ILLINOIS BALLOT INTEGRITY PROJECT

www.ballot-integrity.org

A SURVEY OF SAMPLING METHODS

June 11, 2005

For further information concerning this document, please contact:

Robert A. Wilson
Chairperson, Suburban Cook County Chapter
Illinois Ballot Integrity Project
635 Chicago Ave – Suite 127
Evanston IL 60202
(847) 644-2654
wilson@ballot-itegrity.org

A SURVEY OF SAMPLING METHODS

Why sampling audits are inappropriate for individual precinct counts and how they can be used at the city-wide or county-wide level to detect miscounted precincts

Introduction

This study was undertaken in conjunction with the Ballot Integrity Project, to determine ". . . the number of votes we need to count in every precinct on the night of the election to secure a precinct of 400 actual voters with a 99.9% confidence level." It is our intention to address this question for both the City of Chicago and Suburban Cook County.

It is feasible to address both because the number of registered and actual voters in each area is similar (e.g. in the race for president/vice president, 1,032,878 votes were cast in the City of Chicago and 1,008,910 in Suburban Cook County); the number of precincts (Chicago - 2,709, Cook County – 2,402); creating an average precinct turnout in Chicago of 381 and Cook County, 420. Thus, the population of voters in each of the major areas is very close and the size of precincts, in terms of actual turnout on average is only about 10% larger in Cook County.

The range of votes cast by precinct in Chicago varied from a low of 23 to a high of 1,023, and by ward from 6,684 (12th) to 32,530 (42nd). Average votes per precinct (by ward) varied from a low of 229 (26th) to 623 (44th). [1] In Suburban Cook County, votes cast by precinct showed similar results with a low of 54 and a high of 1,060. Average votes per precinct varied from a low of 282 (Cicero Township) to 608 (Lemont Township). [2] Thus, in terms of number of precincts and actual votes per precinct in the 2004 general election for president/vice president, the City of Chicago and Suburban Cook County are relatively similar in precinct/voter distribution.

Preliminary Observations

Confidence Level/Margin of Error

The **margin of error** is the plus-or-minus figure usually reported in newspaper or television opinion poll results. For example, if you use a margin of error of 4 and 47% percent of your sample picks an answer you can be "sure" that if you had asked the question of the entire relevant population between 43% (47-4) and 51% (47+4) would have picked that answer.

The **confidence level** tells you how sure you can be. It is generally expressed as a percentage and represents how often the true percentage of the population who would pick an answer lies within the confidence interval. The 95% confidence level means you can be 95% certain; the 99% confidence level means you can be 99% certain. Most market research and general opinion polls use the 95% confidence level. Expressed somewhat differently, a 99% confidence level means that if you repeated the survey 99 times, 99 times out of 100, the results would fall within the margin of error (the + or -% indicated – remember that the margin of error is + or -, so that a 1% margin of error actually indicates a 2% range of error.

The margin of error is just a re-expression of the sample size, *N*. The numerators of these equations are rounded to two decimal places:

Margin of error at 99 percent confidence = 1.29/SQRT *N*
Margin of error at 95 percent confidence = 0.98/SQRT *N*
Margin of error at 90 percent confidence = 0.82/SQRT *N*

