Election Confidence

A Comparison of Methodologies and Their Relative Effectiveness at Achieving It

C. Andrew Neff*

Revision 6 – December 17, 2003

1 Introduction

There are two distinct methodologies for auditing electronic election data. The first of these, which we will call *Precinct Count* methodology, is the one most commonly used. It involves comparing, for a small number of *sample precincts*, a "hand recount" of the precinct's *paper ballots* with the electronic tally for that same precinct. The second methodology, which we will call *Vote Receipt* methodology, involves issuing indisputable receipts for either "real ballots", or "test ballots", or both, and comparing, for a small number of *sample ballots*, the receipt values with the electronic ballot data.

In this paper, we first observe that on an abstract level, the "fraud detection power" of these two methodologies are essentially the same. In both cases, establishing "election confidence" comes down to selecting the same number of *sample points* (either precincts or ballots) randomly. We then argue that the primary differences between the two methods are in the areas of logistics, scalability, and the feasibility of implementing a real world system that accurately embodies the corresponding system abstraction. With respect to these characteristics, the Vote Receipt methodology has a significant advantage.

*aneff@votehere.net

2 Basic Model

We begin by looking at the *abstract*, or *theoretical* detection power of any random audit process. Each comparison of audit data with electronic data either detects compromise (if the comparison fails), or adds a certain amount of "confidence" that there has been no compromise.

For the remainder of this paper we will use the generic term "population" to describe a set of elements – precincts, ballots, etc. – that are to be sampled and "tested". For example, under the precinct count methodology, "testing" means:

PC1. Hand recount paper ballots for the specified precinct.

PC2. Compare recount total with same precinct's electronic total.

PC3. If totals are equal, SUCCESS; else FAIL.

Under the receipt verification methodology, "testing" means:

R1. Compare receipt data with electronic ballot for specified Ballot ID.

R2. If values are identical, SUCCESS; else FAIL.

Assuming that all tests are chosen *randomly*, one can precisely compute the probability of detection for any given fraud rate, ρ, as a function of the number of test performed. (See the appendix Detection Probability for one way to perform this analysis.)

3 California Statewide Benchmark

3.1 Precinct Count

In the state of California (CA), there are a total of 25702 precincts. As a reasonable approximation, we may assume that all precincts have about the same number of voters. In this case, an overall ballot fraud rate, ρ, can potentially be achieved by compromising the maximum number of ballots in $25702\,\rho$ precincts. So, if T precincts are selected *at random* to be recounted, the worst case detection probability for ballot fraud rate ρ is given by $p_D(25702\,,\,25702\,\rho\,,\,T)$.[1]

Figure 1 summarizes election confidence levels (i.e. fraud detection probabilities) attempted fraud rates of 0.1%, 1%, 2.5%, 5%, 10%, and 20%.

