Percentage-Based versus Statistical-Power-Based Vote Tabulation Audits

John McCarthy, Howard Stanislevic, Mark Lindeman, Arlene S. Ash, Vittorio Addona, and Mary Batcher

Several pending federal and state electoral-integrity bills specify hand audits of 1% to 10% of all precincts. However, percentage-based audits are usually inefficient, because they require large samples for large jurisdictions, even though the sample needed to achieve good accuracy is much more affected by the closeness of the contest than population size. Percentage-based audits can also be ineffective, since close contests may require auditing a large fraction of the total to provide confidence in the outcome. We present a plausible statistical framework that we have used in advising state and local election officials and legislators. In recent federal elections, this audit model would have required approximately the same effort and resources as the less effective percentage-based audits now being considered.

KEY WORDS: Election audits; Election recounts; Electronic voting; Precinct sampling.

1. INTRODUCTION

Electronic vote tally miscounts arise for many reasons, including hardware malfunctions, unintentional programming errors, malicious tampering, or stray ballot marks that interfere with correct counting. Thus, Congress and several states are considering requiring audits to compare machine tabulations with hand counts of paper ballots in randomly chosen precincts. Audits should be highly effective in detecting mis-

counts large enough to alter election outcomes; and they should be efficient—no larger than necessary to confirm the winners. While financial and quality control audits set sample sizes that are very likely to detect errors large enough to cause harm, most proposed election auditing laws specify sampling fixed or tiered percentages of precincts. For example, Connecticut has just adopted a law (Public Act 07-194) requiring random audits of 10% of voting districts (precincts) in selected contests. We believe that the laws are written this way because most nonstatisticians have unrealistic fears about the inadequacy of small-percentage audits; because the authors have not measured the statistical effectiveness of percentage-based schemes in general; and because statisticians have—thus far—rarely been involved in drafting audit options for legislators. Statisticians of course know that we can measure the effectiveness of sampling strategies by their statistical power, which principally depends on the number of units sampled and the size of the effect to be detected. Thus, fixed-percentage audits are inefficient (too large) in the vast majority of contests, especially in statewide contests that involve many hundreds of precincts and that are not close; also, they are ineffective (too small) in the rare contests with small winning margins. However, most statisticians know little about election procedures and are not well-equipped to respond when asked "what percentage shall we put in the bill?" We hope that this article helps fill that gap.

Vote tabulation audits entail supervised hand-to-eye manual counts of all voter-verified paper ballots in a subset of precincts, randomly selected shortly after an election and before results are certified. We assume that a "hard copy" record of each voter's choice, one that was reviewable by the voter for accuracy before the vote was officially cast, is available for comparison with the electronic tally. In particular, this excludes Direct Recording Electronic (DRE) voting with no paper trail. This assumption is consistent with the draft standards of the Election Assistance Commission Technical Guidelines Development Committee (*http://www.eac.gov/vvsg*). For convenience we use the word "precinct" throughout, although the appropriate audit unit is the smallest cluster of votes that is separately tallied and reported. Thus, a batch of early votes cast at a central location is a "precinct." Or, if a precinct's votes combine tallies from two readily distinguished machines, then each machine could—and for efficiency should—be treated as a separate "precinct." From a legal perspective, a 100% audit may not equal a "recount" after a candidate disputes election results, since different procedures may apply. Whole-precinct audits are not designed to estimate the entire vote count directly,

but rather to independently confirm (or challenge) the accuracy of precinct-level electronic tallies.

Manual audits should also be done in precincts with obvious problems, such as machine failure, and for routine fraud deterrence and quality improvement monitoring, even in the absence of doubt about who won. Importantly, election officials and candidates should be empowered to choose additional precincts with apparent anomalies for auditing, just as financial and quality audits examine "high-interest" units as well as random samples. Comprehensive auditing should also examine many other parts of the electoral process, as outlined by Marker, Gardenier, and Ash (2007) and Norden et al. (2007). Another important question not pursued here is the need for mandatory follow up when an initial audit casts doubt upon an electoral outcome.

Our simplified framework assumes that the outcome of an audit is dichotomous: if the audit sample includes one or more miscounted precincts, additional action is taken; otherwise, the election is confirmed. Here, the audit's power equals the probability of sampling at least one miscounted precinct *whenever there are enough precincts with miscounts to have altered the outcome*. Statistically based audit protocols should seek to fairly and efficiently use resources to achieve a prespecified high power (99%, if feasible) in all elections—from statewide contests to those in a single Congressional district or county.

