

Election Auditing: How Much Is Enough?

The Society for Political Methodology
26th Annual Summer Meeting
Institution for Social and Policy Studies
Yale University
23 July 2009

Philip B. Stark

Collaborators: Joe Hall, Mike Higgins, Luke Miratrix
Special thanks: Elaine Ginnold, Freddie Oakley, Gail Pellerin
& staff

*In theory, there's no difference between theory and practice.
But in practice, there is.*

Jan L.A. van de Snepscheut

[Election Leak] [Voting Machine Wins]
[Homer Votes—sort of]

Abstract: Abstract: Free, fair and accurate elections are the cornerstone of democracy. Post-election audits—which compare reported totals with hand counts of a trustworthy audit trail in a random sample of batches—can limit the risk of certifying an incorrect electoral outcome. The risk is the maximum chance the audit stops short of a full hand count when a full hand count would show that the outcome is wrong.

Post-election audits can measure and improve the accuracy of elections, increase transparency and bolster public confidence. Several states have post-election audits. Federal legislation requiring audits has been introduced several times. But to the best of my knowledge, no current or proposed audit law controls the risk of certifying an incorrect outcome.

Risk-limiting audits can be built as sequential tests: Data are collected. If they give strong evidence that the outcome is right, the audit stops. Otherwise, more data are collected. Eventually, the audit stops or there has been a full hand count. The strength of the evidence can be quantified using the maximum P -value of the hypothesis that the outcome is wrong given the audit data. The maximum is over all ways in which the outcome could be wrong—a nonparametric hypothesis. The P -value depends on the sampling scheme, the choice of test statistics, the number of overstatement errors in each audited batch, and more.

Abstract, cont'd

Multiplicity can be dealt with in many ways, for instance, by considering the sample to be “telescoping” and using Bonferroni’s inequality, or by conditioning on the results at each stage. The latter facilitates incorporating “targeted” sampling, which can substantially reduce requisite sample sizes. Several races can be audited simultaneously by using a test statistic that summarizes the error across races in each batch of ballots. This approach controls the familywise error rate at a cost that can be lower than controlling the per-comparison error rate using independent audits of each race.

The most efficient approach so far draws audit batches independently using probability proportional to bounds on the relative margin overstatement error in the batches. The overstatement error in the each sampled batch is expressed as “taint,” the ratio of the actual overstatement error to the bound. A sufficient condition for the outcome of the race to be correct can be expressed in terms of the expected taint. Nonparametric sequential tests can be constructed via an inequality due to Kaplan that combines a Martingale inequality and Markov’s inequality.

Abstract, cont'd

There have been four risk-limiting audits, all in California in 2008: Marin County (a small measure in February requiring a supermajority and a county-wide measure in November), Santa Cruz County (County Supervisor, District 1, November), and Yolo County (bond measure in November). We designed these audits, which were conducted in collaboration with elections officials in the counties. Several sampling techniques were tested. The audits ensured at least a 75% outcome was confirmed at the first stage; full hand counts were not required. Many lessons were learned. Clear, precise and timely communication between the auditors and the elections officials is key. Optimality is a low priority compared with simplicity. The biggest barrier to wide-scale risk-limiting audits is the inability of current election management systems (vote tabulation systems) to export data in a useful, machine-readable format. Adopting standard terminology and data formats would be extremely helpful.

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.

Did “the right thing” in 4 contests in California in 2008; affordable, but methods maturing. More tests in November 2009.

Need data plumbing first, then better audit laws.

Humboldt County CA, 2008

Serious Error in Diebold Voting Software Caused Lost Ballots in California County, by Kim Zetter

Election officials in a small county in California discovered by chance last week that the tabulation software they used to tally votes in this year's general election dropped 197 paper ballots from the totals at one precinct. The system's audit log also appears to have deleted any sign that the ballots had ever been recorded.

Premier has acknowledged . . . its software caused the system to delete votes. The company has apparently known about the problem since 2004 . . .

[RoV] Crnich would never have discovered the problem through her standard canvassing procedures . . . nor would she have discovered it while conducting a mandatory manual audit that California counties are required to do.

Crnich discovered the missing ballots only because she happened to implement a new and innovative auditing system this year that was spearheaded by members of the public who helped her develop it.

Wired News, 8 December 2008 <http://blog.wired.com/27bstroke6/2008/12/unique-election.html>

SoS Bowen's response

SACRAMENTO Secretary of State Debra Bowen today announced she has withdrawn state approval of Premier Election Solutions Global Election Management System (GEMS) version 1.18.19, which contains serious software flaws.