[1] See Appendix A
[2] See Appendix B

Sample Size

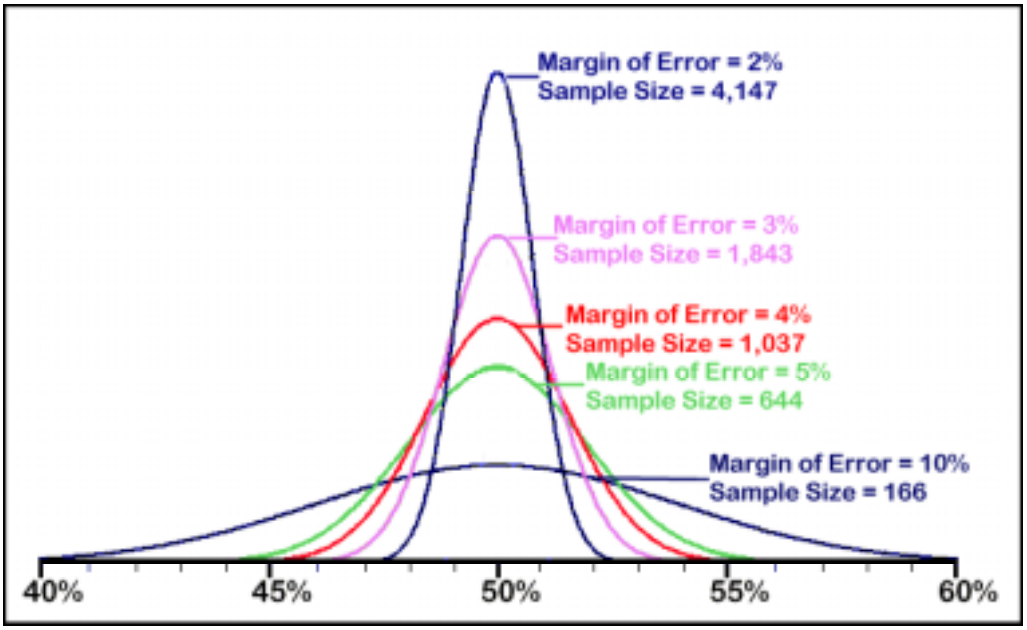
Margin of error decreases as the sample size increases, but only to a point. A very small sample, such as 50 voters, has about a 14 percent margin of error while a sample of 1,000 has a margin of error of 3 percent. The size of the population (number of ballots being surveyed) does not matter. (This statement assumes that the population is larger than the sample.) There are, however, diminishing returns. By doubling the sample to 2,000, the margin of error only decreases from plus or minus 3 percent to plus or minus 2 percent.

Although a 95 percent level of confidence is an industry standard for surveys and polls, a 90 percent level may suffice in some instances; or in others, where greater accuracy is desired, a 99 percent level may be used. A 90 percent level can be obtained with a smaller sample, which usually translates into a less expensive survey or poll.

It's an interesting mathematical fact that this margin of error depends ***only*** on the sample size and ***not*** on the population, provided that the population is significantly larger than the sample size (there are corrections that can be made for finite sample sizes). Thus a sample of 4,147 randomly sampled voters would yield essentially the same margin of error (2% with a 99% level of confidence) regardless of whether the population of voters consisted of 100,000 people or 100,000,000 people.

While this may seem counter-intuitive as each person in the population is unique and in a very large population only a very small fraction of people would actually be sampled, it would therefore seem that the sample is not capturing enough data. However, because the question involves only a very specific vote for a specific race, there is only one relevant attribute in the population that needs to be considered (yes/no - Candidate A or B [and occasionally C, D & E]). This means that any individual's vote is effectively equivalent to those of a large number of other voters, some fraction of which will be sampled.

Here is a graphic interpretation of sample sizes needed at the 99% confidence level that will produce a given margin of error: [3]



NOTE: For this graph, we have not shown the 1% margin of error as it would require the graph to run off the top of the page. The sample size required is 16,341, or roughly four times the sample size for the 2% margin of error (at the 99% confidence level).

[3] Graph adapted from Wikepedia's discussion: http://en.wikipedia.org/wiki/Margin_of_error

Further Considerations Regarding Margin of Error

- The margin of error is a simple transformation of the number of respondents into an ambiguous term that is neither a "margin" nor the whole of "error".
- The margin of error is often confused with the confidence level in reported percentages.
- The 99 percent confidence interval radius is smaller than the margin of error for any percentage other than 50 percent; and it is much smaller and more asymmetric for very high and very low percentages.
- It is not a "margin" at all as the probability of the true percentage being outside the margin of error is low but still has a value greater than zero.
- When the purpose of surveys is to compare percentages, the use of the margin of error is tempting but inappropriate.

Why Sampling Won't Work

As expressed above, the objective of this inquiry was to determine ". . . the number of votes we need to count in every precinct on the night of the election to secure a precinct of 400 actual voters with a 99.9% confidence level."

The problem of random sampling is that to obtain a high enough confidence level and low enough margin of error at the precinct level, the needed sample size equals the population, even when making minor adjustments to the formulae used to provide for a finite population, such as the number of voters in a precinct. Thus random sampling is an inappropriate methodology for determining the accuracy of the votes cast by precinct (either optically scanned or the Sequoia touch screens).

Random sampling only makes sense over a very large population. For example, the entire City of Chicago. Assuming approximately 1.1 million votes cast, a sample with 1% margin of error and a 99% confidence level would be the same 16,341 as above, or about six (6) votes per precinct. Increasing the confidence level to 99.9% only requires about 10 votes per precinct, or a sample of 26,419.

