations of Verified Voting and the American Statistical Association*. US Count Votes Web site. <http://electionmathematics.org/ucvAnalysis/US/paper-audits/Critique-McCarthy-NJ-Audits.pdf>.
- Dopp, Kathy. 2009 (Mar.). *Checking the accuracy of election outcomes — post-election audit procedures*. not published yet.
- Dopp, Kathy, & Baiman, Ron. 2005–2006 (June). *How can independent paper audits ensure election integrity?* US Count Votes Web site. [http://electionarchive.org/ucvAnalysis/US/paper-audits/Paper\\_Audits.pdf](http://electionarchive.org/ucvAnalysis/US/paper-audits/Paper_Audits.pdf).
- Dopp, Kathy, & Stenger, Frank. 2006 (Sept.). *The election integrity audit*. US Count Votes Web site. <http://vote.nist.gov/ElectionIntegrityAudit.pdf>.
- Dopp, Kathy, & Straight, Joycelynn. 2006–2008. *Mandatory vote count audit — A legislative & administrative proposal*. US Count Votes and Utah Count Votes Web site. <http://electionarchive.org/ucvAnalysis/US/paper-audits/legislative/VoteCountAuditBillRequest.pdf>.
- Hall, Joseph. 2008a (Mar.). *Dice binning calculator for post-election audits*. Web site. <http://www.josephhall.org/dicebins.php>.
- Hall, Joseph. 2008b (Dec.). *Policy mechanisms for increasing transparency in electronic voting*. Ph.D. thesis, University of California at Berkeley. <http://josephhall.org/papers/jhall-phd.pdf>.
- Hall, Joseph Lorenzo, Miratrix, Luke W., Stark, Philip B., Briones, Melvin, Ginnold, Elaine, Oakley, Freddie, Peaden, Martin, Pellerin, Gail, tom Stanionis, & Webber, Tricia. 2009 (July). *Implementing risk-limiting post-election audits in california*. Web site. [http://arxiv.org/PS\\_cache/arxiv/pdf/0905/0905.4691v3.pdf](http://arxiv.org/PS_cache/arxiv/pdf/0905/0905.4691v3.pdf).
- Jones, Douglas W. 2004. Auditing elections. *Communications of the acm*, **47**(10), 46–50. <http://doi.acm.org/10.1145/1022594.1022622> <http://www.cs.uiowa.edu/~jones/voting/cacm2004.shtml>.

- McBurnett, Neal. 2008. *CO election auditing project — auditable reports for Hart Intercivic and Hart CCOS machines*. Web site. <http://bcn.boulder.co.us/~neal/elections/boulder-audit-08-11/>. Contains links to programs to calculate the sample size, and randomly select the sample and the contests to audit at <http://bazaar.launchpad.net/~nealmcb/electionaudits/trunk/files>, <http://bcn.boulder.co.us/~neal/electionaudits/>.
- McCarthy, John, Stanislevic, Howard, Lindeman, Mark, Ash, Arlene, Addona, Vittorio, & Batcher, Mary. 2007 (Nov.). *Percentage-based vs. SAFE vote tabulation auditing: A graphic comparison*. Web site. <http://electionaudits.org/files/Percentage-based%20vs%20SAFE%20Vote%20Tabulation%20Auditing%20Paper.pdf>.
- Mebane, Jr., Walter R., Sekhon, Jasjeet S., & Wand, Jonathan. 2003 (Oct.). *Detecting and correcting election irregularities*. Stanford University Web site. <http://wand.stanford.edu/research/detecting.pdf>.
- Neff, C. Andrew. 2003 (Dec.). *Election confidence — A comparison of methodologies and their relative effectiveness at achieving it*. Web site. <http://www.electionmathematics.org/em-audits/US/Neff-ElectionConfidence.pdf>.
- NIST staff. 2007. *Voluntary Voting System Guidelines, recommendations to the Election Assistance Commission, August 2007. Prepared at the request of the Technical Guidelines Development Committee*. NIST Web site. <http://vote.nist.gov/DraftWhitePaperOnSIinVVSG2007-20061120.pdf> <http://vote.nist.gov/VVSG-0807/Final-TGDC-VVSG-08312007.pdf>.
- Norden, Lawrence, Burstein, Aaron, Hall, Joseph Lorenzo, & Chen, Margaret. 2007a (Aug.). *Post-election audits: Restoring trust in elections*. Brennan Center for Justice with the Samuelson Law, Technology & Public Policy Clinic. [http://www.brennancenter.org/dynamic/subpages/download\\_file\\_50227.pdf](http://www.brennancenter.org/dynamic/subpages/download_file_50227.pdf) Executive Summary [http://www.brennancenter.org/dynamic/subpages/download\\_file\\_50228.pdf](http://www.brennancenter.org/dynamic/subpages/download_file_50228.pdf) [http://www.brennancenter.org/dynamic/subpages/download\\_file\\_50109.pdf](http://www.brennancenter.org/dynamic/subpages/download_file_50109.pdf).
- Norden, Lawrence, Burstein, Aaron, Hall, Joseph Lorenzo, Dill, David, Hoke, Candice, Mebane, Jr., Walter, Oakley, Freddie, Rivest, Ronald L., & Wagner, David. 2007b (Feb.). *Thoughts on mandatory audits*. The Brennan Center Web site. [http://www.brennancenter.org/dynamic/subpages/download\\_file\\_47860.pdf](http://www.brennancenter.org/dynamic/subpages/download_file_47860.pdf).