Premier GEMS 1.18.19 contains the “Deck Zero” anomaly, a software error that can delete the first batch of optically scanned ballots under certain circumstances without alerting elections officials to the deletion. In addition, the systems audit logs fail to record important events and clear buttons permit deletion of key records, both of which violate federal standards. . . .

“Clearly, a voting system that can delete ballots without warning and doesn’t leave an accurate audit trail should not be used in California or anywhere,” said Secretary Bowen, California’s chief elections officer.

<http://www.sos.ca.gov/admin/news-releases.htm>

Polk County NC, 2008

Owens victory in Polk is in doubt, by Times-News staff

Ted Owens went to sleep Tuesday night thinking he had earned another term . . . A recount Wednesday showed he may not have. . . .

Computer software initially displayed figures that were different than those shown by the voting machines . . .

The software installed in the stand-alone computer that ballot results are fed into was the problem . . . [Elections Director Dale Edwards] said there was no explanation as to why the computer counted the wrong numbers, and no one is at fault.

BlueridgeNow.com Times-News, 6 November 2008 <http://www.blueridgenow.com/article/20081106/NEWS/811050255>

Santa Clara County, CA, 2008

Few problems reported in area despite record turnout,
by Karen de Sá and Lisa Fernandez

Record-high voting in the Bay Area on Tuesday mostly defied predictions of unwieldy waits and overwhelmed polls. But in Santa Clara County, concerns about touch-screen voting machines will likely increase following significant malfunctions.

Fifty-seven of the county's Sequoia Voting Systems machines failed on Election Day, resulting in hourslong delays before replacements arrived.

Mercury News, 4 November 2008 http://www.mercurynews.com/elections/ci_10901166?nclick_check=1

Leon County, FL, 2008

Ballots not being recorded at two Leon County polling places, by Angeline J. Taylor

Leon County Supervisor of Elections Ion Sancho has reported that ballots . . . are not being read properly. The problem, he said, rests with a new machine that has been purchased for polling sites throughout the state. . . .

“Certain ballots are being rejected across the state,” he said. . . . If the machine reads the ballot card as too long, the . . . machine will simply not read the card.

Tallahassee Democrat, 20 October 2008 [http://www.tallahassee.com/
article/20081020/BREAKINGNEWS/81020024](http://www.tallahassee.com/article/20081020/BREAKINGNEWS/81020024)

Palm Beach County, FL, 2008

Florida Primary Recount Surfaces Grave Voting Problems One Month Before Presidential Election, by Kim Zetter

At issue is an August 26 primary election in which officials discovered, during a recount of a close judicial race, that more than 3,400 ballots had mysteriously disappeared after they were initially counted on election day. The recount a week later, minus the missing ballots, flipped the results of the race to a different winner.

...officials found an additional 227 ballots that were never counted on election day
...in boxes in the county's tabulation center.

Palm Beach County was using new optical-scan machines that it recently purchased from Sequoia Voting Systems for \$5.5 million.

Palm Beach County, FL, 2008, cont'd

[In a re-scan of ballots the machines had rejected] [o]fficials expected the machines would reject the same ballots again. But that didn't happen. During a first test of 160 ballots, the machines accepted three of them. In a second test of 102 ballots, the machines accepted 13 of them . . . When the same ballots were run through the machines again, 90 of the ballots were accepted.

[T]he county then re-scanned two batches of 51 ballots each that had initially been rejected for having no vote cast in the judicial race, but that were found in a manual examination to contain legitimate votes for one candidate or the other. The first batch of 51 ballots were found to have legitimate votes for Abramson. The second batch of 51 ballots were found to have legitimate votes for Wernet.

In the first batch of 51 ballots . . . 11 of the ballots that had previously been rejected as undervotes were now accepted . . . the remaining 40 ballots were rejected as having no votes. In the second batch of 51 ballots . . . the same machine accepted 2 ballots and rejected 49.

Palm Beach County, FL, 2008, cont'd

The same two batches of ballots were then run through the second ... machine. [I]n the first batch ... the machine accepted 41 ... and rejected 10 others. In the second batch ... the machine accepted 49 of the ballots and rejected 2—the exact opposite of the results from the first machine.

Wired News, 7 October 2008, <http://blog.wired.com/27bstroke6/2008/10/florida-countys.html>

Washington, DC, 2008

Report Blames Speed In Primary Vote Error; Exact Cause of Defect Not Pinpointed, by Nikita Stewart

Speed might have contributed to the Sept. 9 primary debacle involving thousands of phantom votes, according to a D.C. Board of Elections and Ethics report issued yesterday. . . . [T]he report does not offer a definitive explanation. . . .