Reducing the margin of error, however, requires substantially increasing sample sizes. For example, decreasing the margin of error to 0.5% at a 99% confidence level requires 62,675 ballots to be sampled. Increasing the confidence level of the 0.5% margin of error sample to 99.9% would require 98,573 ballots to be examined, or about 36 per precinct. Even a slight gain in margin of error (from 0.5% to 0.4%) would require a 48.8% increase in the sample size to 146,630. To reach a 0.1% margin of error at a 99.9% confidence level would require a sample of 782,156 or 71% of the total votes cast.

The problem with sampling is that it's least reliable when you need it most, in those races that are "too close to call." The margin of error increases as the proportions tend toward even (50-50). All of the above examples of sample sizes are based on such an even division. Thus, you can see that the generally recommended 2% sample (22,000 ballots) has some merit in a large population such as the entire city as it would yield approximately a 0.86% margin of error at a 99% confidence level. [4]

This leads to the inescapable conclusion that at the precinct level, sampling cannot work because the sample size needed at this level either approaches or equals 100 percent, for the smallest/largest precincts:

99% Confidence Level – 1% Margin of Error:

Precinct Size = 23 - Sample Size = 23 - Precinct Size = 1,060 - Sample Size = 997

4 For a fuller discussion of confidence levels and margin of error, try here:
www.isixsigma.com/library/content/c040607a.asp

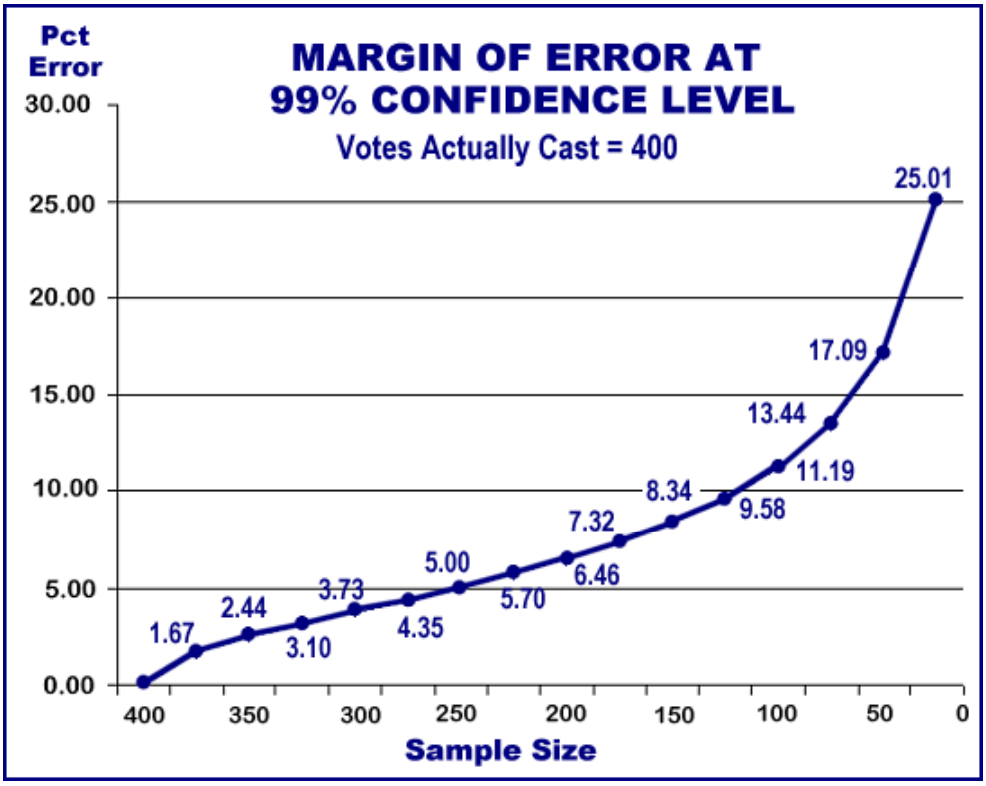
99% Confidence Level – 0.1% Margin of Error:

Precinct Size = 23 - Sample Size = 23
Precinct Size = 100 - Sample Size = 100
Precinct Size = 200 - Sample Size = 200
Precinct Size = 300 - Sample Size = 300
Precinct Size = 400 - Sample Size = 400
Precinct Size = 500 - Sample Size = 500
Precinct Size = 600 - Sample Size = 600
Precinct Size = 700 - Sample Size = 700
Precinct Size = 800 - Sample Size = 800
Precinct Size = 900 - Sample Size = 900
Precinct Size = 1,000 - Sample Size = 999
Precinct Size = 1,100 - Sample Size = 1,099
Precinct Size = 1,100 - Sample Size = 1,099
Precinct Size = 1,200 - Sample Size = 1,199

Conversely, at a 99% confidence level in a precinct of 400, we find the following Margin of Error: [5]

MARGIN OF ERROR AT 99% CONFIDENCE LEVEL - POPULATION = 400								
Pct Sample	Sample Size	Margin of Error	Pct Sample	Sample Size	Margin of Error	Pct Sample	Sample Size	Margin of Error
5	20	28.15%	40	160	7.91%	75	300	3.73%
10	40	19.37%	45	180	7.14%	80	320	3.23%
15	60	15.37%	50	200	6.46%	85	340	2.71%
20	80	12.92%	55	220	5.84%	90	360	2.15%
25	100	11.19%	60	240	5.27%	95	380	1.48%
30	120	9.86%	65	260	4.74%	100	400	0.00%
35	140	8.80%	70	280	4.23%			

This graph shows the relationship between Margin of Error and Sample size for a finite population of 400:



[5] There are a number of available online calculators available to compute sample size and margin of error at both the 95% and 99% confidence levels. Try Here: http://www.surveysystem.com/sscalc.htm

The conclusion we can draw from the foregoing is simply that even at a fairly large sample size (75%) the Margin of Error of 0.1% is still substantial, even at a 99% confidence level (±3.73%). This margin of error creates a range that is outside the parameters of any reasonable audit trigger(s). For this reason alone, sampling of ballots to detect potential errors or fraud is contra-indicated as a reasonable methodology. While there is a strong argument for using statistically relevant samples over an entire population of voters such as city-wide or county-wide, this does not fall within range of solutions that can be applied at the precinct level and effectuated on the day of election.

In short, the sample size needed to achieve the desired confidence level and margin of error is 100%. Random sampling of ballots to detect error or fraud doesn't work!

Other Considerations

The voting industry has spent millions of dollars in lobbying legislators and boards of elections throughout the United States. Their message is that voting machines are a safe, secure, reliable and inexpensive means of capturing and reporting votes. There is ample evidence that they are ***none*** of these. A report on the security issues of Diebold Optical Scanning Machines was issued by Black Box Voting, a non-profit, non-partisan Washington-based organization: http://www.blackboxvoting.org/BBVreport.pdf on July 5, 2005. This report outlines the near total lack of security and the "open doors" left in the system which allow hackers, both inside and outside the electoral system to easily tamper with results. Of especial interest is their conclusion that such tampering can go ***undetected*** by election officials. What this means is that the security of these and other systems is so poor that no trail is left and thus "triggers" that might indicate the necessity for audit or recount simply will not be initiated. Additional discussions are here: http://www.bbvforums.org/forums/messages/1954/5921.html.

"I've been saying all along that my biggest fear is that someone would program a machine to give a wrong answer. If that were to happen, the machine would still work fine - we just wouldn't know it."

Avi Rubin - Professor, Computer Science - Johns Hopkins University

Many other organizations have documented fraud in the 2000 and 2004 elections all across the country, not just in Florida and Ohio. A worthwhile site is: http://www.votersunite.org. A new report from them, *Myth Breakers*, is available in PDF format. While the report is 70 pages, it's available free for download and is worth the reading. http://www.votersunite.org/MB2.pdf .

For information on the vast fraud which resulted in the stealing of the Ohio election in 2004, the report, *Preserving Democracy: What Went Wrong in Ohio*, issued January 5, 2005, by the House Judiciary Committee Democratic staff is required reading. It's available (again in PDF Format) online by going here: http://www.house.gov/judiciary_democrats/ohiostatusrept1505.pdf. 102 pages, but again worth the effort for exceptional documentation of election fraud in all its aspects, not just the failure of machines to do the job.

