

## Step-by-Step Instructions for a Risk Limiting Audit Using the Ballot Comparison Method

### Introduction

The Risk Limiting Audit (RLA) involves comparing the interpretation of ballots by the voting system to human interpretations of the same ballots. Differences between the two interpretations are noted. Determining whether the audit can stop depends on the number and nature of those differences, the number of ballots examined so far, the risk limit, and the vote margin; the smaller the risk limit or the smaller the margin, the larger the number of ballots that must be audited.

The comparison of interpretations can be neutral, an *understatement*, or an *overstatement*, depending on the effect of changing the voting system interpretation of the ballot to match the hand interpretation. Consider the pairwise margin between each winner and each loser in a contest. For instance, a city council election might involve voting for three candidates from a pool of ten, to fill three seats on the council. Each of the three winners can be paired with each of the seven losers, giving twenty-one pairwise margins in that contest. If changing the interpretation of a ballot according to the voting system to make it match the human interpretation of the ballot would widen every pairwise margin in every contest under audit, that ballot has an *understatement*. Understatements do not call the outcome into question. If changing the interpretation according to the voting system to match the human interpretation would narrow any pairwise margin in any contest under audit, the ballot has an *overstatement*. If enough ballots have overstatements, the outcome could be wrong.

Efficient RLAs generally perform ballot interpretation comparisons until there is convincing evidence that the outcome according to a full hand count would agree with the outcome under audit. If convincing evidence is not forthcoming, the audit progresses to a full hand count, which is used to correct the outcome under audit if the two disagree.

Because the mock election will use previously prepared ballots, the exercise will not accurately mimic a real election in all ways. For example, the prepared ballots have no write-ins or overvotes and write-ins can't be added because any such modification will create overvotes that are treated differently than real write-in ballots. The only write-ins that might occur will be from the 300 ballots filled out by EIC members. While overvotes could be added to the mock election ballots, they would not have any effect on the election or RLA because they would just need to be rescanned and assigned new sequential ID numbers.

As this mock election RLA is a pilot study, we do not need to conduct a full audit for all contests. We can select any sample size, but the objective is to obtain an estimate of the time and resources involved in conducting RLAs for larger more realistic elections.

For this mock election, we will begin by conducting a Logic and Accuracy (L&A) test that will consist of about 300 ballots filled out by EIC members. Each member will provide a spreadsheet report of the vote totals in their ballot set. These will be combined to obtain expected tallies for

the full 300 ballot set. These 300 ballots will be added to the general population of 29,699 mock election ballots.

## **Step-by-Step Scanning and RLA Instructions**

**1. Scan ballots and track ballot storage locations.** Scan ballots in batches, imprinting a unique serial number on each ballot. Keep ballots in the order in which they were scanned and create labeled stacks or boxes of scanned ballots with labeling designed so that individual ballots selected during the audit can be retrieved efficiently. If ballots need to be rescanned for any reason, take precautions to rotate the ballot so that the new serial number is not printed on top of an older one. Create a “ballot manifest” that lists how scanned ballots are organized (labeling of stacks, listing the number of ballots in each stack – see Step 7). Write-in ballots must be tracked separately from the general population. Conduct quality control measures to ensure that each stack of ballots contains the number of ballots listed in the manifest. During the audit, ballots will be randomly selected from the ballot manifest. Reconcile the ballot manifest against other accountings of the number of ballots and against poll books to the extent possible. Resolve any discrepancies before the audit begins.

**2. Create a Statement of Votes Cast (SOVC)** or other report or canvass with final vote tallies for each contest.

**3. Create a list of Cast Vote Records (CVRs)**, a spreadsheet in which each record specifies how an individual physical ballot was interpreted by the tabulation system. Each record must include the CVR number and the ballot serial number. The audit will compare a sample of physical ballots with the CVRs that correspond to those ballots. The serial numbers on the physical ballots will be used to identify the corresponding CVR.