2. MODEL ASSUMPTIONS

We assume that every vote is cast in one and only one precinct and that precincts to be audited will be chosen randomly (with equal chance of selection) *after* an election has taken place and after the unofficial vote counts for each auditable unit are publicly reported. We also assume that net miscounts in more than 20% of the votes cast in a precinct would trigger a suspicion-based "targeted audit." This implies that a *random* audit, to be successful, only needs to detect at least one of a set of precincts, with shifts of at most 20% each, that together would change the electoral outcome. The *Within-Precinct Miscount*, or WPM, of 20% is a parameter that could be reset as experience accumulates. Saltman (1975) described WPM as the "maximum level of undetectability by observation." Setting an upper bound for the WPM allows us to determine the *minimum* number of precincts that must contain miscounts if an election outcome with a reported margin of victory were to be reversed. The hypergeometric distribution can then be used to calculate the sample size needed to achieve a specified probability that at least one miscounted precinct appears in the sample. (While the same approach can be used without using WPM, it typically requires substantially larger samples.) Our general findings do not depend on any particular value of WPM; the 20% figure serves as a reasonable baseline for comparative purposes.

When a contest has outcome-altering miscounts, an initial audit is successful only if it finds at least one miscounted precinct, since, if it does not, the wrong outcome would be confirmed. This initial audit does not have to determine what the outcome should have been, so long as it triggers further actions to make

that determination. Procedures for deciding when an audit that uncovers small discrepancies should still confirm an election outcome is the subject of ongoing research. See, for example, Stark (2007).

3. CALCULATION OF STATISTICAL POWER

Let:

- n = number of randomly selected precincts.

- N = total number of precincts.

- m = margin of victory in a particular election contest— that is, the difference between the number of votes for the winning candidate in a single seat contest (or the number of votes for the winning candidate with the smallest number of votes in a multiseat contest) and the number of votes for the runner up candidate, divided by the total number of ballots cast.

- B_{min} = minimum number of miscounted precincts (out of the N precincts) needed to overturn the election result for a particular contest.

- X = number of miscounted precincts in the sample of n.

Then we define:

$$\text{Power} = P(X > 0 | B_{min} \text{ miscounted precincts in the population}).$$

The value of B_{min}, the minimum number of miscounted precincts that could alter the outcome, depends upon the possible extent of miscounts in each precinct. Assuming equally sized precincts with a maximum WPM of 20% (a 40-point shift in the percentage margin within that precinct), a proportional winning margin of m could be overcome by switching votes in

$$B_{min} = \left\lceil \left(N \cdot \frac{m}{2 \text{ WPM}} \right) \right\rceil,$$

or

$$\lceil (N \cdot m/0.4) \rceil$$

precincts. For instance, if all precincts contain the same number of votes, a 10-percentage-point margin ($m = 0.1$) could be overcome by switching 20% of votes in at least $0.25N$ or 25% of all precincts.

In practice, precincts contain varying number of votes. It is instructive to consider how B_{min} (and therefore audit power) is affected by the alternative assumption that all the miscounts reside in the largest precincts. For instance, under the "largest precinct assumption," for a margin of 10 points, B_{min} will be the smallest number of large precincts that together contain at least 25% of all *votes*. This figure can be calculated directly if the distribution of precinct-level vote counts is available (from a preliminary report of precinct-level election returns), or it can be estimated based on the distribution of precinct-level votes in the previous election; a reference distribution (such as the Ohio CD-5 distribution shown in Figure 3); or using an heuristic approximation. (For one such heuristic, see McCarthy et al. (2007).) As we demonstrate in the following discussion, this assumption can markedly reduce estimates of audit power.
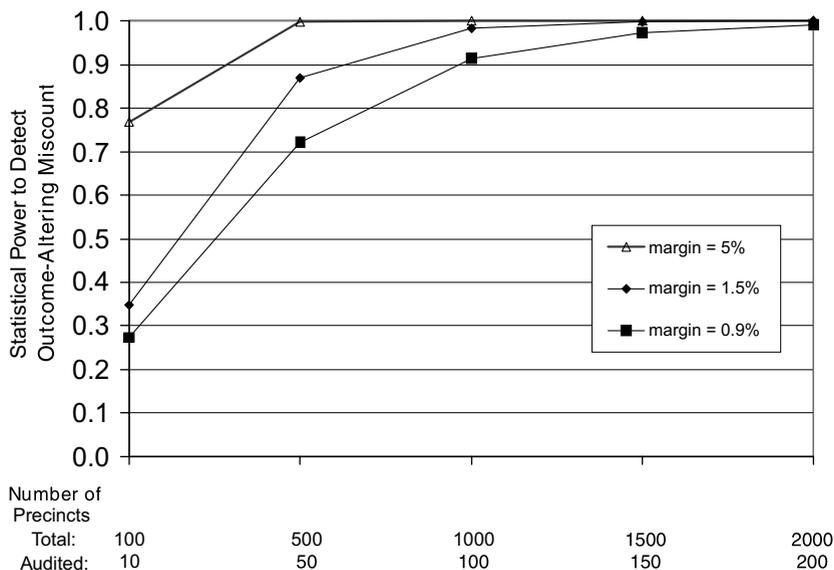
Figure 1. Statistical Power of 10% Audits for Districts: By Number of Precincts Audited and Margin of Victory. (Power = the probability of finding at least one miscounted precinct when the number of miscounted precincts equals the fewest average-sized precincts with 20% shifts needed to overturn the election.)