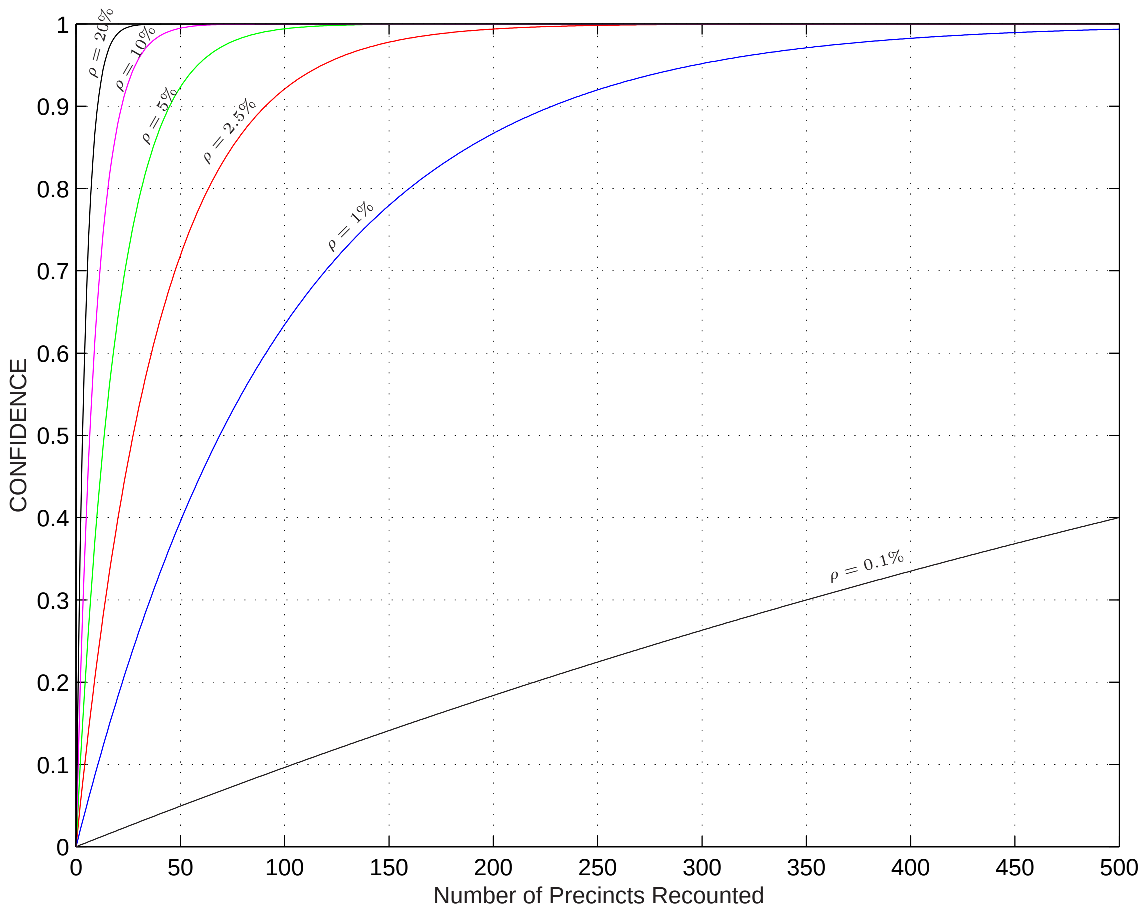

Figure 1: Confidence for California Statewide Election : $p_D(25702\,,\,25702\rho\,,\,T)$

To correlate this with standard practice, California selects about 257 (1% of precincts) for "hand-recount" (i.e. random audit)[2]. So there is roughly a 93% level of confidence (0.93 detection probability) that no more than 1% of the ballots have been changed, and roughly a 23% level of confidence that no more than 0.1% of the ballots have been changed.

[1] In a later section we will discuss heuristic reasoning which intuitively suggests that this worst case estimate may be somewhat pessimistic. Nevertheless, it is possible.

[2] California Elections Code, Section 15360.

Remark 1 The graphs in Figure 1 assume that the recount precincts are selected randomly and *uniformly* from all precincts participating in the election. Actual California law specifies a slightly different precinct selection distribution, one where each *county* (represented by its election official) selects *its own* 1% of precincts for recount. Though this may not seem like an important distinction, it actually has a small effect on confidence levels. In fact, it does have the effect of *reducing* detection probabilities. Under the California configuration of counties and precincts, the effect is small enough that the difference between the graphs of Figure 1 (uniform selection) and the ideal graphs (selection of 1% per county) are hardly noticeable. In a subsequent revision of this paper we will include a figure containing the ideal graphs for the sake of comparison.

3.2 Receipts

Amazingly, the same level of confidence can be obtained by auditing 257 randomly selected (real) ballot receipts (state wide). Much higher levels of confidence can be obtained if more receipts are audited.

If only control receipt tests are used, the detection probabilities do change slightly, but imperceptibly. As evidence, Figure 2 shows graphs for the same fraud rates for $T = T_c$, the total number of control receipt tests. It assumes that there are 360 ballots cast per precinct (roughly a 60% voter turnout), giving a vote total of $9,252,720$.