**4. Validate the CVRs.** Use a program to read the CVRs and tabulate the votes in all the contests contained therein. The results of this process should yield final tallies that are identical to those shown in the SOVC. If they are not, the CVRs and/or the SOVC are inaccurate. However, if the ballot manifest and the CVRs agree on the total number of ballots in each contest, and agree on the winners of each contest, the audit can proceed. In that case, the margin that should be used in the computation below is the margin according to the re-tabulation of the CVRs, rather than according to the SOVC.

**5. Determine the contests to be audited.** The contests to be audited may be determined by state or local laws or they may be selected randomly by agreement among local observers, political parties and election officials. If most of the contests are on the ballots of most precincts in the jurisdiction, there is not much time penalty to pay in auditing all the contests simultaneously. If however, a specific contest involves only a small part of the jurisdiction, it may be more efficient to audit that contest separately.

**6. Determine the size of the initial audit sample.** The initial sample size depends on the “diluted margin” which is the margin of victory in votes divided by the total number of ballots cast in each contest to be audited (not the number of valid votes in a contest). Dividing by the

number of ballots accounts for the possibility of confusing undervotes or overvotes for a valid vote, or vice versa. Go to the [Post-Election Audit Tools Website](#) and follow the instructions to enter contest data. The elections official may instead develop software tools based on the audit model (See Appendix). For purposes of the pilot program, the risk limit setting should be left at 10%.

Illustration: For zero expected over/understatements, the math is simple. The number 4.8 divided by the diluted margin provides the initial sample size for a 10% risk limit:

| Diluted Margin | Equation | Initial Sample Size |
|----------------|----------|---------------------|
| 50%            | 4.8/.5   | 10                  |
| 30%            | 4.8/.3   | 16                  |
| 10%            | 4.8/.1   | 48                  |
| 2%             | 4.8/.02  | 240                 |

### 7. Randomly select ballots for the sample:

1. Ask public observers to roll ten to twenty 10-sided dice to generate a “seed” number for a public-source pseudo-random number generator.
2. Enter the seed on the “Random Sampling” section of the [Audit Tools Website](#).
3. Enter the number of ballots in the contest(s) to be audited.
4. Click “draw sample” to select ballots.
5. Each random number corresponds to one ballot. For the mock election RLA the audit tools website can be used to randomly select CVR numbers. Each CVR number corresponds to a unique serial number. The serial numbers should be sorted to identify the set of serial numbers in each box of ballots so that ballot boxes are only visited once.
6. For each serial number, check the CVR to see if the ballot is a write-in, in which case the search for the specific ballot can be limited to the separated write-in subset.
7. Given the selected serial numbers, the ballots are retrieved by a small team of “retrievers.” Insert colored paper as a place-holder for each ballot so that extracted ballots can be easily reinserted after Step 8.

Note: the “ballot manifest” is any tracking tool that facilitates these steps. We are expecting that this will take the form of a spreadsheet that can be sorted, searched, and highlighted.

**8. Compare ballots in the sample with the corresponding CVR entries.** Because risk-limiting audits involve comparing individual ballots to the voting system results for each ballot, the audits do not involve a “tally” or count in the traditional way a 1% manual tally is conducted. That is, a 1% manual tally ballot counters are assembled to tally entire precincts or batches of ballots and compare the hand tally totals to the totals produced by the voting system for the same precinct or batch. For risk-limiting audits, a human eye interpretation of each ballot is compared to the CVR for that ballot as recorded by the voting system, so ballots are not “tallied” or counted up and totaled in the usual manner. Instead, we will track any discrepancies we find.