4. STATISTICAL POWER OF FIXED PERCENTAGE AUDITS

Figure 1 shows the power of a 10% audit for jurisdictions with varying numbers of precincts and winning margins. We assume here that all precincts contain the same number of votes or, equivalently, that the average number of votes for the precincts with miscounts is the same as for all precincts. Thus, for instance, to overcome a 5% margin, 20% (WPM) of votes in at least 12.5% of precincts must be switched from the reported winner to the reported loser. For context, Connecticut has 769 voting districts; New Hampshire's 1st Congressional District has only 114 reporting units (usually entire towns); West Virginia and Iowa have just under 2,000 precincts; and 28 states are off the scale in Figure 1, including five with over 10,000 precincts.

Clearly, 10% audits have limited power when electoral margins are close and/or when the total number of precincts is small. The fixed percentage approach also examines too many precincts when the election is not close and/or the number of precincts is large. For example, in California, a 10% state-wide audit would involve 2,200 precincts; this is an order of magnitude larger than needed, under an equal sized precinct assumption, even with a margin as small as 0.9%.

5. STATISTICAL POWER OF TIERED PERCENTAGE AUDITS

Tiered percentage audits specify a percentage of precincts to be audited, depending on the reported winning margin. For example:

1. audit 3% of precincts if the winning margin is 2% or more of the total votes cast;

2. audit 5% of precincts if the winning margin is at least 1% but less than 2%; and

3. audit 10% of precincts if the winning margin is less than 1% of the total votes cast.

This approach addresses the need to audit more when margins are narrower, but does not solve the fundamental problem with fixed percentage audits. Most congressional districts contain fewer than 500 precincts. In these congressional districts, and in other modest-sized legislative districts, even sampling 10% of precincts does not achieve 75% power when the winning margin is at most 1%. A tiered audit also allows precipitous drops in power at the tier thresholds, which appear as "saw-teeth" in Figure 2.

