How Big Should an Election Audit Be?

Fixed Rate Audits Do Not Work For Elections

http://electionarchive.org/ucvAnalysis/US/paper-audits/ElectionAuditEstimator.pdf
Kathy Dopp
kathy@electionarchive.org
January 17, 2007

Abstract

This paper presents a simple formula for estimating vote count audit sample sizes to achieve any desired certainty for ensuring the integrity of election outcomes. The formula described in this paper for estimating election audit sample sizes, was derived first by Ronald Rivest.[1]

In particular, this paper briefly shows how to derive an estimate for vote count audit sample sizes:

1. to achieve any desired probability of detecting vote miscount that could alter an election outcome, and
2. to incorporate the principle of maximum vote shift per machine vote count[2] that is required for logical consistency when auditing elections

Vote count audit sample sizes must be based on margins between the leading two candidates because the smaller the margin, the smaller the amount of miscount that could wrongly alter the election outcome, and the larger the audit sample must be to detect the smaller number of corrupt counts.

In conjunction with this paper, a spreadsheet is available to calculate exact minimum election audit sample sizes necessary to ensure the integrity of election outcomes. Marian Beddill helped to craft this new version of an earlier audit calculator spreadsheet created by Dopp.[3] This new spreadsheet allows any person, without a lot of experience, to enter the important factors and obtain a result and see the efficacy of doing smaller or larger audits. It is available here:

http://ElectionArchive.org/ucvAnalysis/US/paper-audits/HowManyToAudit.xls

Why Assume a Maximum Vote Shift Per Machine Vote Count?

A maximum rate of miscount within any one vote count must be assumed to derive an estimate because 100% of votes cannot be wrongly shifted within each vote count. It would not only be unlikely that 100% of votes are available to target, but it is also unlikely that anyone trying to rig an election would try to steal 100% of the target votes because it would be immediately noticed.[4] However, to avoid detection, a fraudster would corrupt as few counts as possible. So a maximum rate of vote shift per vote count is assumed to calculate the amount of corrupt counts that could wrongly alter an election outcome.

The larger the assumed maximum wrongful vote shift rate per machine vote count, the fewer number of corrupt vote counts could wrongly alter an election outcome; and the larger the audit sample size must be to detect the corrupt counts.

---

[1] Ron Rivest "How Big should a Statistical Audit Be?", November, 2006
http://theory.csail.mit.edu/%7Erivest/Rivest-OnEstimatingTheSizeOfAStatisticalAudit.pdf
Rivest's formula is used to calculate audit sample size by using a formula developed by Dopp, June 2006, "Paper Audits"
http://electionarchive.org/ucvAnalysis/US/paper-audits/Paper_Audits.pdf

[2] The Brennan Center "The Machinery of Democracy: Protecting Elections in an Electronic World", June, 2006

[3] July 2006 http://electionarchive.org/ucvAnalysis/US/paper-audits/AuditCalculator.xls

[4] If vote miscount were made by innocent error, miscount would be more likely to appear in all vote counts and so be detected with any audit amount. Independent audits should be designed to detect deliberate fraud.

We divide the margin between leading candidates by 2 to obtain the overall rate of votes which could be shifted to alter an outcome, and then divide by the assumed maximum wrongful vote shift per vote count to find the minimum percentage of vote counts that must be corrupt to wrongly alter an election outcome.[5]

Historical Background

In 1975 Roy Saltman first introduced the concept of the necessity to base election audit sample sizes on margins between candidates - the closer the margin between candidates, the smaller the amount of vote fraud or miscount and the fewer the number of corrupt counts which could wrongly alter the outcome.

To detect small numbers of corrupt vote counts, larger audit samples are necessary. For example, if the margin between the two leading candidates is 1%, then only approximately 3 corrupt vote counts out of 100 might put the wrong candidate into office. If the margin is 5%, it might take 15 or more corrupt vote counts out of 100 to wrongly alter the outcome. A larger manual audit sample size is needed to uncover one of 3 corrupt counts than to uncover one of 15 corrupt counts.

Unfortunately Saltman's work in 1975 was substantially ignored at the time until the concepts were rediscovered by Kathy Dopp in July 2006. Dopp furthered Saltman's work by applying a maximum vote shift per vote count assumption to derive a formula for calculating the amount of corruption that could alter an election outcome and used it, along with an estimate based on sampling with replacement[6] to create a trial-and-test spreadsheet method to obtain exact optimal election audit sample sizes.[7]

Beginning in July 2006, Dopp and Frank Stenger developed a numerical method to exactly calculate election audit sample sizes, which they released in September 2006, "The Election Integrity Audit"[8]. The Dopp/Stenger method included an optional method for adjusting audit sample sizes for precinct size variation in case miscounts are targeted to the largest precincts. However, the Dopp/Stenger numerical method has not caught on yet, perhaps due to the complexity of using a computer program at a time when the public is demanding transparent, easily-understand verification of election results.

