

Risk-Limiting Audits

Workshop on Post-Election Audits
American Statistical Association
Alexandria, VA
23–24 October 2009

Philip B. Stark

In theory, there's no difference between theory and practice.

But in practice, there is.

Jan L.A. van de Snepscheut

Problem: Any way of counting votes makes mistakes.

If there are enough mistakes, apparent winner could be wrong.

If there's a complete, accurate audit trail, can ensure big chance of fixing wrong outcomes.

Crucial question: when to *stop* counting, not where to start.

Solution: If there's compelling evidence that outcome is right, stop; else, audit more.

Current audit laws have the wrong focus: virtually useless for fixing wrong outcomes.

Need data plumbing first, then better audit laws.

Wrong Focus

Current and proposed laws focus on how big an initial sample to draw.

Heated debates over fixed percentages, tiered percentages depending on the margin, or sample sizes that vary continuously with the margin and depend on batch sizes.

The real issue isn't where to start. It's when to stop.

Can't fix wrong outcomes without counting the whole audit trail.

Risk-Limiting Audits

If the outcome is wrong, there's a at least a [pre-specified] minimum chance of a full manual count,
no matter what caused the outcome to be wrong.

The *risk* is the maximum chance that there won't be a full hand count when a full hand count would show that the apparent outcome is wrong.

“Wrong” means disagrees with what a full hand count would show: presupposes accurate & complete audit trail, secure chain of custody, etc. Nontrivial.

Null hypothesis: outcome is wrong.

Control Type I error rate.

Role of statistics: Less counting when the outcome is right, but big chance of a full hand count when outcome is wrong.

Persistent idea that only the initial sample matters, not the errors the sample finds.

E.g., Holt bill.

Essential that voters create complete, durable, accurate audit trail.

Essential that voting systems enable auditors to access reported results (total ballots, counts for each candidate, registered voters) in auditable batches.

Essential to select batches at random, *after* the results are posted. (Can supplement with “targeted” samples.)

Need a plan for dealing with discrepancies, possibly leading to full count. “Explaining” or “resolving” isn’t enough.

Current audit laws do not limit risk.

Compliance audits vs. materiality audits.

Sampling Designs

Simple

Stratified (by county, voting method, other)

PPEB

NEGEXP

Stratified PPEB?

Sampling scheme affects choice of test statistic—analytic tractability

Weighted max for simple & stratified sampling.

More efficient choices possible for PPEB.

Sequential risk-limiting test

0. Calculate error bounds $\{u_p\}$, U . Set $n = 1$. Pick $\alpha \in (0, 1)$ and $m > 0$.

1. Draw a batch using PPEB. Audit it if not audited previously.

2. Find $T_n \equiv t_p \equiv e_p/u_p$, taint of the batch p just drawn.

3. Compute

$$P_n \equiv \prod_{j=1}^n \frac{1 - 1/U}{1 - T_j}. \quad (1)$$

4. If $P_n < \alpha$, stop; report apparent outcome. If $n = m$, audit remaining batches. If all batches have been audited, stop; report known outcome. Else, $n \leftarrow n + 1$ and go to 1.

This sequential procedure is risk-limiting

If outcome is wrong,

$$\mathbb{P}\{\text{stop without auditing every batch}\} < \alpha.$$

Chance $\geq 1 - \alpha$ of fixing wrong outcome by full hand count.

Remarkably efficient (in simulations).

Assessing Evidence

How strong is the evidence that the outcome is correct, given how the sample was drawn, the margin, the errors found, etc.?

What is the biggest chance that—if the outcome is wrong—the audit would have found as little error as it did?

(The definition of “little” differs across sampling methods, etc.)

5 February 2008 Marin County Measure A

First election ever audited to attain target level of confidence in the result.

Thanks to Elaine Ginnold!

Audited to attain 75% confidence that a full manual recount would find the same outcome.

Required 2/3 majority to pass. Margin 298 votes.

3 strata: in-precinct, VBM, provisionals

Confirmed outcome at $\leq 25\%$ risk (quite conservative)

Marin Measure A audit timeline

Milestone	Date
Election day	5 February
Polling place results available	7 February
Random selection of polling place precincts	14 February
VBM results available	20 February
Random selection of VBM precincts	20 February
Hand tally complete	20 February
Provisional ballot results available	29 February
Computations complete	3 March

Costs:

\$1,501, including salaries and benefits for 4 people tallying the count, a supervisor, support staff to print reports, resolve discrepancies, transport ballots and locate and retrieve VBM ballots from the batches in which they were counted.

\$0.35 per ballot audited. $1\frac{3}{4}$ days.

Marin Measure B and Santa Cruz Supervisor District 1, November 2008

Used PPEB sampling and trinomial bound.

County	Ballots	Winner	Loser	Prcts	Batches	Batches Audited	Ballots Audited	% Ballots Audited
Marin	121,295	51%	35%	189	544	14	3,347	3%
SC	26,655	45%	37%	76	152	16	7,105	27%

Marin: no discrepancies. 2 days, total cost \$1,723, \$0.51 per audited ballot.

Santa Cruz: “taints” 0.036, 0.007, -0.002, -0.003, -0.005, -0.007, -0.012; twelve were 0.

3 days, total cost \$3,248, \$0.46 per audited ballot.

Miscommunication about provisional ballots in Santa Cruz; treated as error.

Yolo County Measure W, November 2008

Davis school bond. Required simple majority. Used SRS.

batches	yes	no	undervote	overvote	margin
114	25,297	8,118	3,001	2	17,179

Stop if no batch has more than 14 overstatements.

Assumed “small” batches were entirely in error; sampled from remaining 95.

Counted about 2,500 ballots by hand on 17 November 2008.
1 extra “yes” and 1 extra “no.”

Logistical issues: stratification, etc.

Samples for different counties drawn independently: stratified.

VBM, absentee & provisional ballots not counted right away.

Makes sense to start with a uniform sampling rate, then escalate as necessary.

Simultaneous audits?

Coordination across jurisdictions?

Wrinkles

Optimal attacks against stratified samples: sharp bounds by combinatorial optimization.

Optimal sequential tests against various alternatives.

False discovery rate: limit the fraction of certified outcomes that are wrong.

Simplicity matters more than optimality!

Activists & lawmakers want tables/spreadsheets.

Hard to pass sensible laws.

Recap

- Good audits can limit the risk of certifying a wrong outcome. Sometimes requires full hand counts; else, can't fix wrong outcomes.
- Current auditing laws do not limit risk.
- There are practical ways to conduct risk-limiting audits—we've done it.
- Data plumbing is crucial! First step for any jurisdiction.
- Everything should be as simple as possible, but not wrong.