

Summary of Analysis of the 1% Manual Tally of the June 2010 Primary Election in Los Angeles County

In the accompanying table and graphs, we delineate which are the worst counted precincts in our sample. The error rate for reading the markings on ballot cards is approximately one missed mark in 100 ballot marks. This "error rate per ballot mark" or "global errors per mark" is 5,000 times above the Federal Election Commission's voluntary guideline of one miscounting error in 500,000 ballot marks. Most of the errors we saw are due to failure to detect ballot marks, not insertion of marks that voters did not intend. Analysis of the data indicates that the errors are *heterogeneously distributed* among the voting precincts. Several precincts in our sample are *outliers*, in that they incurred a *much* higher error rate – up to 100-fold higher – than most of the others. For these precincts, the inaccuracy of the count impacted all the contests and measures voted on. The error rate of the outlier precincts' ballots is staggering for an advanced industrialized country: For example, in the Vote By Mail (VBM) tally of the statewide contests and measures common to all precincts, the worst four precincts had error rates of 32, 25, 17 and 7 percent. These error rates were so high that they almost tripled the overall error rate of the 60 precincts in our sample of the VBM tally from .66% to 1.76%. The Election Night (EN) tally did not have such extreme outliers, although the fraction of precincts with zero errors was about half that of the VBM. In our EN sample, four precincts had error rates above 2%: 2.2%, 2.8% and 3.4%, and 4.3%.

Our analysis shows that each precinct incurred a characteristic error rate that is roughly uniform for all the contests and measures on the ballot. The distribution of error rates of the EN and VBM counting are quite different. Since VBM voters mark their ballots with pens at home, the poor performance of the VBM tally would seem to be due to malfunctioning machines, not light ink on all the cards in the precinct. The unreliability of the VBM count suggests that voters might be better served by using the booth on election day, not a VBM envelope that gets opened a month after the election is decided.

Because our data were from only 1% of all the precincts in Los Angeles County, smaller jurisdictions such as an Assembly District (AD) provided only one or two precincts to the dataset. Therefore we initially focused on the analysis of contests and measures common to all precincts in our sample. However, we calculated the error rates for contests and measures not common to all precincts separately, and found that the accuracy for those races and measures is about the same as the accuracy of that precinct's tally for races and measures common to all precincts. All the ballot marks in an individual precinct incurred errors with a probability *characteristic of the precinct as a whole*.

The observed pattern of error rates strongly suggests that several machines used to count both EN and VBM ballots in the June 2010 election functioned very poorly compared to the rest (and thousands of times below the Federal guidelines). We urge that the machines associated with the poorly counted precincts be traced, their performance assessed, and fixed or taken out of use before the November election.

Antonie K. Churg, Ph. D.

October 14, 2010

2010 June Audit Analysis. doc

Table 1: Parameters describing the error rates calculated for the June 2010 1% Manual Tally

| | Election Night | | Vote By Mail | |
|-----------------------------------------------|------------------------|-----------------------------------------|------------------------|--------------------------------|
| | in Common (§) | Not in Common | in Common | Not in Common |
| Global errors per mark, % ^(†) | 0.8386 | 0.7796 | 1.3036 | 2.9611 |
| Total # ballot marks | 105531 | 30913 | 42573 | 11685 |
| Total # missed or added marks ^(‡) | 885 | 241 | 555 | 346 |
| Mean error rate by precinct, % ^(§) | 0.8181 | 0.7160 | 1.7653 | 3.7174 |
| Median error rate, % | 0.6639 | 0.3929 | 0.3433 | 1.0663 |
| Worst outlier precincts | 1797, 3460, 3684 | 329, 1797, 3103, 3460, 3684 | 651, 4117, 4123, | 651, 3219, 4117, 4123 |
| Mean without worst, % ^(§) | 0.6565 | 0.4713 | 0.6652 | 1.9700 |
| Median without worst, % | 0.5449 | 0.2814 | 0.3058 | 0.6061 |
| Global rate without worst, % ^(¶) | 0.6446 | 0.4926 | 0.6991 | 2.2410 |
| Total # of precincts | 55 | 55 | 60 | 60 |
| # ballots cast | 6246 | 6246 | 2530 | 2530 |

Notes and definitions:

(§) "In Common" means "contest and measures common to all precincts". In this election, these were the statewide offices, U.S. Senate, and the state ballot Propositions. "Not in Common" means "contest and measures that were *not* voted on by all precincts in the sample". For example, the votes for State Assembly were generally sampled in only one precinct.

