This month's guest columnists, Arlene Ash and Philip B. Stark, have been very active in election integrity work. I'm grateful for their instrumental role in shaping and guiding the ASA science policy work in this area, particularly when it comes to interacting with election officials. I especially appreciate their willingness to write this month's column, which provides an update on this rapidly developing area and suggests a road map for further activity.

~Steve Pierson, ASA Director of Science Policy, *pierson@amstat.org*

Thinking Outside the Urn

Statisticians make their marks on US ballots

Arlene Ash, Boston University School of Medicine, and Philip B. Stark, University of California at Berkeley

Free, fair, and accurate elections are the cornerstone of democracy, yet troubling failures of equipment, software, and procedures continue in the United States. For example, we have all heard that some ballots "went missing" in Minnesota's 2008 famously close Senate race recount.

Elsewhere in 2008: Florida lost more than 3,400 ballots initially counted on Election Day; a bug in commercial election software dropped 197 ballots from the totals in Humboldt County, California; the same voting machines used in 34 states lost votes in Ohio; several states reported vote-flipping on electronic voting machines; and thousands of phantom votes were reported in Washington, DC, inflating the apparent number of votes to 4,759 from a group of 326 actual votes.

U.S. elections are complicated, involving at least the following:

— Registering voters and maintaining up-to-date registries

— Certifying (the usually many) candidates and measures to be on each ballot

— Designing, testing, and printing (or programming) ballots, with up to hundreds of variants by language and jurisdiction

— Designing, printing, and mailing election information, such as sample ballots

— Managing early, absentee, and Election Day voting

— Educating poll workers and the public

— Allocating staff and equipment for polling places

— Preparing and delivering poll books and other voting materials and equipment to election sites

— Maintaining a secure chain of custody of all relevant materials while developing preliminary, first-reported, and certified vote counts

— Reporting Election Day results

— Performing a full canvass and reporting final election results

— Conducting routine audits, and sometimes full recounts of election tallies

— Quality checking and public reporting

Statisticians have much to offer election administration, including specifying data collection and reporting requirements, monitoring the integrity of data and data processing, designing and computing with large databases, conducting routine quality-control tests, and—ultimately—assessing the uncertainty in election outcomes.

The Help America Vote Act of 2002, responding to problems in the 2000 presidential election, encouraged jurisdictions to retire punchcard voting. Many switched to electronic voting. Unfortunately, most of the new "direct-recording electronic" (DRE) voting machines produced no paper trail, so the only possible checks on machine totals rely on the accuracy of the same machines being tested.

Independent verification is the keystone of good electoral auditing. Hand counting paper records is a good check on machine tallies, as the two methods tend to err for different reasons and in different ways. Jurisdictions that scan voter-marked paper ballots or use DREs that produce a voter-verifiable paper audit trail

PhD student Luke Miratrix and Philip Stark (right) audit the votes for Measure W in Yolo County, California, in November of 2008. Miratrix works with Stark developing risk-limiting methods.

(VVPAT) can independently check the accuracy of machine subtotals, typically subtotals by precinct. Vote-tabulation audits (also called post-election audits) that compare hand and machine counts of a random sample of subtotals can help ensure incorrect outcomes are caught and corrected. Indeed, post-election audits uncovered many of the problems cited above.

Statisticians helped draft *Principles and Best Practices for Post-Election Audits* (see *http://electionaudits.org/principles*), and the American Statistical Association Board of Directors endorsed its statistical content. The document states the following goals of vote-tabulation audits:

— To deter fraud

— To promote public confidence in elections

— To find error, whether accidental or intentional

— To verify that vote-tallying machines are functioning properly

—To provide for continuous improvement in the conduct of elections

Statisticians can help. For example, they can do the following:

— Characterize the distribution and nature of vote-counting errors for different voting technologies and different methods of hand counting

— Develop and improve the efficiency of practical methods to control the risk that an incorrect outcome will go undiscovered

— Estimate the cost and logistical requirements of various approaches to auditing and the trade-offs among rigor, risk, time, and cost

— Inform policy when there are competing risks and benefits—for example, between the risk of disenfranchisement and that of vote fraud, between cost and waiting time at the polls, between cost and accuracy of vote-count technology, between security and ease of use

— Identify processes and technologies that can improve accuracy or enable less painful compromises

— Develop a knowledge base

The role of statisticians was discussed by David Marker, John Gardenier, and Arlene Ash in the June 2007 issue of *Amstat News*. Here, we review recent progress, propose a road map for further progress, and suggest ways for statisticians to contribute.

A Challenging Statistical Problem

At first glance, election auditing appears to be a straightforward sampling problem. In fact, audits raise questions about nonparametric tests for the mean of nonstandard distributions, exact inference from stratified samples, sampling with probability proportional to error bounds, multiple testing, sequential testing, and combinatorially complex games. Audit protocols must also meet constraints of cost, time, complexity, and transparency to diverse audiences—the public, election integrity advocates, legislators, and election officials. Because audits must work within the practical limitations of elections and the legislative process, statisticians should be prepared to collaborate across many disciplines: political science, computer science, the social and behavioral sciences (usability studies and survey research), forensics, and "usable security"—security that is adequate to the task without impairing usability. Statisticians must "think outside the urn."