The infamous Precinct 141 cartridge “had inexplicably added randomly generated numbers to the totals that had been reported,” according to the report written by the elections board’s internal investigative team.

. . . 4,759 votes were reflected instead of the actual 326 cast there.

Washington Post, 2 October 2008; Page B02

see also hearings at http://www.octt.dc.gov/services/on_demand_video/channel13/October2008/10_03_08_PUBSVRC_2.aspx

New Jersey 2008

County finds vote errors: Discrepancies discovered in 5% of machines, by Robert Stern

Five percent of the 600 electronic voting machines used in Mercer County during the Feb. 5 presidential primary recorded inaccurate voter turnout totals, county officials said yesterday . . .

23 February 2008, New Jersey Times

Ohio 2004

Machine Error Gives Bush Thousands of Extra Ohio Votes, by John McCarthy

COLUMBUS, Ohio – An error with an electronic voting system gave President Bush 3,893 extra votes in suburban Columbus, elections officials said. Franklin County's unofficial results had Bush receiving 4,258 votes to Democrat John Kerry's 260 votes in a precinct in Gahanna. Records show only 638 voters cast ballots in that precinct. Bush's total should have been recorded as 365.

5 November 2004, Associated Press

Florida 2004

Broward Machines Count Backward, by *Eliot Kleinberg*

[E]arly Thursday, as Broward County elections officials wrapped up after a long day of canvassing votes, something unusual caught their eye. Tallies should go up as more votes are counted. That's simple math. But in some races, the numbers had gone . . . down.

Officials found the software used in Broward can handle only 32,000 votes per precinct. After that, the system starts counting backward.

. . . The problem cropped up in the 2002 election. . . . Broward elections officials said they had thought the problem was fixed.

5 November 2004, The Palm Beach Post

California Elections Code §15360

[T]he official conducting the election shall conduct a public manual tally of the ballots tabulated by those devices, including absent voters' ballots, cast in 1 percent of the precincts chosen at random by the elections official . . .

The elections official shall use either a random number generator or other method specified in regulations . . .

The official conducting the election shall include a report on the results of the 1 percent manual tally in the certification of the official canvass of the vote. This report shall identify any discrepancies between the machine count and the manual tally and a description of how each of these discrepancies was resolved . . .

NJ S507 [1R] (Gill)

[officials] shall conduct random hand counts of the voter-verified paper records in at least two percent of the election districts where elections are held for federal or State office . . .

Any procedure designed, adopted, and implemented by the audit team shall be implemented to ensure with at least 99% statistical power that for each federal, gubernatorial or other Statewide election held in the State, a 100% manual recount of the voter-verifiable paper records would not alter the electoral outcome reported by the audit . . .

[procedures] shall be based upon scientifically reasonable assumptions . . . including but not limited to: the possibility that within any election district up to 20% of the total votes cast may have been counted for a candidate or ballot position other than the one intended by the voters[.]

Say what?

Others

Oregon and New Mexico have audit laws that allow the sample (of races and/or ballots) to be selected before the election.

Maryland's pending legislation has elaborate tables of sample sizes.

Rep. Rush Holt has proposed federal legislation that has tiered sampling fractions, depending on the margin—but no requirement for followup if errors are found.

Legislation should enunciate *principles*, not *detailed methods*.

Methods should be left to regulation, so that they can be improved, fixed, etc.

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 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.

Risk-Limiting Audits are Easy

Toss p -coin. If coin lands heads, count the whole race.

Risk = $1 - p$.

Simple & understandable, but not efficient: Counts too many ballots when outcome is right.

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.

Hard to convince community that need big chance of a full count when the outcome is wrong, else never fix wrong outcomes.

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.