4 California Congressional District Benchmark

Figure 1 applies to a *statewide race* like Governor or U.S. Senator, since all precincts participate in that election. However, suppose a malicious insider were determined to change a single Congressional race? In this case, fewer ballots need to be changed to swing the race by any given percent of *its* cast vote total. Using the fact that there are 53 California Congressional districts, the typical district has about 476 precincts, so a 1% recount would mean “hand counting” about 5 precincts. In this case, the graphs of Figure 1 are ideal (that is, remark 1 does not apply), but the range of interest for the “Number of Random Audits” axis (i.e. number of precincts recounted) is different. The estimates for confidence levels are shown “blown up” in Figure 3.

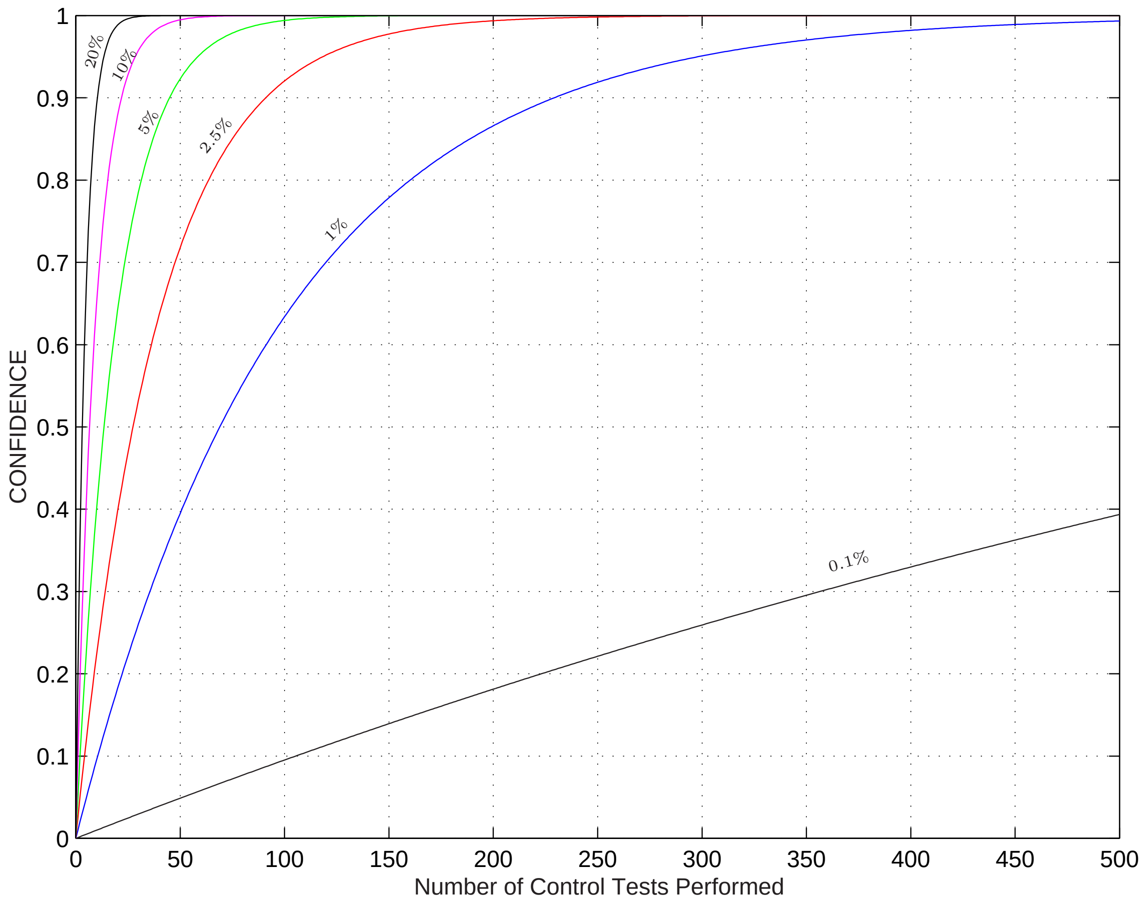

Figure 2: Confidence for California Statewide Race : $q_D(9252720\,,\,9252720\rho\,,\,0\,,\,T)$

Again correlating this with standard California practice, this means that there is only about a 40% level of confidence that a given Congressional race has not been changed by 10%, and only about a 5% level of confidence that it has not been changed by 1%.