Methodological research in election auditing is relevant to financial auditing, experimental design and analysis, and many other fields in which one seeks to draw inferences about a finite population that satisfies *a priori* bounds, but for which parametric approximations could be inaccurate. Preliminary results suggest that techniques recently developed for election auditing are more efficient than those currently used in financial auditing.

Risk-Limiting, Vote-Tabulation Audits

According to *Principles and Best Practices for Post-Election Audits*, a risk-limiting audit has a prespecified, minimum chance of a full manual count whenever the outcome of the election is wrong. By definition, an outcome is "wrong" if it disagrees with the outcome a full hand count would show. A risk-limiting audit is only as good as its audit trail. No trail, no audit. Only with a complete and accurate audit trail—ensured through a secure chain of custody—can any audit provide real assurance.

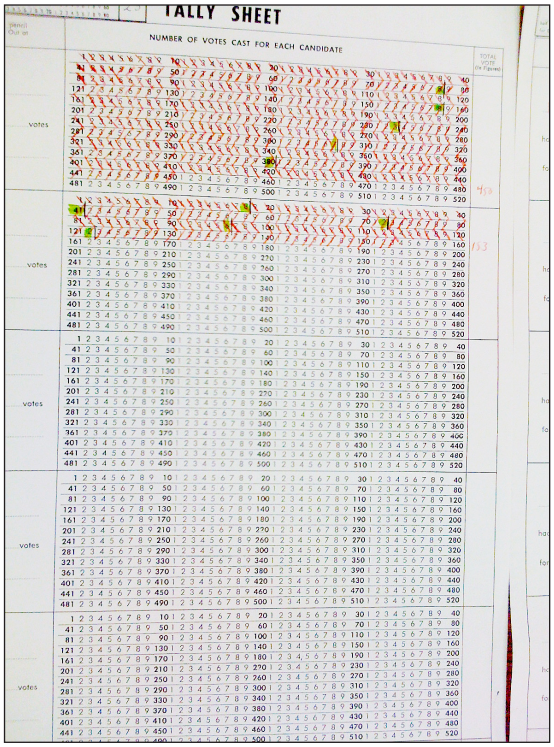
A risk-limiting audit can be couched as a hypothesis test. The "risk" is the chance that the audit stops before a full hand count when the outcome is wrong. To control that risk statistically, it makes sense to take the null hypothesis to be that a full hand count would contradict the apparent outcome. A Type I error occurs if the audit stops short of a full hand count when the outcome is wrong. The goal is to efficiently control the Type I error rate—that is, to count as few ballots as possible while limiting the risk of stopping too early when a full count would change the outcome.

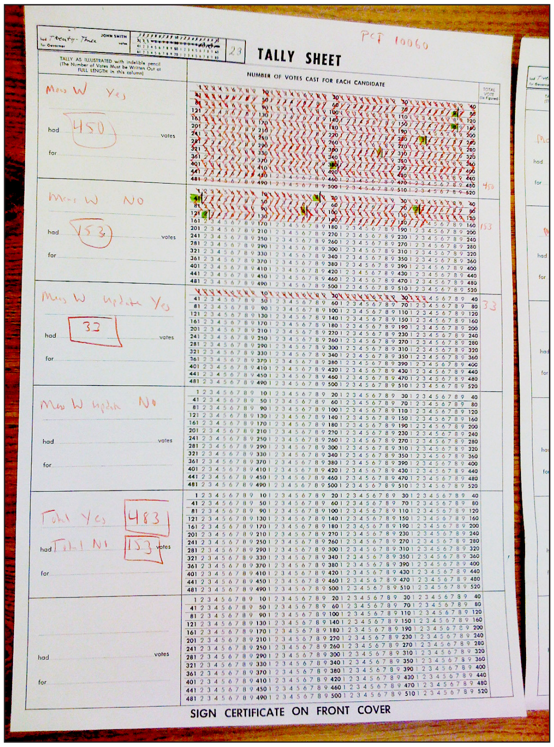
Early statistical work on vote-tabulation audits includes SAFE (statistically accurate, fair, and efficient), which gave the number of audit batches to select by simple random sampling as a function of the margin of victory, precinct sizes, and a parameter specifying the maximum plausible level of error in any precinct. SAFE ensures that if the outcome is wrong, there is a large chance the audit will uncover at least one precinct with an error, even if the error is concealed in as few precincts as possible. SAFE is not a fully fleshed-out, risk-limiting protocol; it provides no guidance for what to do when the audit finds discrepancies, which are virtually inevitable even when the outcome is right.

Philip B. Stark of the University of California at Berkeley and his colleagues have developed several methods for risk-limiting audits. In collaboration with county election officials, they conducted four risk-limiting audits in California in 2008: Marin County (a small measure requiring a supermajority and a county-wide measure), Santa Cruz County (County Supervisor, District 1), and Yolo County (a bond measure). The audits limited the risk—the Type I error rate—to at most 25%. All four confirmed the provisional outcomes without requiring a full hand count.