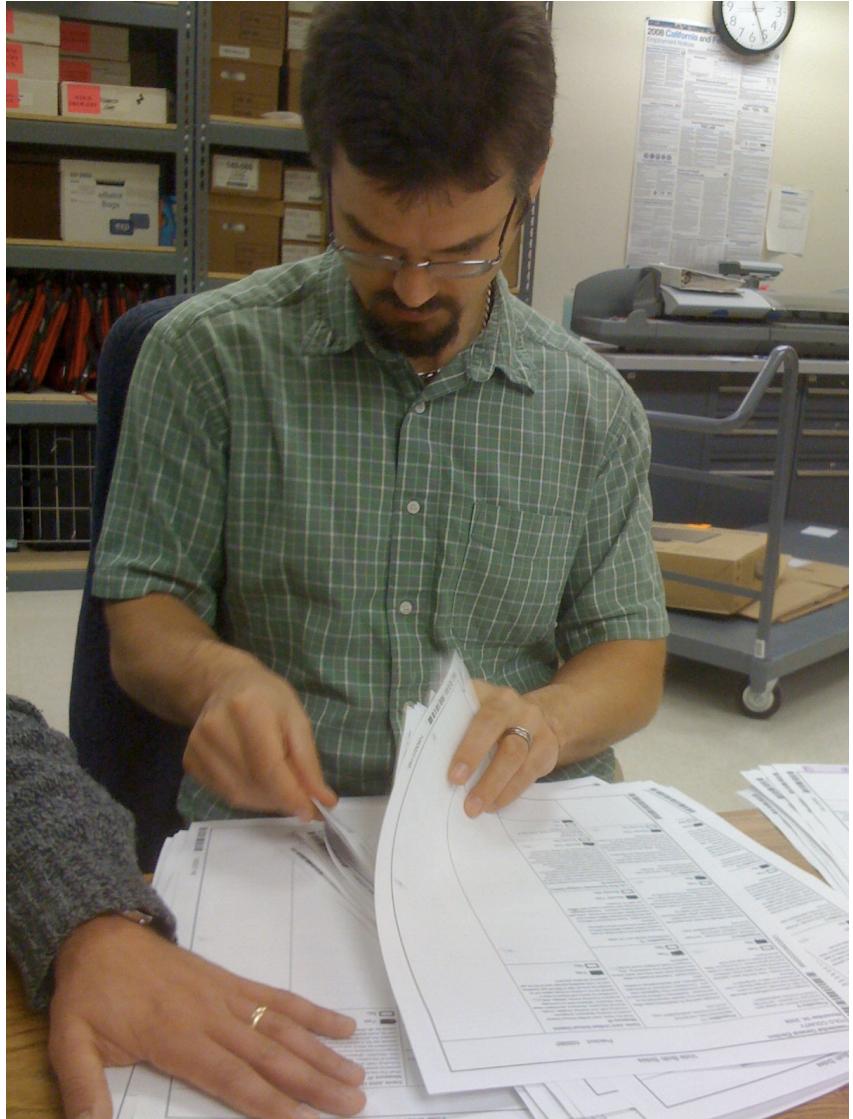
Essential to select batches at random, *after* the results are posted. (Can supplement with “targeted” samples.)

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.

2008 Yolo County, CA Measure W Audit





SIGN CERTIFICATE ON FRONT COVER

SIGN CERTIFICATE ON FRONT COVER

Vote Both Sides

**Presidential General Election
YOLO COUNTY
November 04, 2008**

Proposition 8
ELIMINATES RIGHT OF SAME-SEX COUPLES TO MARRY. INITIATIVE CONSTITUTIONAL AMENDMENT.

Changes California Constitution to eliminate the right of same-sex couples to marry. Provides that only marriage between a man and a woman is valid or recognized in California.

Fiscal Impact: Over next few years, potential revenue loss, mainly sales taxes, totaling in the tens of millions of dollars, to state and local governments. In the long run, likely little fiscal impact on state and local governments.

Yes
 No

Proposition 9
CRIMINAL JUSTICE SYSTEM, VICTIMS' RIGHTS, PAROLE INITIATIVE CONSTITUTIONAL AMENDMENT AND STATUTE.

Requires notification to victim and opportunity for input during phases of criminal justice process. Increases fines, penalties, sentencing and parole. Establishes victim safety as consideration for bail or parole.

Fiscal Impact: Potential loss of state savings on prison operations and reduced county jail costs amounting to tens of millions of dollars annually. Potential net savings in the low tens of millions of dollars annually on parole procedures.

Yes
 No

Proposition 10
ALTERNATIVE FUEL VEHICLES AND RENEWABLE ENERGY BONDS. INITIATIVE STATUTE.

Authorizes \$5 billion in bonds paid from state's General Fund, to help consumers and others purchase certain vehicles, and to fund research in renewable energy and alternative fuel vehicles.

Fiscal Impact: State cost of about \$10 billion over 30 years to repay bonds. Increased state and local revenues, potentially totaling several tens of millions of dollars annually. Potential state administrative costs up to about \$10 million annually.

Bonds Yes
 Bonds No

Los Rios Community College District Measure M
"Shall the Los Rios Community College District be authorized to issue \$475 million in bonds at the lowest available interest rates to improve student academic performance by building classrooms, faculty and labs throughout the district including for teaching green technologies; nursing and healthcare programs; architecture, engineering and construction management; computer sciences; early childhood development; and fire and police public safety programs at the American River, Cosumnes River, El Dorado, Folsom, and Sacramento City College campuses?"