6. VARIATIONS IN PRECINCT SIZE

The power of an audit also can be reduced when precincts differ markedly in numbers of voters (Saltman 1975; Stanislevic 2006; Dopp and Stenger 2006; and Lobdill 2006), because fewer large miscounted precincts can alter enough votes to change the outcome. For reference, Figure 3 shows the vote distribution in the 640 precincts of Ohio's Fifth Congressional District (CD-5) in the 2004 general election, where votes per precinct ranged from 132 to 1637. Only about 8.6% (55) of Ohio CD-5's largest precincts are needed to encompass 15% of the vote. Applying Ohio CD-5's precinct size distribution to any number of precincts, we can estimate the power of audits in a "worst-case scenario" when miscounts occur in the smallest possible number of precincts (making it harder to find "at least one" in the audit sample). For instance, consider an election with 500 total precincts and a 6% margin, so that $B_{\min}$ equals the number of precincts containing at least 15% of the vote. A 3% audit sample is estimated to have 92% power if mis-
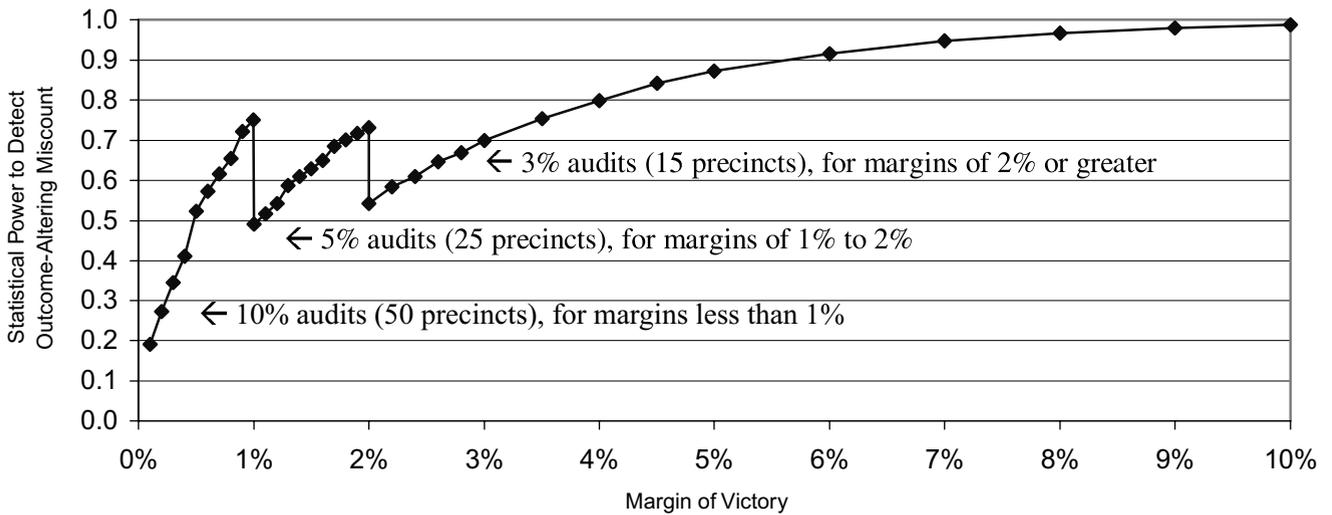
Figure 2. Statistical Power of Three-Tiered 3–5–10% Audits in a 500-Precinct Jurisdiction: By Margin of Victory. (Assumes the number of precincts with miscounts equals the minimum number needed to overturn the election if all miscounts are in average-sized precincts, each with a vote shift of 20%.)

counts are equally common among small and large precincts, but just 75% power if they occur in the largest precincts. (In the first case, $B_{\min} = 0.15(500) = 75$; in the second, $B_{\min} = 0.086(500) = 43$.)

7. STATISTICAL POWER-BASED AUDITS

Statisticians and a growing number of election experts have urged replacing fixed percentage audits with audits that employ a statistically grounded criterion of efficacy. Here we present a power-based audit which determines the number of precincts that must be sampled to achieve a specified power level for each election contest. In addition to the desired power, the sample size will depend on the reported victory margin, the value of WPM, and both the number and size-distribution of the precincts. Given the total number of precincts in a contest and the minimum number of miscounted precincts that could alter the outcome, the sample size for a given power can be obtained using the hypergeometric distribution or an appropriate approximation to simplify calculations so they can be readily implemented by election officials (Aslam, Popa, and Rivest 2007). In
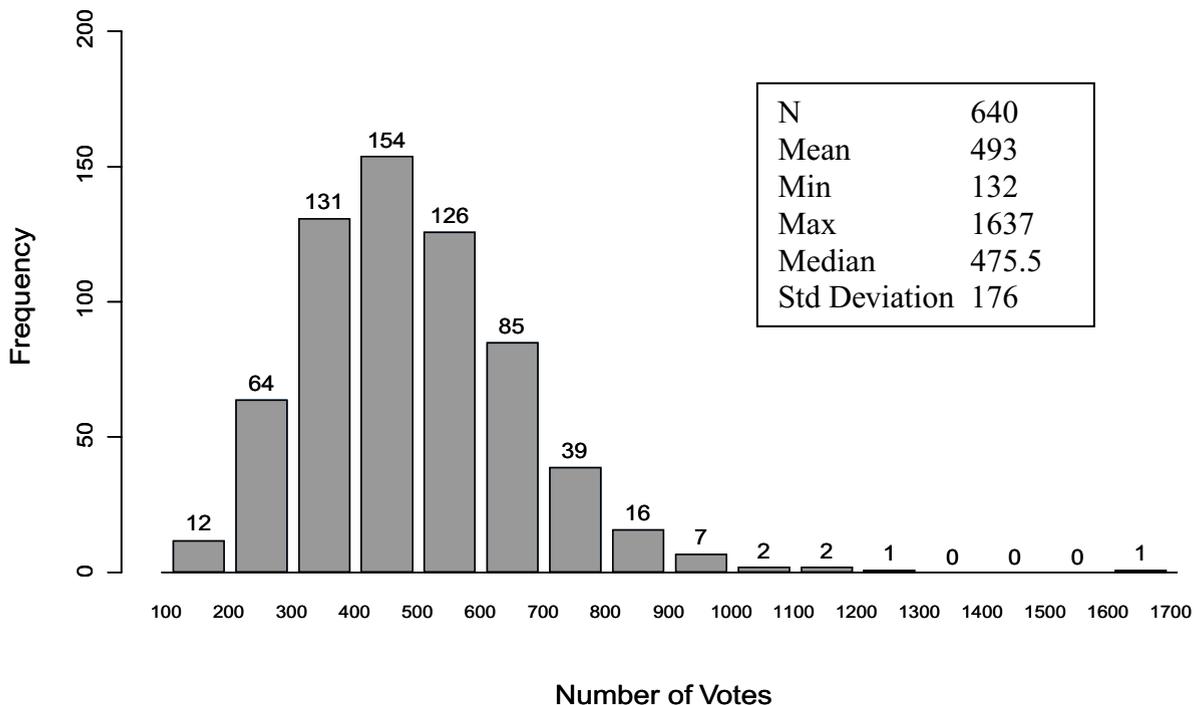