Contrast this with the data in Figure 2. If only 500 control receipts are checked in each Congressional district, there would be better than 99.9% confidence that the election count has not been changed by more than 1%. Since there are roughly 476 precincts per Congressional district, this amounts to only about *one receipt check per precinct*.

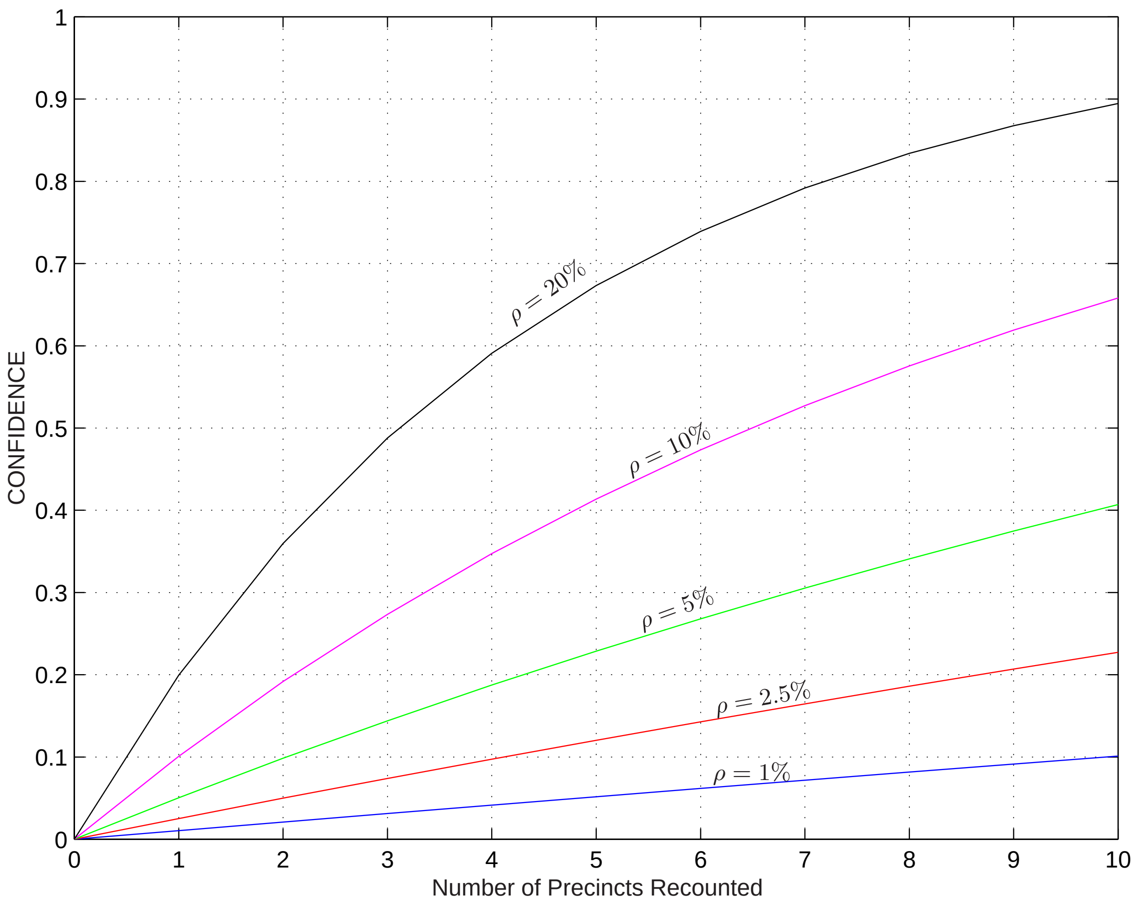

Figure 3: Confidence for California Congressional Race : $p_D(476, 476\rho, T)$

5 Error Margins

A common way to look at data of this kind is by way of error margins. The following two tables summarize approximate error margins for various different test totals at the 90% and 99% confidence levels.