Bonds Yes
 Bonds No

City of Davis Measure N
"Shall the Proposed Charter of the City of Davis Be Adopted?"

Yes
 No

1063031149

Vote Both Sides

Vote Both Sides

**Presidential General Election
YOLO COUNTY
November 04, 2008**

Precinct 100063

Proposition 8
ELIMINATES RIGHT OF SAME-SEX COUPLES TO MARRY. INITIATIVE CONSTITUTIONAL AMENDMENT.

Changes California Constitution to eliminate the right of same-sex couples to marry. Provides that only marriage between a man and a woman is valid or recognized in California.

Fiscal Impact: Over next few years, potential revenue loss, mainly sales taxes, totaling in the several tens of millions of dollars, to state and local governments. In the long run, likely little fiscal impact on state and local governments.

Yes
 No

Proposition 11
REDISTRICTING INITIATIVE CONSTITUTIONAL AMENDMENT AND STATUTE.

Changes authority for establishing state office boundaries from elected representatives to commission. Establishes multilevel process to select commissioners from registered voter pool. Commission comprised of Democrats, Republicans, and representatives of neither party.

Fiscal Impact: Potential increase in state redistricting costs once every ten years due to two entities performing redistricting. Any increase in costs probably would not be significant.

Yes
 No

Proposition 9
CRIMINAL JUSTICE SYSTEM, VICTIMS' RIGHTS, PAROLE INITIATIVE CONSTITUTIONAL AMENDMENT AND STATUTE.

Requires notification to victim and opportunity for input during phases of criminal justice process, including bail, pleas, sentencing and parole. Establishes victim safety as consideration for bail or parole.

Fiscal Impact: Potential loss of state savings on prison operations and increased county jail costs amounting to hundreds of millions of dollars annually. Potential net savings in the low tens of millions of dollars annually on parole procedures.

Yes
 No

Proposition 12
VETERANS' BOND ACT OF 2008.

This act provides for a bond issue of nine hundred million dollars (\$900,000,000) to provide farm and home aid for California veterans.

Fiscal Impact: Costs of about \$1.8 billion to pay off both the principal (\$900 million) and interest (\$556 million) on the bonds; costs paid by participating veterans. Average payment for principal and interest of about \$59 million per year for 30 years.

Bonds Yes
 Bonds No

Davis Joint Unified School District Measure W
"Shall the Davis Joint Unified School District preserve existing classroom programs including math and science, English, music, physical education, librarians, secondary class size reduction, athletics and co-curricular programs including drama, debate, and journalism by being authorized to levy a special tax for a period of three years not to exceed the annual rate of \$50.00 per dwelling unit for multi-dwelling parcels and \$120.00 per parcel for all other parcels?"

Fiscal Impact: Over next few years, potential revenue loss, mainly sales taxes, totaling in the several tens of millions of dollars, to state and local governments. In the long run, likely little fiscal impact on state and local governments.

Davis Joint Unified School District Measure W
"Shall the Davis Joint Unified School District preserve existing classroom programs including math and science, English, music, physical education, librarians, secondary class size reduction, athletics and co-curricular programs including drama, debate, and journalism by being authorized to levy a special tax for a period of three years not to exceed the annual rate of \$50.00 per dwelling unit for multi-dwelling parcels and \$120.00 per parcel for all other parcels?"

Yes
 No

Proposition 13
VETERANS' BOND ACT OF 2008.

This act provides for a bond issue of nine hundred million dollars (\$900,000,000) to provide farm and home aid for California veterans.

Fiscal Impact: Costs of about \$1.8 billion to pay off both the principal (\$900 million) and interest (\$556 million) on the bonds; costs paid by participating veterans. Average payment for principal and interest of about \$59 million per year for 30 years.

Bonds Yes
 Bonds No

Los Rios Community College District Measure M
"Shall the Los Rios Community College District be authorized to issue \$475 million in bonds at the lowest available interest rates to improve student academic performance by building classrooms, faculty and labs throughout the district including for teaching green technologies, nursing and healthcare programs; architecture, engineering and construction management; computer sciences; early childhood development; and fire and police public safety programs at the American River, Cosumnes River, El Dorado, Folsom, and Sacramento City College campuses?"

Fiscal Impact: State cost of about \$10 billion over 30 years to repay bonds. Increased state and local revenues, potentially totaling several tens of millions of dollars through 2019. Potential state administrative costs up to about \$10 million annually.

Bonds Yes
 Bonds No

City of Davis Measure N
"Shall the Proposed Charter of the City of Davis Be Adopted?"