Compare each ballot as follows:

1. Retrieve the ballots chosen for the sample. Those designated to retrieve ballots should not have access to the CVRs for the ballots they retrieve in order to ensure the integrity of the audit.
2. Retrieve the CVR entry for each ballot using the serial number on the ballot (the highest serial number if there are multiple serial numbers) and determine whether the entry matches the human eye interpretation of the votes on the corresponding ballot. Existing ballot interpretation methods may be used, including local rules for determining voter intent.
3. Ensure public observers have the opportunity to compare the CVR entry with the physical ballot. If the public observers or auditors disagree about the interpretation of the votes on the ballot, record the disagreement in a log of the audit process.
4. Record the interpretation of the vote on the physical ballot as determined by the auditors in the log of the audit process. This will be used if the audit escalates to all ballots.
5. Document and share with the public any differences found between the human interpretation and the voting system interpretation. Keep accurate track of all observed discrepancies and the specific candidates they apply to. Tally the understatements and overstatements, if any; they are needed to size the escalated audit sample, if necessary.
6. Document and share with the public any instances in which one or more public observers disagreed with the comparison.
7. Establish procedures to handle observer challenges to the audit. The public must be allowed to observe, verify and point out procedural problems without interfering with the process.
8. Return each sample ballot to their original boxes.

**9. Stop or escalate the sample size if necessary.** Depending on the number and type of overstatements and understatements found in the initial sample, the audit may need to be expanded to look at more ballots. To determine how many more ballots should be hand tallied, if any, assuming a similar rate of over/under statements:

1. Go to the [Post-Election Audit Tools Website](#) and follow the instructions to determine whether escalation is necessary.
2. If escalation is necessary, the elections official should explain to the public that the audit might lead to a full hand count if significant differences persist.
3. Retrieve the additional ballots, if any, required by the escalation instructions, and continue the ballot interpretation comparison process until either the process terminates (no additional escalation) or all ballots have been compared.

**10. Finish and publish results.** If the audit process terminated before all ballots have been compared (the most likely case), then the outcomes (winners) reported by the tabulation system in the SOVC are accepted. If the process terminated because all ballots were compared, then the set of vote discrepancies observed during the audit and the log of ballot interpretations can be used to adjust the outcomes reported in the SOVC. In this case, the outcomes of the election are those resulting from the audit. In either case:

1. Release the results of the audit to the public and the Secretary of State.
2. Record and report the time it took to conduct the audit, with a breakdown of the time needed to scan ballots compared to the time needed to conduct the RLA.
3. Record and report the cost of the audit.

**Appendix.** Detailed Procedures for the RLA in case someone wants to implement all the steps rather than using the online Audit Tools Website.

1. Validating the CVRs in Step 4. This step requires a program that accumulates ballot selections for each contest. This program will need to be specific to the format of the CVRs or the list of CVRs in spreadsheet form (LVR).
2. Initial sample size in Step 6. The initial sample size is given by the formula:

$$n_0 = -2g \log_e(a) / ((m + 2g(r_1 \log_e(1 - 1/(2g)) + r_2 \log_e(1 - 1/g) + s_1 \log_e(1 + 1/(2g)) + s_2 \log_e(1 + 1/g))))$$

with

m = diluted margin,

a = risk limit,

g = 1.03905,

r<sub>1</sub> = expected rate of 1-vote overstatements per ballot,

r<sub>2</sub> = expected rate of 2-vote overstatements per ballot,

s<sub>1</sub> = expected rate of 1-vote understatements per ballot, and

s<sub>2</sub> = expected rate of 2-vote understatements per ballot.

The diluted margin is the smallest margin in votes, divided by the total number of ballots cast for each contest, including undervoted and overvoted ballots.

The number n<sub>0</sub> is then adjusted to take into account the fact that differences must be round numbers, as follows: The expected number of differences in the sample of each type is n<sub>0</sub> times the expected rate of those differences. The expected numbers are either rounded to the nearest whole number, or rounded up. Then those numbers of discrepancies are plugged into the stopping rule described below, to determine how many ballots would have to be audited if the estimated number of differences of each type were to be observed in the sample. That number is then used once again to estimate the number of differences of each type the sample would contain; the results are rounded to the nearest integer and plugged into the stopping rule a second time. The result is then the starting sample size.