More Articles and Links

There are a number of shorter articles that can give you a flavor of how private companies like Diebold, ES&S and Sequoia influence election methods, voting and tallying:

How A Private Company Counts Our Votes On Election Night - by Christopher Bollyn
http://www.votefraud.org/how_a_private_company_counts_our_votes.htm

The Big Fix 2004, How to Fix a Presidential Election - by Daniel Hopsicker
http://www.madcowprod.com/mc6912004.html

Pandora's Black Box: Did It Really Count Your Vote? - by Phil O'Halloran,
http://www.votefraud.org/relevance_o'halloran_pandora's_box.htm

Inside A U.S. Election Vote Counting Program - By Bev Harris
http://www.votefraud.org/inside_election_program.htm

A House Without Doors – Voter Fraud in America - by James J. Condit, Jr.
http://www.votefraud.org/chronicles_house_without_doors.htm

Eternal Vigilance: Why America Must Say NO! - Right Now - To Electronic Voting,
http://www.libertyforall.net/2005/may1/Vigilance.html (Original, Pt. 1 of 3)
http://www.libertyforall.net/2005/may14/Eternal_Vigilance.htm (Original, Pt. 2 of 3)
http://www.libertyforall.net/2005/june14/RedFlags.htm (Original, Pt. 3 of 3)

National Ballot Integrity Project
http://www.ballotintegrity.org/

The above links include dozens of additional links that will get you started into an investigatory world that's been swirling about America for the past decade. A few hours of poking about should allow you to become relatively well-versed in the nature and scope of the myriad of problems associated with mechanical/electronic voting machines.

Fixes Don't Work

The Dec. 7th conference in Washington, D.C., Voting 2004: A Report to the Nation on America's Election Process, sponsored by Common Cause, The Century Foundation, and LCCR (Leadership Conference on Civil Rights) examined ways to "fix" the voting process,. They adopted the position that the problem is to fix the process and the machines. The conference endorsed the VerifiedVoting.org and Congressman Rush Holt's (D-NJ) prescription for voting integrity. It is beyond worthless.

The problem is that it gives people false hope, instead of a sensible solution. Holt's legislation calls for ballot printers and audits. First, that leaves the machines in the voting process - ready, willing, and able to malfunction, break down, or not show up - causing chaos and confusion. Ballot printers won't fix that. Second, it proposes spot audits, which leaves the counting of ballots in the hands of the very election officials who prove with each new election how truly inept or completely evil they really are. And third, the only time paper ballots will be counted is in case of a "close" election, ensuring that perpetrators of vote fraud will steal a sufficient number of votes to avoid triggering a recount. [6]

Conclusion

Sampling of voting in individual precincts won't do the job of securing voting accuracy at the precinct level. Even sampling of larger populations is unlikely to uncover error or fraud. The entire process of electronic vote gathering and compilation is rife with opportunities for hackers of even moderate capabilities to corrupt the electoral process.

The solution is to restore complete transparency to the system where every qualified voter has an opportunity to vote, using a paper ballot that is publicly deposited and publicly counted, with the results for each and every precinct publicly posted.

NOTE: The foregoing is not intended to aspire to a scholarly discourse on sampling methods, but rather to provide a straight-forward discussion of the problems inherent in statistical sampling methods because of the built-in margin of error, even at relatively high levels confidence. Our votes are vitally important in securing democracy in America, we deserve no less than to have every vote counted, accurately and quickly.

[6] This discussion is taken directly from an article***,*** *Voting Rights Groups 'Block' Talk of Machine-Free Elections* **-** by Lynn Landes: http://www.ecotalk.org/OpenDebate.htm

Further Notes on Audit Triggers for Recounts

There has been considerable discussion on techniques for sampling precincts at the city/county level to determine if error or fraud has occurred. The National Election Data Archive Project (NEDA) has produced a paper regarding the probability of finding a miscounted precinct with both 2% and 5% audits of county-wide elections, *How Can Independent Paper Audits Detect and Correct Vote Miscounts*, NEDA – July 30, 2005. [7]

This paper is easily available in PDF format at: http://www.electionarchive.org/ucvAnalysis/US/paper-audits/Paper_Audits.pdf The paper discusses the methodology for auditing paper ballots in a small number of precincts in order to detect errors. This method requires that a complete (100%) audit of the ballots be conducted in a number of precincts county-wide. While this method can be used, it requires that a hand count of paper ballots be conducted immediately upon poll closing and before ballots are transmitted to the central counting location.