Yes
 No

1063031149

Vote Both Sides

VOTE DATES

Precinct 100063

Da Joint Unified School District
Measure W
The Davis Joint Unified School District
proposes to expand existing classroom programs
including math and science, English, music,
and education, librarians, secondary class
size reduction, athletics and co-curricular
programs including drama, debate, and
jazz. It will do this by being authorized to levy a
new tax for a period of three years not to
exceed the annual rate of \$50.00 per dwelling
on multi-dwelling parcels and \$120.00 per
dwelling for all other parcels?

Yes

No

others
native
million
nd state
several
o about

technologies; nursing and medical
programs; architecture, engineering and
construction management; computer sciences;
early childhood development; and fire and
police public safety programs at the American
River, Cosumnes River, El Dorado, Folsom,
and Sacramento City College campuses?"

Bonds Yes

Bonds No

~~City of Davis
Measure N
Shall the Proposed Charter of the City of
Davis Be Adopted?~~

Yes

No

Selecting batches at random

Software pseudo-random number generators:
not transparent, hackable.

One ticket per precinct:
hard to verify; hard to mix (Vietnam draft).

2008 Oregon law uses one ticket per precinct; allows
selections before publishing election results.

10-sided dice (Marin County) [\[Roll 1\]](#) [\[Roll 2\]](#)

Ping-pong balls (Alameda County) [\[Static\]](#) [\[Tumbling\]](#)

Alameda County: 1204 precincts. Pick 1s digit, 10s, 100s.
If result is between 205 and 999, stop.
Else, remove 2–9 & pick 1000s digit from {0, 1}.

Unintended consequences?

Logistic tradeoff

Errors and time for dice rolling, look-up tables (especially for PPS sampling).

Hybrid selection

November 2008 Marin & Santa Cruz audits)

Roll 10-sided dice to get a 6-digit seed.

Use “good” open source PRNG to generate a sequence of numbers from the seed in a reproducible way.
(Used the Mersenne Twister implemented in R)

Notation

1 contest at a time (for now), vote for $\leq f$ of K candidates.

Results subtotalled in N auditable batches, $p = 1, \dots, N$.

a_{kp} : actual votes for candidate k in batch p .

v_{kp} : reported votes for candidate k in batch p .

$A_k \equiv \sum_{p=1}^N a_{kp}$; $V_k \equiv \sum_{p=1}^N v_{kp}$.

$V_{w\ell} \equiv V_w - V_\ell$: margin of w over ℓ

\mathcal{W} : the f apparent winners; \mathcal{L} : the $K - f$ apparent losers.

If $w \in \mathcal{W}$ and $\ell \in \mathcal{L}$, then $V_{w\ell} > 0$.

Summarizing error

Candidate w really beat candidate ℓ if $A_w - A_\ell > 0$.

Relative overstatement of the margin between $w \in \mathcal{W}$ and $\ell \in \mathcal{L}$ in batch p :

$$e_{w\ell p} = \frac{v_{wp} - v_{\ell p} - (a_{wp} - a_{\ell p})}{V_{w\ell}}.$$

The outcome of the race is correct if $\forall w \in \mathcal{W}, \ell \in \mathcal{L}$,

$$\sum_{p=1}^N e_{w\ell p} < 1.$$

Apparent winner(s) are real winner(s) if no margin was overstated by 100% or more of *that* margin.

Maximum relative overstatement of pairwise margins

Let

$$e_p \equiv \max_{w \in \mathcal{W}, \ell \in \mathcal{L}} e_{w\ell p}.$$

Now

$$\max_{w \in \mathcal{W}, \ell \in \mathcal{L}} \sum_p e_{w\ell p} \leq \sum_p \max_{w \in \mathcal{W}, \ell \in \mathcal{L}} e_{w\ell p} = \sum_p e_p.$$

Apparent outcome is right if

$$E \equiv \sum_{p=1}^N e_p < 1.$$

Can have slack, but then might be good idea to hand count anyway.

Testing $E \geq 1$

How strong is the evidence that $E < 1$? Maximum P -value.

Need upper bounds on $\{e_p\}$; otherwise P -value for $E \geq 1$ large unless we audit most batches.

b_p : bound on valid votes for any candidate in batch p .*

Reported margin vs. all b_p votes really for ℓ :

$$\begin{aligned} e_{w\ell p} &\leq \frac{v_{wp} - v_{\ell p} - (0 - b_p)}{V_{w\ell}} \\ &= \frac{b_p + v_{wp} - v_{\ell p}}{V_{w\ell}}. \end{aligned}$$

*E.g., from voter registrations, accounting of ballots, pollbook signatures.