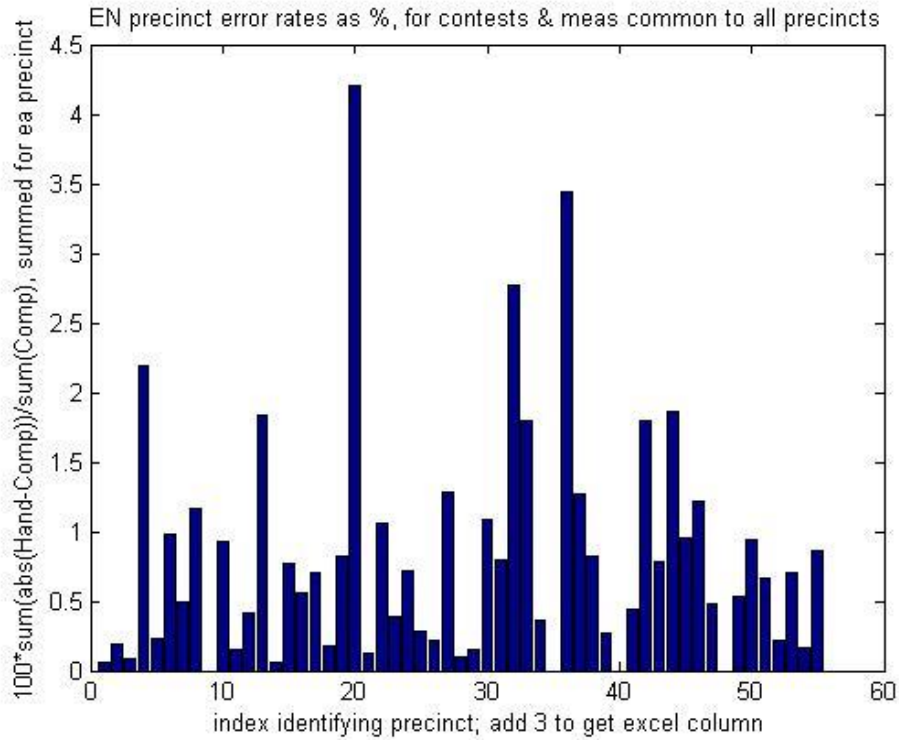
(†) Global errors per mark, or global error rate, is simply the total # of missed or added marks divided by the total # of ballot marks. We report the error rate in units of percent, not as a fraction. The formula for the global error rate in percent is 100 times the ratio of two sums: $100 * \frac{\text{sum}(\text{absolute value}(\text{Hand count} - \text{Computer count}))}{\text{sum}(\text{Computer count})}$. In each precinct, for each candidate or measure position there is a Hand count and a Computer count, or tally. The numerator involves the tally difference, the denominator just the Computer count. For the error rate we call "global errors per mark", the "sum" is a two-step sum: In the first step, for each precinct, take a sum over all candidates and measures being considered. In the second step, sum those sums over all precincts in the sample. Do these two-step sums *separately* for numerator and denominator and *then* form the ratio.

(‡) $\text{sum}(\text{absolute value}(\text{Hand count} - \text{Computer count}))$. This is also a double sum; see (†). Most of the discrepancies between hand and computer are due to missed marks, but occasionally the computer records marks that were not present. We use absolute values of the tally differences because otherwise the erroneously added marks would compensate for the missed marks.