Risk-limiting audits are typically performed in stages. Each stage involves drawing a probability sample of batches, comparing the preliminary results with hand counts of the audit trail for those batches, and calculating a *p*-value for the hypothesis that the outcome of the election is incorrect. The calculation involves the reported votes by batch for the entire race, the observed discrepancies between the preliminary counts and the hand counts, the sampling design (including sample sizes, stratification, and the protocol for advancing from one stage to the next), and the desired limit on the risk. If the discrepancies in a large enough sample are sufficiently small, there is strong evidence that the outcome is correct, so the audit can stop; otherwise, the audit progresses to the next stage—to collect more evidence, perhaps eventually requiring a full hand count.

This sample-and-test strategy requires a transparent, public, trustworthy method of generating random samples, public counting, published procedures, and calculations and theorems that few elections officials or others have the background to understand. Perhaps the biggest open problems for methodologists in post-election audits now concern logistical barriers to adoption: developing simpler, more efficient, and more usable methods; developing efficient ways to audit dozens of contests simultaneously; and developing data 'plumbing,' software tools and turn-key solutions that make the methods accessible to elections officials and transparent to the public. Another important task is to estimate the costs and specific resources needed to support credible audits.





Tally sheets for auditing Measure W, Yolo County, California, November 2008

Political Climate for Post-Election Audits

Some states have enacted vote tabulation audit laws that require comparing machine tallies with a hand count of a sample of ballot batches. Other states are considering such laws. Federal legislation requiring audits has been introduced several times. The laws and bills vary in their goals and effect on election integrity. Some have no enunciated goal, but require that the audit trail receive some scrutiny. Some seek 'quality control,' in a loose sense. Some seek to ensure that electoral outcomes are correct. States draw their initial audit samples differently; few require additional sampling if the initial audit finds errors. Some use a fixed number of batches per county across all contests, or a number that increases with population. Some require auditing a fixed percentage of batches or votes. Some have tiered designs, whereby one of three percentages of ballots are audited, depending on the apparent margin of victory.

A few bills now being considered seek to limit the risk of certifying an incorrect electoral outcome. Although no current or pending legislation actually requires risk-limiting audits, we commend states for moving toward statistically motivated and justified audits. A common flaw of many bills is that they attempt to legislate details of sampling and statistical calculations, rather than enunciating principles (e.g., "the procedure shall have at least a 90% chance of requiring a full hand count whenever that count would show a different outcome") and leaving implementation details to regulation. The push for overly detailed prescriptions seems to arise from mutual suspicion among the stakeholders, including local, county, and state election officials, legislators, and election integrity advocates.

Vote-Tabulation Audit Road Map

From our experience in talking with election officials, auditing elections, and working with election integrity advocates, we see the following priorities:

Build the information infrastructure for good audits (i.e., data plumbing). Audits can only be done efficiently if data are promptly and publicly available in machine- and human-readable formats at the level of auditable batches of ballots. Surprisingly, this is rare. Commercial election management software seems to be the bottleneck. Adopting standard terminology and data formats would help states ask for useful data and make it easier for voting machine and software vendors to supply it. National or international standards should be considered.

Legislate auditing principles, not implementation details. Methods for risk-limiting audits are evolving quickly. Overly prescriptive laws prevent jurisdictions from using the most efficient, reliable, and transparent procedures available. Legislators might be unwilling to revisit election audits repeatedly, so jurisdictions that legislate methods can be stuck with poor methods indefinitely.

Collect election and audit data and make it accessible. A central repository for data—including election results at the precinct level or below, voting system failures, errors uncovered by audits, etc.—would be extremely valuable. Metadata should include details about the voting technology, the vote-counting technology, the audit, the hand-counting procedures, and so on.

Help the public and elections officials understand basic statistical principles that apply to audits. For instance, it is a common misconception that correcting the error found in a random sample of batches corrects the error in the entire contest—the idea that each item in a random sample stands for a larger number of items in the population is not universally appreciated. Similarly, jurisdictions have passed laws requiring that the audit sample be selected before the preliminary results are published. Statisticians can help explain why that undermines the security the audit would otherwise give.

Statisticians possess the expertise to discover and characterize the nature, frequency, and sources of problems in our elections and reduce the risk that electoral outcomes do not truly reflect the will of the voters. The ASA as an organization and several of its members are working to develop and disseminate model language, techniques, and best practices for our elections. Election officials have begun to come to us for guidance. We welcome your help to meet the important challenge of measuring, improving, and ensuring the accuracy and integrity of U.S. elections. ■

ASA Science Policy Actions

- The ASA sends a letter regarding effects of IT consolidation on the IRS Statistics of Income (SOI) Division and a letter regarding the next SOI director
- The 50 slots for the ASA congressional visits filled in just two weeks, with representatives being 25% students and coming from 25 states
- The ASA signs a letter to congressional leaders thanking them for their efforts to pass the Edward M. Kennedy Serve America Act (H.R. 1388)

Further information can be found at *www.amstat.org/outreach/scipolicyletters.cfm*.