3. Random number generator in Step 7. The AuditTools page implements a good pseudo-random number generator based on the SHA-256 hash function. There are [many other possibilities](#), both programs and servers.
4. Stopping rule and escalation size in Step 9. The stopping rule implements the following formula:

$$\text{stopping sample size} = -2g(\log_e(a) + o_1 \log_e(1 - 1/(2g)) + o_2 \log_e(1 - 1/g) + u_1 \log_e(1 + 1/(2g)) + u_2 \log_e(1 + 1/g)) / m$$

with

m = diluted margin,

$a$  = risk limit,  
 $o_1$  = number of 1-vote overstatements in the sample,  
 $o_2$  = number of 2-vote overstatements in the sample,  
 $u_1$  = number of 1-vote understatements in the sample, and  
 $u_2$  = number of 2-vote understatements in the sample,  
 $g = 1.03905$ , but any value greater than one can be used. For  $g = 1.03905$ , a two-vote overstatement increases the sample size by five times as much as a one-vote overstatement. The estimates based on differences continuing to occur at the observed rate are based on the method described above for estimating the initial sample size, including the method of rounding the expected number of differences of each type.

5. Outcome adjustment, if necessary, in Step 10. If the audit goes to a full hand count, the vote totals in the SOVC will need adjustment according to the recorded discrepancies (under-counts and over-counts for each candidate), together with the manual tracking in step 8.4.

### **The January 2017 Mock Election**

The mock election, as originally planned, consists of 29699 ballots, all with the same 6 contests. Since these ballots have already been scanned, we know the official outcomes:

#### Best Picture (vote for 3)

|                   |       |
|-------------------|-------|
| Bridge of Spies   | 15500 |
| Mad Max Fury Road | 17699 |
| The Big Short     | 16499 |
| The Brooklyn      | 8200  |
| The Martian       | 13400 |
| The Revenant      | 12000 |
| The Room          | 5000  |
| The Spotlight     | 799   |

#### Best Actor in a Leading Role (vote for 1)

|                    |       |
|--------------------|-------|
| Eddie Redmayne     | 5000  |
| Leonardo DiCaprio  | 799   |
| Matt Damon         | 12000 |
| Michael Fassbender | 11900 |

#### Best Actress in a Leading Role (vote for 1)

|                    |      |
|--------------------|------|
| Brie Larson        | 5500 |
| Cate Blanchett     | 5000 |
| Charlotte Rampling | 6900 |
| Jennifer Lawrence  | 5799 |
| Saoirse Ronan      | 6500 |

Best Actor in a Supporting Role (vote for 2)

|                    |       |
|--------------------|-------|
| Christian Bale     | 16900 |
| Mark Ruffalo       | 10500 |
| Mark Rylance       | 10500 |
| Sylvester Stallone | 7299  |
| Tom Hardy          | 14199 |

Best Actress in a Supporting Role (vote for 2)

|                      |       |
|----------------------|-------|
| Alicia Viander       | 14200 |
| Jennifer Jason Leigh | 12699 |
| Kate Winslet         | 6500  |
| Rachel McAdams       | 11299 |
| Rooney Mara          | 14700 |

Oscars be Funded by the US Government? (vote for 1)

|     |       |
|-----|-------|
| No  | 14999 |
| Yes | 14700 |

These ballots will be scanned again during the mock election so it is possible that a few ballots will be interpreted differently. There are no undervotes on these ballots.

The diluted vote margin for each of these contests and the corresponding initial sample sizes (for a 10% risk limit and zero anticipated understatements and overstatements) are as follows:

| Category                | Diluted Margin               | Initial Sample Size (4.8/margin) |
|-------------------------|------------------------------|----------------------------------|
| Best Picture            | $(15500-13400)/29699=0.071$  | 68                               |
| Best Actor              | $(12000-11900)/29699=0.0034$ | 1426                             |
| Best Actress            | $(6900 - 6500)/29699=0.0135$ | 357                              |
| Best Supporting Actor   | $(14199-10500)/29699=0.125$  | 39                               |
| Best Supporting Actress | $(14200-12699)/29699=0.051$  | 95                               |
| Oscars Funded by U.S.?  | $(14999-14700)/29969=0.010$  | 480                              |

These 29699 ballots will be augmented with 300 ballots filled out by Commission members. The outcomes (winners) are unlikely to change, but the margins and initial audit sizes may be adjusted slightly.