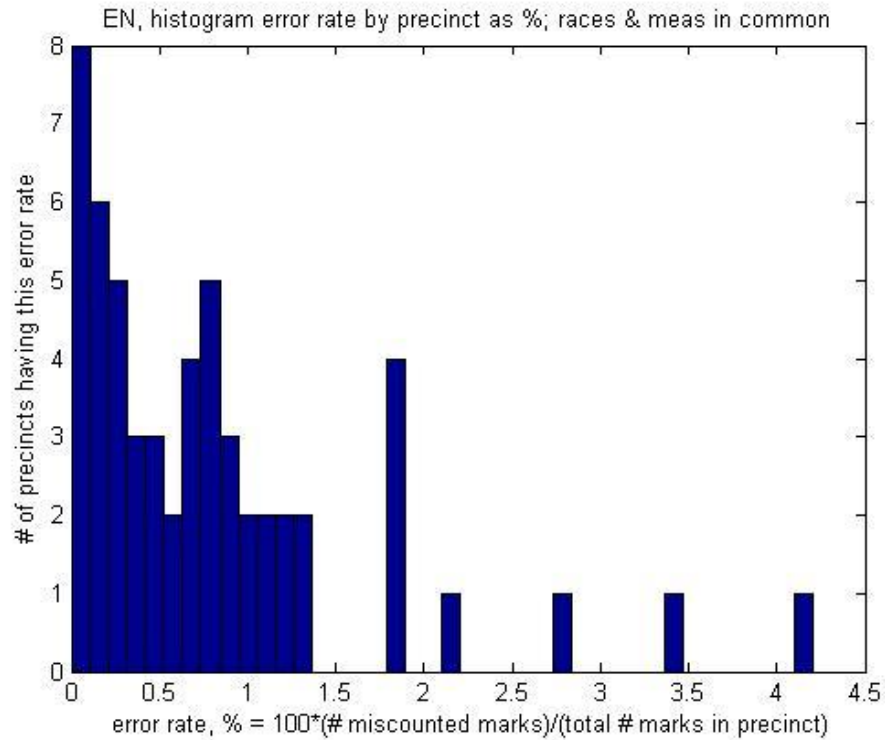
(§) This parameter is not a rigorous measure of the global error rate, but it is in the same ballpark as the global errors per mark. It is the average of the quantities resulting from doing only "step one" in (†) above, and then forming the ratio for each precinct. It would be exact if all the precincts had the same total Computer count. It is included here mainly because the median precinct error rate is a legitimate quantity, and the large discrepancy between the mean and the median indicates that the errors do *not* follow a Gaussian distribution.

(¶) The global error rate without the worst precincts is calculated with the same formula as in (†), except that the "worst" precincts are removed from the dataset.

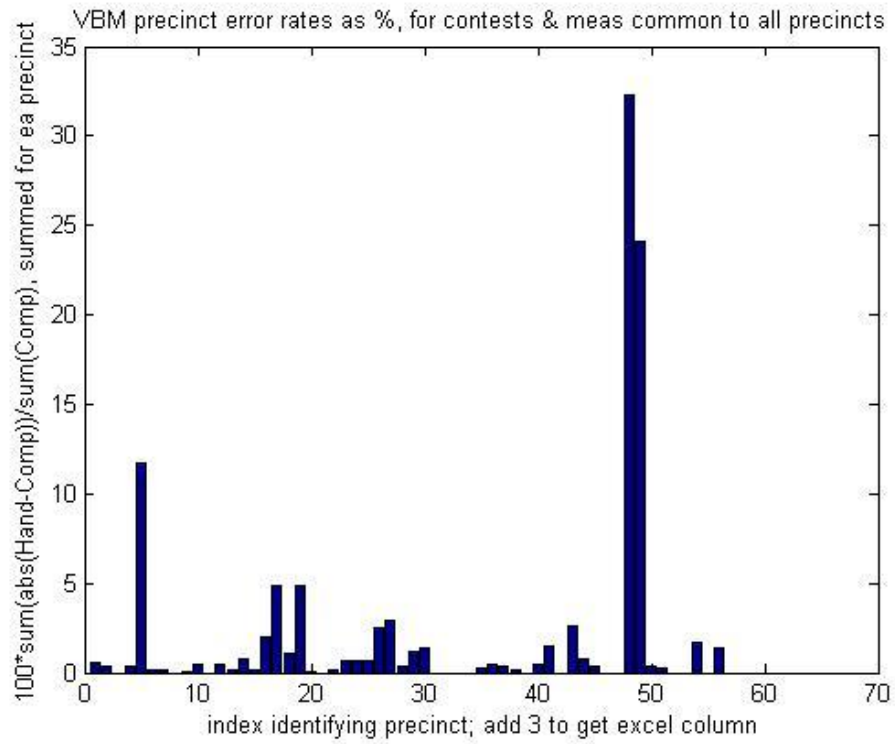
enBarErrorRatesByPrecinctCommon10-13-10.jpg