# **Tests** (precincts or ballots)	**Error Margin (%)**
5	36
250	1.2
500	0.5
2000	0.1

Figure 4: Error Margins at 90% Confidence

# **Tests** (precincts or ballots)	**Error Margin (%)**
5	60
250	2.3
500	0.9
2000	0.2

Figure 5: Error Margins at 99% Confidence

For example, one sees that if, in a given race, 250 precincts are successfully recounted, or 250 receipts are successfully checked, one can be 90% certain that the electronic count is correct to within 1.2% (Figure 4), or 99% certain that it is correct to within 2.3% (Figure 5).

For the Congressional district example, roughly 5 precincts would be recounted. *At the 99% confidence level (Figure 5), the error margin is an astounding 60%!*

6 Theory vs. Practice

The initial analysis only gives levels of confidence assuming that no tests fail. In practice, one must consider both

1. How likely is it for a test to fail due to "natural" (i.e. non-malicious) causes.

2. What confidence levels can be obtained if the model is modified to allow for these occasional "natural" failures.

On the first point, receipts are inherently better. Digital signature techniques can be employed to protect receipts from changing in any potentially confusing way after they are issued, *and* a few lost receipts does not effect the detection probability in any appreciable way. On the other hand, a recount problem leads to the kind of unresolvable disputes we saw in Florida during Election 2000.

7 Improving Precinct Count by Heuristic

In practice, the estimates we have given for precinct count detection probabilities *may be* pessimistic for the following reason. A precinct reporting a *unanimous* vote for one candidate or issue might be immediately flagged as suspect, and slated for recount. Hence, very large corruptions of a precinct total *might* be caught with near certainty if a heuristic is added to the precinct count strategy – *always hand-recount those precincts with a margin of victory greater than some predetermined threshold, τ.* This creates a dilemma however. Set τ too high, and machines can still alter a significant fraction of ballots without detection. On the other hand, set τ too low, and the jurisdiction could be faced with an unmanageable manual recount process.

As a very rough guess, for example, we might suggest $\tau = 20\%$. In this case, in order to change 1% of the total, it turns out that machines would have to change at least 1.7% of the precincts (possibly fewer if one is willing to assume that "naturally occurring" victory margins are *always* confined to a narrow range). This would indicate that a recount of about 84 precincts – *in addition to all those precincts with margin of victory greater than* 20% – would be sufficient to achieve the 90% confidence level at detecting a 1% fraud rate. Again, assuming roughly 360 ballots cast per precinct, this still requires handling roughly $30,240$ ballots (with 100% accuracy) however. Still considerably more than the 230 ballots required by the control receipt strategy. Moreover, 90% confidence is not particularly satisfying when it comes to a highly contested election.

Obviously there is much work that must be done to define the optimal details of this heuristic strategy and analyze its true power to detect election fraud. However, the fact that its actual effectiveness is based on an *assumption* about voting patterns will always be unsettling.

8 Conclusion

This analysis shows that for auditing election data, a Vote Receipt methodology has significant advantages over the commonly used Precinct Count methodology. In fact, the shortcomings of Precinct Count, as practiced in California, are disturbing enough to warrant giving careful consideration to the idea of moving to Vote Receipt based systems.

We hope this analysis will serve as a point of departure for establishing a figure-of-merit by which to objectively judge all election verification methodologies.