In the City of Chicago, a 2% audit of precincts would detect at least one miscounted precinct 99.99% of the time if 54 precincts were audited and at least 406 (15%) precincts were corrupted. [8] A 5% audit would require hand counting in 135 precincts and would detect at least one miscounted precinct 99.92% of the time if at least 135 precincts (5%) were corrupted. This methodology would require the counting of 20,574 votes (assuming the average precinct size of 381) for a 2% sample and 51,435 votes for a 5% sample.

Results are similar for Suburban Cook County where a 2% sample would yield at least one miscounted precinct 99.98% of the time with an audit of 48 precincts, assuming that at least 360 (15%) were corrupted. A 5% sample would require hand counting 120 precincts and would uncover a miscounted precinct 99.62% of the time if 5% of precincts were corrupt (120). [9] This would require hand counting of 20,160 votes at the 2% level and 50,400 at the 5% level, assuming an average actual precinct vote of 420.

Therefore, an audit (100% hand count) of a relatively small number of precincts (135 or 120) would yield a fairly high probability that at least one miscounted precinct could be detected, assuming that corruption was reasonably widespread (15% of precincts corrupt). Further, hand counting of approximately 50,000 ballots in each of the City of Chicago and Suburban Cook County would have a reasonable chance to detect error or fraud assuming that such occurred in more than a few precincts.

This methodology has a good opportunity for success because the relatively low number of voters per precinct (400). For example, 5% of the votes in an average precinct would be 20. If more than 400 precincts were corrupted by 20 votes, that would be 8,000 corrupt votes that would probably trigger a recount. However, even a relatively small level of irregularities over a large number of precincts has little opportunity of going undetected as the probability of detection approaches 100% as the percentage of corrupt precincts rises. Essentially, with a 5% sampling of precincts, the number of precincts that could be tampered with must be less than 5% to avoid detection, assuming that the precincts to be audited are truly random within the populations of 2,402 or 2,709 precincts.

Secondary Conclusion

While sampling within precincts is inappropriate for determining the accuracy of the precinct count, a sampling (100%) of 5% of precincts within a county has a reasonably degree of probability to uncover a miscounted precinct within the entire population, assuming that such miscounts are sufficiently widespread.

[7] See Appendix E
[8] See Appendix C
[9] See Appendix D

Robert A. Wilson – 635 Chicago Ave – Suite 127 – Evanston IL 60202 – bobwilson@consultant.com