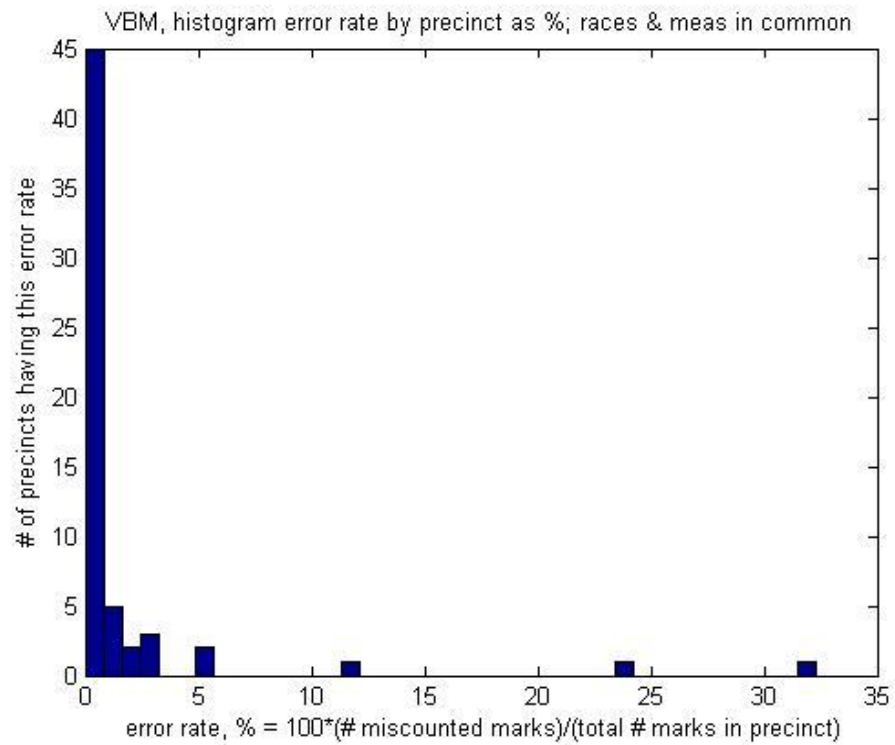
enErrorRateHistogramCommon10-13-10.jpg



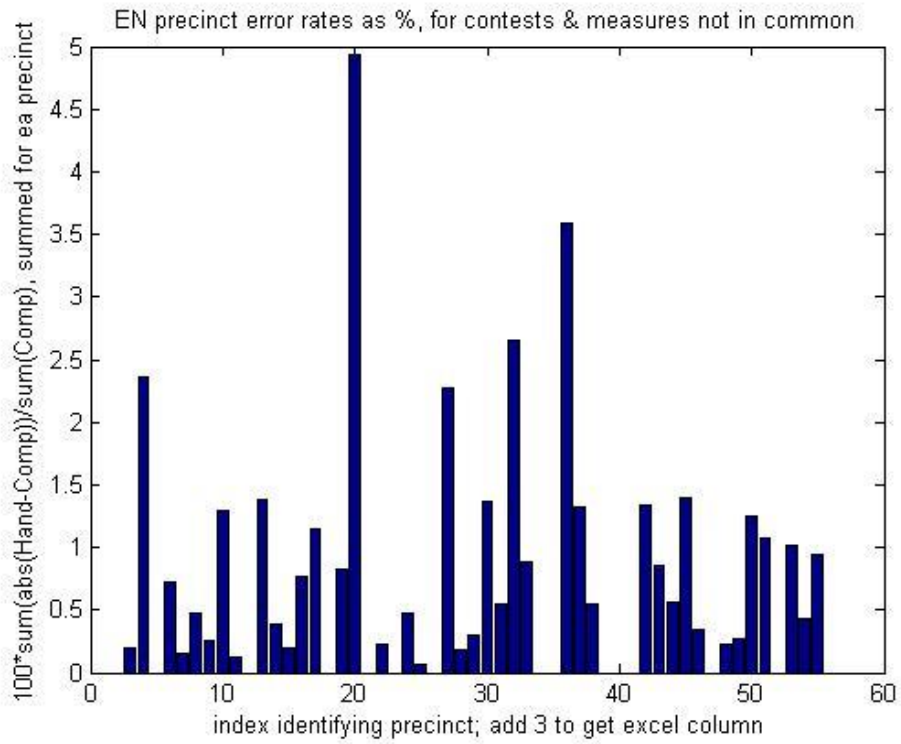
vbmBarErrorRatesByPrecinctCommon10-13-10.jpg