A Detection Probability

We will use the following variables as notation:

N	Population size
F	Number of elements of population which are "compromised" (i.e. elements that would produce FAIL if tested.)
T	Number of "test" elements

Assuming that test elements are selected randomly and independently[3], we can derive an exact expression for the probability of a detecting a compromised element as a function of N, F and T. The number of T-element subsets that *do not* contain a compromised element is

$$\binom{N-F}{T} = (N-F)!\,/\,T!\,(N-F-T)!$$

while the total number of T element subsets of the population is

$$\binom{N}{T} = N!\,/\,T!\,(N-T)!$$

By assumption, all subsets are equally likely to appear as the test subset, meaning that the probability of *not* detecting a compromised element is

$$\bar{P}_D(N,F,T) = \frac{\binom{N-F}{T}}{\binom{N}{T}} = \frac{(N-F)!\,(N-T)!}{N!\,(N-F-T)!} \tag{1}$$

for all N, F, and T with $F+T \leq N$. (Clearly $\bar{P}_D = 0$ if $F+T > N$.)

So the *detection probability*, P_D, is

$$P_D(N,F,T) = 1 - \bar{P}_D = \begin{cases} 1 - \frac{(N-F)!\,(N-T)!}{N!\,(N-F-T)!} & \text{if } F+T \leq N \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

[3] The same results will also hold under weaker assumptions.

In order to evaluate this probability for "reasonably sized" values of N (i.e. $N \approx 20$ or greater), we recall Stirling's asymptotic formula for the factorial function:

$$n! \;\approx\; \sqrt{2\pi}\, n^{n+1/2}\, e^{-n} \tag{3}$$

If we define $S(x) = (x + 1/2)\log x$, a good approximation for P_D when $F + T \leq N$ is given by

$$P_D(N, F, T) \;\approx\; 1 - \exp\left(\begin{array}{c} S(N-F) + S(N-T) \\ -S(N) - S(N-F-T) \end{array}\right) \tag{4}$$

A.1 A Useful Approximation

Assuming that $T + F \ll N$ (which is nearly always the case in practice), the function P_D can be roughly approximated by

$$P_D(N, F, T) \;\approx\; 1 \;-\; (1 - F/N)^T \tag{5}$$

From this one can see, for example, that to detect a compromise of 100 elements of the population with 95% confidence, one must test roughly 3% of the elements.

A.2 Fraud Rate

For election confidence, the absolute magnitude of F does not matter so much as its magnitude *relative to* N, since this determines the likelihood that the final winner will be changed by compromise. For example, in an election of 1000 voters, a change in 100 ballots can *easily* throw the election, whereas in an election of $10,000,000$ ballots, as long as there are not thousands of candidates running for the same race, 100 ballots can be considered an almost imperceptible fraction of the "will of the majority."

For this reason, we will focus on the behavior of the function

$$p_D(N, \rho, T) \;=\; P_D(N, \rho N, T) \tag{6}$$

In words, this is the probability of detecting a *fraud rate* $0 \leq \rho \leq 1$.

A.3 Real Tests vs. Control Tests

Test elements can come in two different types. If an element that is tested is also an element to be included in the final result, we call this a *real* test element. With the precinct count methodology, all test elements are real since recounted precincts are also included in the final tally. On the other hand, with the receipt methodology, it is possible to test either actual voted ballots, or “control ballots” (which are intentionally removed, as a part of the audit protocol, from the final count), or both. In order to study these more general audit processes, we need to introduce another detection function derived from P_D

$$q_D(N\ ,\ \rho\ ,\ T_r\ ,\ T_c)\ =\ P_D(N+T_r\ ,\ \rho N\ ,\ T_r+T_c) \tag{7}$$

This function measures the detection probability for a given election fraud rate, ρ, assuming that T_r real tests and T_c control tests are performed, for a total of $T = T_r + T_c$ tests. In the extreme case, when $T_c = 0$ (so $T = T_r$), q_D is identical to p_D.

We will see that for “reasonable” values of the parameters, q_D does not differ significantly from p_D. Nevertheless, we need to study both functions in order to demonstrate this fact.