A priori batch error bounds

Define

$$u_p \equiv \max_{w \in \mathcal{W}, \ell \in \mathcal{L}} \frac{b_p + v_{wp} - v_{\ell p}}{V_{w\ell}}.$$

Then

$$e_p \leq u_p.$$

Surprisingly controversial among EI advocates.

Less controversial among elections officials. Hmm.

Extends to simultaneous audits of several races (MARROP); controls FWER with cost comparable to controlling PCER.

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.

PPEB sampling

total error bound: $U \equiv \sum_{p=1}^N u_p.$

“taint” of batch p : $t_p \equiv e_p/u_p \leq 1.$

Draw n times iid, chance u_p/U of drawing batch p .

$T_j = t_p$ if batch p is selected in draw j

$$\mathbb{E}T_j \equiv \sum_{p=1}^N t_p \frac{u_p}{U} = \frac{E}{U}.$$

Outcome must be right if $\mathbb{E}T_j < 1/U$.

Test whether mean of upper-bounded r.v. $< 1/U$ from iid sample.

Martingale: X_1, X_2, \dots such that $\mathbb{E}|X_j| < \infty$ and

$$\mathbb{E}(X_{j+1}|X_1, \dots, X_j) = X_j \text{ (a.s.)}.$$

If X_1, X_2, \dots is a Martingale and $x > 0$,

$$\mathbb{P}\left(\max_{1 \leq j \leq k} X_j > x\right) \leq \mathbb{E}|X_k|/x.$$

Will use this to get a P -value for the hypothesis $E \geq 1/U$.

IID rvs to martingales

Suppose $\{Y_j\}$ iid, $\mathbb{P}\{Y_j \geq 0\} = 1$; $\mathbb{E}Y_j = \mathbb{E}|Y_j| = \mu < \infty$.

Let

$$X_j \equiv \prod_{i \leq j} Y_i / \mu.$$

Note $X_{j+1} = X_j \cdot Y_{j+1} / \mu$, $\mathbb{E}X_j = 1$, and

$$\mathbb{E}(X_{j+1} | X_1, \dots, X_j) = X_j \mathbb{E}Y_{j+1} / \mu = X_j.$$

So, X_1, X_2, \dots is a Martingale.

Kaplan-Markov-martingale P -value

Substuting definition of X_j and letting $x = 1/P$:

$$\mathbb{P} \left(\max_{1 \leq j \leq k} \prod_{i=1}^j Y_i / \mu > 1/P \right) \leq P$$

Want P -value for $\mathbb{E}T_j \geq 1/U$, where $\mathbb{P}\{T_j \leq 1\} = 1$.

Let $Y_j \equiv 1 - T_j$. Gives P -value

$$\min_{j \leq k} \prod_{i=1}^j \frac{1 - 1/U}{1 - T_i}.$$

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).

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?

Simultaneous audits: MARROP

N batches of ballots cover R races.

Not every race is on every ballot.

Voters in race r may vote for up to f_r of the K_r “candidates” in race r .

\mathcal{W}_r : the f_r apparent winners of race r .

\mathcal{L}_r : apparent losers of race r .

a_{kp} : actual vote for k in batch p

v_{kp} : reported vote for k in batch p .

Simultaneous audits: MARROP cont'd.

$$A_k \equiv \sum_{p=1}^N a_{kp}; V_k \equiv \sum_{p=1}^N v_{kp}.$$

If candidates w and ℓ are contestants in the same race r , the reported margin of apparent winner $w \in \mathcal{W}_r$ over apparent loser $\ell \in \mathcal{L}_r$ is

$$V_{w\ell} \equiv V_w - V_\ell > 0.$$

$$A_{w\ell} \equiv A_w - A_\ell.$$

MARROP: sufficient condition

All R apparent outcomes are right if

$$\min_{r \in \{1, \dots, R\}} \min_{w \in \mathcal{W}_r, \ell \in \mathcal{L}_r} A_{w\ell} > 0.$$

$$e_{pw\ell} \equiv \begin{cases} \frac{(v_{wp} - v_{\ell p}) - (a_{wp} - a_{\ell p})}{V_{w\ell}}, & \text{if ballots in batch } p \text{ contain race } r \\ 0, & \text{otherwise.} \end{cases}$$

If any outcome is wrong,

$\exists r \in \{1, \dots, R\}$, $w \in \mathcal{W}_r$ and $\ell \in \mathcal{L}_r$ s.t.

$$\sum_{p=1}^N e_{pw\ell} \geq 1.$$

MARROP, cont'd

$$e_p \equiv \max_{r \in \{1, \dots, R\}} \max_{w \in \mathcal{W}_r, \ell \in \mathcal{L}_r} e_{pwl}.$$

$$\max_{r \in \{1, \dots, R\}} \max_{w \in \mathcal{W}_r, \ell \in \mathcal{L}_r} \sum_{p=1}^N e_{pwl} \leq \sum_{p=1}^N e_p \equiv E.$$

E is maximum across-race relative overstatement of pairwise margins (MARROP).