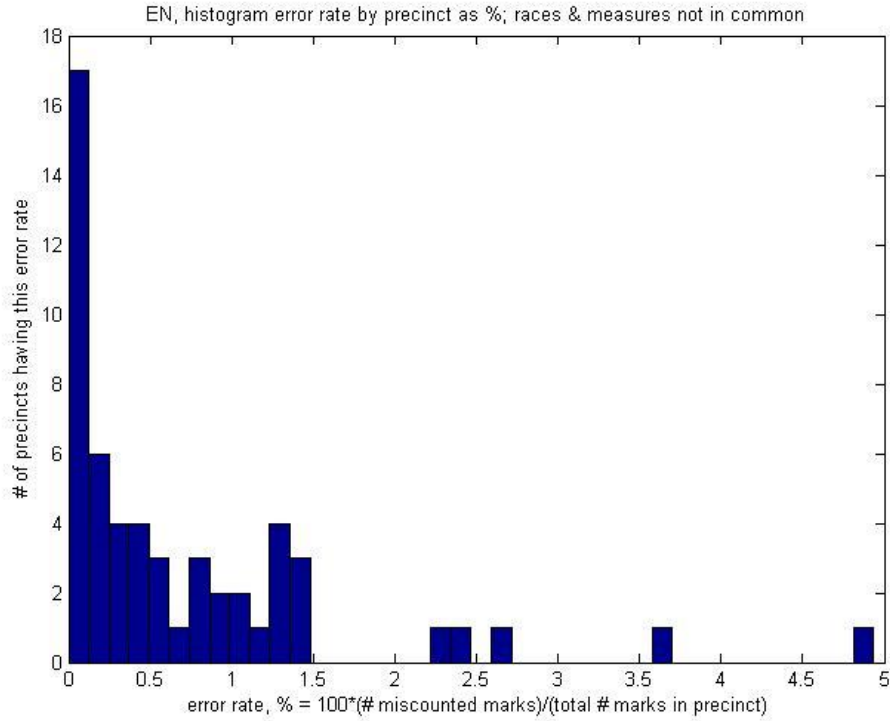
vbmErrorRateHistogramCommon10-13-10.jpg



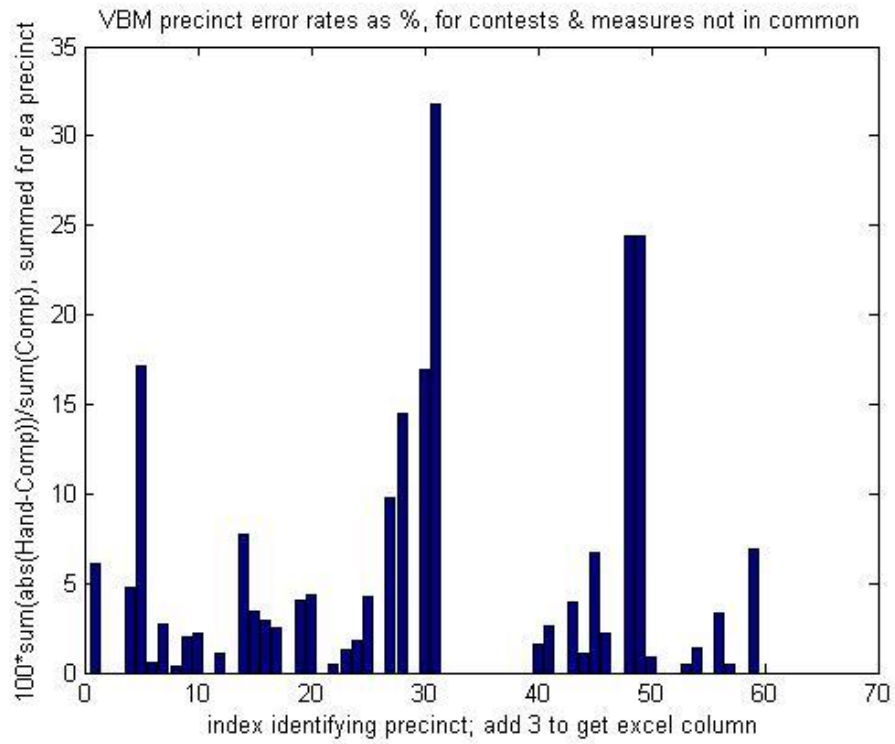
enBarTrueErrorsByPrecinctNotInCommon10-12-10.jpg



enErrorRateHistogramNotInCommon-10-12-10.jpg



vbmBarTrueErrorsByPrecinctNotInCommon10-08-10.jpg



vbmErrorRateHistogramNotInCommin10-08-10.jpg