- Project ACCURATE. 2005 (Sept.). *Public comment on the 2005 Voluntary Voting System Guidelines*. Submitted to the United States Election Assistance Commission, Samuelson Law, Technology & Public Policy Clinic, University of California, Berkeley. [http://accurate-voting.org/accurate/docs/2005\\_vvsg\\_comment.pdf](http://accurate-voting.org/accurate/docs/2005_vvsg_comment.pdf).
- Rivest, Ronald L. 2007 (Nov.). *A simple rule of thumb for election audit size determination*. MIT Web site. <http://people.csail.mit.edu/rivest/Rivest-ASimpleRuleOfThumbForElectionAuditSizeDetermination.pdf>.
- Rivest, Ronald L. 2008 (Apr.). *A Sum of Square Roots (SSR) pseudorandom sampling method for election audits*. MIT Web site. <http://people.csail.mit.edu/rivest/Rivest-ASumOfSquareRootsSSRPseudorandomSamplingMethodForElectionAudits.pdf>.
- Rivest, Ronald L., & Wack, John P. 2006 (July). *On the notion of “software independence” in voting systems*. NIST Web site. <http://vote.nist.gov/SI-in-voting.pdf>.
- Saltman, Roy G. 1975 (Mar.). *Effective use of computing technology in vote-tallying*. Technical Report NBSIR 75-687. National Bureau of Standards, Information Technology Division. [http://csrc.nist.gov/publications/nistpubs/NBS\\_SP\\_500-30.pdf](http://csrc.nist.gov/publications/nistpubs/NBS_SP_500-30.pdf).
- Stanislevic, Howard. 2006. *Random auditing of E-voting systems: How much is enough?* <http://www.votetrustusa.org/pdfs/VTTF/EVEPAuditing.pdf>.
- Stark, Philip B. 2008a (Aug.). *CAST: Canvass Audit by Sampling and Testing*. Department of Statistics, University of California, Berkeley. <http://statistics.berkeley.edu/~stark/Preprints/cast08.pdf> and <http://www.stat.berkeley.edu/~stark/Seminars/ksu08.pdf>.
- Stark, Philip B. 2008b. Conservative statistical post-election audits. *The annals of applied statistics*, **2**(2), 550–581. <http://statistics.berkeley.edu/~stark/Preprints/conservativeElectionAudits07.pdf>.
- Stark, Philip B. 2008c. *Election audits by sampling with probability proportional to an error bound: dealing with discrepancies*. Department of Statistics, University of California, Berkeley. <http://statistics.berkeley.edu/~stark/Preprints/ppebwrwd08.pdf>.
- Stark, Philip B. 2008d. A sharper discrepancy measure for post-election audits. *The annals of applied statistics*, **2**(3), 982–985. <http://statistics.berkeley.edu/~stark/Preprints/pairwise08.pdf>.

- Stark, Philip B. 2009a (May). *Auditing a collection of races simultaneously*. Department of Statistics, University of California, Berkeley. <http://www.stat.berkeley.edu/~stark/Preprints/simultaneous09.pdf>.
- Stark, Philip B. 2009b (Feb.). *Risk-limiting post-election audits: P-values from common probability inequalities*. Department of Statistics, University of California, Berkeley. <http://www.stat.berkeley.edu/~stark/Preprints/pvalues09.pdf>.
- The Collaborative Audit Committee. 2007 (Apr.). *Collaborative public audit of the November 2006 General Election*. Center for Election Integrity for the Cuyahoga County Board of Elections. [http://electionmathematics.org/em-audits/OH/2006Audit/cuyahoga\\_audit\\_report.pdf](http://electionmathematics.org/em-audits/OH/2006Audit/cuyahoga_audit_report.pdf).
- The Election Audits Task Force. 2009 (Jan.). *Report on election auditing*. [http://www.lwv.org/Content/ContentGroups/Membership/ProjectsTaskforces/Report\\_ElectionAudits.pdf](http://www.lwv.org/Content/ContentGroups/Membership/ProjectsTaskforces/Report_ElectionAudits.pdf).
- US General Accounting Office. 2005. *Federal efforts to improve security and reliability of electronic voting systems are under way, but key activities need to be completed*. US GAO Web site. <http://www.gao.gov/new.items/d05956.pdf>.
- Walmsley, Paul. 2005. *Requirements for statistical live auditing of optical-scan and VVPAT records — and for live auditing for vote tabulation*. Web site. <http://www.booyaka.com/~paul/ea/eac-20050930/live-audit-overview.txt> <http://www.booyaka.com/~paul/ea/eac-20050930/interpretation-live-audit.txt> <http://www.booyaka.com/~paul/ea/eac-20050930/tabulation-live-audit.txt>.
- Wand, Jonathan. 2004. *Auditing an election using sampling: The impact of bin size on the probability of detecting manipulation*. Stanford Web site. <http://wand.stanford.edu/elections/probability.pdf>.