If $E < 1$, all R race outcomes are right.

Level- α test of $E \geq 1$ tests all R outcomes at FWER $\leq \alpha$.

Error bounds for MARROP

Valid ballots cast in batch p for race $r \leq b_{rp}$.

$a_{wp} \geq 0$ and $a_{\ell p} \leq b_{rp}$, if ℓ is a candidate in race r .

Hence, $e_{pwl} \leq (v_{wp} - v_{\ell p} + b_{rp})/V_{w\ell}$, and so

$$e_p \leq \max_{r \in \{1, \dots, R\}} \max_{w \in \mathcal{W}_r, \ell \in \mathcal{L}_r} \frac{v_{wp} - v_{\ell p} + b_{rp}}{V_{w\ell}} \equiv u_p.$$

u_p bounds relative overstatement of *any* margin in batch p .

Simultaneous audit: hypothetical

Race	prcts	batches	ballots	votes overall			IP	winner	loser	VBM	
				winner	loser	margin				winner	loser
A	200	400	120,000	60,000	54,000	6,000	200	200	180	100	90
B	100	200	60,000	30,000	24,000	6,000	200	200	160	100	80
C	60	120	36,000	18,000	12,600	5,400	200	200	140	100	70

Race A: entire jurisdiction, 200 precincts.

Race B: 100 of the precincts.

Race C: 60 of the precincts; 30 in race B.

Each precinct is 2 batches: 400 ballots cast in-precinct (IP), 200 cast by mail (VBM).

Some undervotes and invalid ballots.

Simultaneous audit workload: example

	U	n	expected batches	FWER		n	expected batches	PCER	
				expected ballots	expected votes			expected ballots	expected votes
A	21.00	52	48.49	16,074.23	16,074.23	33	31.58	10,488.77	10,488.77
B	11.00	28	26.01	8,615.69	8,615.69	17	16.27	5,402.16	5,402.16
C	7.67	19	17.50	5,795.81	5,795.81	12	11.41	3,787.51	3,787.51
all			85.13	28,038.26	30,485.73		56.38	18,649.98	19,678.44
MRP	22.72	36	34.30	11,387.29	20,617.68				

Independent vs. simultaneous audits controlling FWER and PCER.

Threshold taint for “escalation” 0.04.

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.

References

“Best Practices” for audits <http://electionaudits.org/principles>

Ash, A., S. Pierson and P.B. Stark, 2009. Thinking outside the urn: Statisticians make their marks on U.S. Ballots. *Amstat News*, 384. 37–40. http://www.amstat.org/outreach/pdfs/SP_ANJun09.pdf

Jefferson, D., K. Alexander, E. Ginnold, A. Lehmkuhl, K. Midstokke and P.B. Stark, 2007. *Post-Election Audit Standards Working Group: Report to California Secretary of State Debra Bowen*. http://www.sos.ca.gov/elections/peas/final_peaswg_report.pdf

Miratrix, L.W. and Stark, P.B., 2009. Election Audits With a Trinomial Bound. *IEEE Transactions on Information Forensics and Security*, accepted. <http://statistics.berkeley.edu/~stark/Preprints/trinomial09.pdf>

Stark, P.B., 2009. Auditing a collection of races simultaneously. <http://arxiv.org/abs/0905.1422v1>

Stark, P.B., 2009. Risk-limiting audits: P -values from common probability inequalities. <http://statistics.berkeley.edu/~stark/Preprints/pvalues09.pdf>

Stark, P.B., 2009. CAST: Canvass audits by sampling and testing. *IEEE Transactions on Information Forensics and Security*, accepted. <http://statistics.berkeley.edu/~stark/Preprints/cast09.pdf>

References, cont'd

Stark, P.B., 2008. A sharper discrepancy measure for post-election audits, *Annals of Applied Statistics*, 2, 982–985. <http://arxiv.org/abs/0811.1697>

Stark, P.B., 2008. Conservative Statistical Post-Election Audits, *Annals of Applied Statistics*, 2, 550–581. <http://arxiv.org/abs/0807.4005>

More: <http://statistics.berkeley.edu/~stark/Vote>