Ronald Rivest of MIT derived a formula that more accurately estimates vote count audit sample sizes[9] than the one based on sampling with replacement suggested in the Brennan Center report because it gives a smaller over-estimate of the exact minimum audit required to ensure the integrity of election outcomes. For more detailed history and derivation of the formula, see Rivest's paper.

This paper more simply describes the derivation of the Rivest estimate and shows more explicitly how to use Rivest's formula to estimate audit sample sizes for specific margins and assumed maximum vote shift per vote count.

The minimum audit sample size necessary to ensure election outcome integrity to any desired certainty level can be found using this easy-to-use spreadsheet:
http://electionarchive.org/ucvAnalysis/US/paper-audits/HowManyToAudit.xls

[5] Vote counts with more than the assumed maximum vote shift relative to prior elections or partisan active voter registration records in voter history files, must also be included in manual audits as well as randomly selected counts.
[6] A sampling with replacement estimate was suggested in the appendices on parallel election day machine testing in the Brennan Center's June 2006 report "The Machinery of Democracy…"
[7] The Brennan Center introduced the idea of a maximum wrongful vote shift per machine in its appendix on sampling voting machines for Election Day random testing of paperless voting machines.
[8] http://electionarchive.org/ucvAnalysis/US/paper-audits/ElectionIntegrityAudit.pdf
[9] Ibid 1.

Formula for Estimating Vote Count Manual Audit Amounts

The following formulaic estimate can be used to find the exact minimum number of vote counts that must be manually audited to ensure the integrity of election outcomes.

Here are the variables used to calculate the estimate and used in the HowManyToAudit.xls spreadsheet:

N = the total number of vote counts (number of precinct, DRE, or batch vote counts)
m = the margin between the leading two candidates in any election contest
P = the desired probability of detecting one or more corrupt vote counts (suggest 95% to 99%)
v = the maximum wrongful rate of miscount that could occur on any one vote-counting device without raising immediate suspicion[10] (suggest 15% to 20%)
S = audit sample size – number of vote counts to manually count
C = the number of corrupt vote counts that could wrongly alter an outcome
c = the rate of corrupt vote counts that could wrongly alter an outcome

Note that the rate of corrupt vote counts that could wrongly alter an election outcome is $c = \frac{m}{2v}$ and[11] that $C = cN$. See "The Election Integrity Audit"[12] for a more detailed explanation of how to calculate the minimum amount of corrupt vote counts that could wrongly alter an election outcome.

Then a formula for estimating sample sizes for manual audits of vote counts is:

Equation 1: $S = N(1 - e^{\frac{\ln(1-P)}{Nc}})$ or

Equation 2: $S = N(1 - e^{\frac{2v\ln(1-P)}{Nm}})$.

Alternatively, for those who want to first exactly calculate the minimum number of corrupt precincts, C, that could wrongly alter an election outcome, using the number of ballots cast in each precinct, the formula to use is Equation 3: $S = N(1 - e^{\frac{\ln(1-P)}{C}})$ found by substituting $c = \frac{C}{N}$.

Derivation of Formula for Estimating Vote Count Manual Audit Amounts[13]

The derivation of the formula to estimate the minimum necessary election audit sample size is obtained using a well-known approximation of (1-c) for small values of x between 0 and 1.

i.e. $(1-c) \approx e^{-c}$ so that $(1-c)^S \approx e^{-cS}$.

Estimating our desired probability from sampling with replacement, we begin with the probability of not detecting any miscount: $(1-c)^S \approx 1-P$ so taking the LN gives $\ln(1-P) \approx -cS$ and solving for S gives $S \approx \frac{-\ln(1-P)}{c}$. To further improve the estimate for S, we use the formula for estimating the number of distinct elements, S, in a sample of size t drawn (with replacement) from a set of size N:

$S \approx N(1 - e^{\frac{-t}{N}})$ and replace t by our estimate for S above, resulting in Equation 1: $S \approx N(1 - e^{\frac{\ln(1-P)}{Nc}})$.

[10] It is assumed that if more than this amount of wrongful vote shift occurs in any vote count that comparison of vote counts with comparable results in prior elections or with partisan voter registration in the voter history files would raise immediate suspicion. Vote shift * 2 = Margin shift.
[11] This formula was derived by Dopp in July, 2006.
[12] pp. 8-9 "The Election Integrity Audit" http://electionarchive.org/ucvAnalysis/US/paper-audits/ElectionIntegrityAudit.pdf
[13] ibid 9. For more details on the history and derivation of the formula for estimating vote count audit sample sizes, see Rivest's paper.