Figure 3. Distribution of Votes Counted in 2004 among the 640 Precincts of Ohio's Fifth Congressional District.

Table 1. Federal Elections (2002–2006) Achieving Various Levels of Power by Type of Audit

	Type of Audit					
	Tiered/Fixed Percentage				Power-Based	
	Tiered 3-5-10%	2%	3%	10%	99% power	95% power
Power of the Audit	*Number of elections (percent)*					
at least 99%	1152 (82.7%)	1089 (78.2%)	1152 (82.7%)	1292 (92.7%)	1393 (100%)	–
from 95% up to 99%	77 (5.5%)	98 (7.0%)	74 (5.3%)	31 (2.2%)	–	1393 (100%)
from 50% up to 95%	112 (8.0%)	137 (9.8%)	110 (7.9%)	51 (3.7%)	–	–
less than 50%	52 (3.7%)	69 (5.0%)	57 (4.1%)	19 (1.4%)	–	–
Total hand-counted votes (in millions)	20.5	15.3	19.4	57.6	23.0	19.0

NOTE: Results for 1393 election contests for U.S. president, Senate, and House of Representatives. Vote margins were calculated from FEC data for 2002 and 2004; Dr. Adam Carr's Psephos archive for 2006. Numbers of precincts per election were estimated based on the 2004 Election Assistance Commission's Election Day Survey (*http://www.eac.gov/election_survey_2004/intro.htm*); for House elections, the number of precincts in each state was divided by the number of Congressional Districts to estimate the precincts per District. A minimum of one precinct per county is assumed to be audited under each rule. Power is calculated to protect against miscounts residing in the largest precincts, as described in "Variations in Precinct Size" above.

practice, slightly larger samples may be useful to contend with minimal miscount rates observed even in well-functioning systems (Stark 2007).

8. PERCENTAGE-BASED VERSUS POWER-BASED AUDITS

We have already seen that percentage-based audits can examine too few precincts in some contests, while sampling more than necessary in others. To compare the cumulative costs of different types of audits, we studied all 1,393 federal election contests in 2002, 2004, and 2006, using the precinct size distribution of Ohio CD-5 in 2004 to estimate the extra auditing needed to account for variations in precinct size. Table 1 shows power and resource requirements using tiered, fixed, and power-based audits. Each audit was also made to fulfill a common legislative mandate: to include at least one precinct per county. For example, since Iowa has 99 counties and 1966 precincts, all rules assume audits of at least 99 precincts (5%) for every statewide election there. While 2% audits would have used the fewest resources (15 million ballots), they achieve less than 95% power in almost 15% of all contests. At the other extreme, 10% audits would have had at least 95% power in 95% of contests, but require auditing 57.6 million ballots, while still leaving 19 contests where election-altering miscounts are more likely to be missed than detected. Power-based audits could have achieved 99% power in all federal contests while examining only 23 million ballots.

9. CONCLUSIONS

Effective electoral oversight requires routine checks on the entire voting process, timely publication of the original tallies, and thorough reporting of the methods, data, and conclusions from all audits conducted. Since a key purpose of a post-election vote tabulation audit is to provide a check on the original tabulations, procedures should verify election results without trusting any part of the software used in voting. Reporting should include a post-hoc power calculation that fully describes the alternative hypothesis, so as to be explicit about how confident we are that the winner in the initial electronic tally is the same as would be identified in a 100% hand count of voter-verified paper ballots. Of course, all audits should be conducted within a larger framework of good practice, including following well-specified, publicly observable procedures to ensure public (as well as statistical) confidence in the process, examining both randomly selected precincts and targeted precincts with observed anomalies; and generating publicly accessible auditing data. Exactly one feature distinguishes power-based audits from percentage-based: they sample just enough precincts to make it very likely to detect a miscount if an election-changing miscount has occurred.

Election audits have been conducted in some state and local jurisdictions for years. The *electionline.org* briefing paper "Case Study: Auditing the Vote" discusses the auditing experiences of several states including California and Minnesota, and Saltman (1975) cited the 1% "manual tally" still used in California. Although professional auditors, statisticians and computer scientists should advise on standards and procedures, compe-

tent election officials and staff can implement power-based audit sample size calculations, detailed in McCarthy et al (2007), without special assistance. We encourage the use of statistical power-based audits as a preferred alternative to percentage-based audits. At the very least, we hope the laws being adopted now will allow States to use procedures that achieve high power to confirm all electoral outcomes.

[Received September 2007. Revised November 2007.]

REFERENCES

Aslam, J. A., Popa, R.A, and Rivest, R.L. (2007), "On Estimating the Size and Confidence of a Statistical Audit," April 22, 2007. Available online at *http://theory.csail.mit.edu/~rivest/AslamPopaRivest-OnEstimatingTheSizeAndConfidenceOfAStatisticalAudit.pdf*.

Dopp, K., and Stenger, F. (2006), "The Election Integrity Audit," September 25, 2006. Available online at: *http://electionarchive.org/ucvAnalysis/US/paper-audits/ElectionIntegrityAudit.pdf*.

Electionline.org (2007), "Case Study: Auditing the Vote," March 2007. Available online at: *http://www.electionline.org/Portals/1/Publications/EB17.pdf*.

Lobdill, J. (2006), "Considering Vote Count Distribution in Designing Election Audits," Revision 2, November 26, 2006. Available online at: *http://vote.nist.gov/Considering-Vote-Count-Distribution-in-Designing-Election-Audits-Rev-2-11-26-06.pdf*.

Marker, D., Gardenier, J., and Ash, A. (2007), "Statistics Can Help Ensure Accurate Elections," *Amstat News*, June 2007.

McCarthy, J., Stanislevic, H., Lindeman, M., Ash, A., Addona, V., and Batcher, M. (2007), "Percentage-based vs. SAFE Tabulation Auditing: A Graphic Comparison, November 2, 2007," Available online at: *http://www.verifiedvotingfoundation.org/auditcomparison*.

National Institute of Standards and Technology Technical Guidelines Development Committee (2007), "Voluntary Voting System guidelines Recommendations to the Election Assistance Commission." Draft available for comment online at (*http://www.eac.gov/vvsg*).

Norden, L., Burstein, A., Hall, J., and Chen, M. (2007), "Post-Election Audits: Restoring Trust in Elections." Available online at: *http://brennancenter.org/dynamic/subpages/download_file_50227.pdf*.

Saltman, R. G. (1975), "Effective Use of Computing Technology in Vote-Tallying," National Bureau of Standards, Final Project Report, March 1975, prepared for the General Accounting Office. [See particularly Appendix A, "Mathematical Considerations and Implications in Selection of Recount Quantities."] Available on the Internet at: *http://csrc.nist.gov/publications/nistpubs/NBS_SP_500-30.pdf*.

Stanislevic, H. (2006), "Random Auditing of E-Voting Systems: How Much is Enough?," VoteTrustUSA E-Voter Education Project, August 16, 2006. Available online at: *http://www.votetrustusa.org/pdfs/VTTF/EVEPAuditing.pdf*.

Stark, P. (2007), "Conservative Statistical Post-Election Audits," Technical Report 741, Department of Statistics, University of California, Berkeley. Available online at: *http://statistics.berkeley.edu/~stark/Preprints/conservativeElectionAudits07.pdf*.