2004 GENERAL ELECTION - CHICAGO - PRESIDENT/VICE PRESIDENT **Appendix A**

Ward	Total Votes	Kerry/Edwards		Bush/Cheney		Badnarik/Campagna			
		Votes	Pct	Votes	Pct	Votes	Pct	Precincts	Avg/Pcnt
1	20438	16282	79.67%	3985	19.50%	171	0.84%	40	511
2	23256	19602	84.29%	3516	15.12%	138	0.59%	59	394
3	16596	15709	94.66%	823	4.96%	64	0.39%	63	263
4	23463	22174	94.51%	1202	5.12%	87	0.37%	57	412
5	23573	22336	94.75%	1137	4.82%	100	0.42%	55	429
6	27325	26549	97.16%	726	2.66%	50	0.18%	66	414
7	23307	22433	96.25%	800	3.43%	74	0.32%	65	359
8	28792	27920	96.97%	814	2.83%	58	0.20%	70	411
9	23377	22574	96.56%	744	3.18%	59	0.25%	58	403
10	16503	12602	76.36%	3797	23.01%	104	0.63%	49	337
11	17176	12095	70.42%	4981	29.00%	100	0.58%	50	344
12	6684	5293	79.19%	1345	20.12%	46	0.69%	29	230
13	18811	13218	70.27%	5475	29.11%	118	0.63%	54	348
14	9830	7694	78.27%	2104	21.40%	32	0.33%	37	266
15	16802	15737	93.66%	1008	6.00%	57	0.34%	52	323
16	14319	13513	94.37%	760	5.31%	46	0.32%	47	305
17	23391	22800	97.47%	529	2.26%	62	0.27%	64	365
18	25755	22623	87.84%	3048	11.83%	84	0.33%	64	402
19	30401	19308	63.51%	10936	35.97%	157	0.52%	67	454
20	17269	16637	96.34%	576	3.34%	56	0.32%	55	314
21	29986	29098	97.04%	815	2.72%	73	0.24%	76	395
22	7810	6811	87.21%	966	12.37%	33	0.42%	30	260
23	22360	13737	61.44%	8497	38.00%	126	0.56%	56	399
24	20703	20083	97.01%	565	2.73%	55	0.27%	59	351
25	12750	10170	79.76%	2488	19.51%	92	0.72%	31	411
26	14397	11783	81.84%	2523	17.52%	91	0.63%	63	229
27	21005	18280	87.03%	2627	12.51%	98	0.47%	66	318
28	19858	19109	96.23%	700	3.53%	49	0.25%	73	272
29	21445	20065	93.56%	1331	6.21%	49	0.23%	54	397
30	11594	8673	74.81%	2854	24.62%	67	0.58%	43	270
31	11734	8718	74.30%	2949	25.13%	67	0.57%	51	230
32	27955	19773	70.73%	7923	28.34%	259	0.93%	52	538
33	13694	10624	77.58%	2957	21.59%	113	0.83%	34	403
34	28316	27476	97.03%	783	2.77%	57	0.20%	73	388
35	15115	12007	79.44%	2973	19.67%	135	0.89%	36	420
36	21049	13277	63.08%	7652	36.35%	120	0.57%	55	383
37	19351	18259	94.36%	1052	5.44%	40	0.21%	58	334
38	18537	11979	64.62%	6436	34.72%	122	0.66%	53	350
39	17709	11825	66.77%	5748	32.46%	136	0.77%	47	377
40	18144	13998	77.15%	4034	22.23%	112	0.62%	49	370
41	27818	14625	52.57%	13017	46.79%	176	0.63%	71	392
42	32530	20604	63.34%	11696	35.95%	230	0.71%	69	471
43	29915	19571	65.42%	10134	33.88%	210	0.70%	67	446
44	31131	22553	72.45%	8324	26.74%	254	0.82%	50	623
45	22291	13790	61.86%	8344	37.43%	157	0.70%	53	421
46	23963	19042	79.46%	4750	19.82%	171	0.71%	43	557
47	27547	21515	78.10%	5818	21.12%	214	0.78%	51	540
48	22538	18318	81.28%	4071	18.06%	149	0.66%	56	402
49	17196	14811	86.13%	2260	13.14%	125	0.73%	44	391
50	17369	11823	68.07%	5463	31.45%	83	0.48%	45	386
Total	**1032878**	**839496**	**81.28%**	**188056**	**18.21%**	**5326**	**0.52%**	**2709**	**381**

Source: Chicago Board of Elections Website: 3 Aug 2005

2004 GENERAL ELECTION - COOK COUNTY - PRESIDENT/VICE PRESIDENT								**Appendix B**	
	Total	Kerry/Edwards		Bush/Cheney		Badnarik/Campagna			
Township	Votes	Votes	Pct	Votes	Pct	Votes	Pct	Precincts	Avg/Pcnt
Barrington	7418	2480	33.43%	4886	65.87%	52	0.70%	14	530
Berwyn	15743	10062	63.91%	5559	35.31%	122	0.77%	49	321
Bloom	36863	24642	66.85%	11978	32.49%	243	0.66%	94	392
Bremen	45158	28557	63.24%	16358	36.22%	243	0.54%	105	430
Calumet	7035	6060	86.14%	957	13.60%	18	0.26%	18	391
Cicero	14994	9772	65.17%	5119	34.14%	103	0.69%	53	282
Elk Grove	33080	16593	50.16%	16227	49.05%	260	0.79%	77	430
Evanston	35656	29142	81.73%	6245	17.51%	269	0.75%	71	502
Hanover	26316	13219	50.23%	12910	49.06%	187	0.71%	59	446
Lemont	9125	3541	38.81%	5523	60.53%	61	0.67%	15	608
Leyden	29627	15894	53.65%	13505	45.58%	228	0.77%	91	326
Lyons	45416	23480	51.70%	21621	47.61%	315	0.69%	119	382
Maine	54023	28746	53.21%	24926	46.14%	351	0.65%	137	425
New Trier	32196	18120	56.28%	13877	43.10%	199	0.62%	72	447
Niles	45109	28826	63.90%	16080	35.65%	203	0.45%	106	426
Northfield	43597	22826	52.36%	20546	47.13%	225	0.52%	86	507
Norwood Park	10443	5574	53.38%	4807	46.03%	62	0.59%	31	337
Oak Park	27714	21751	78.48%	5782	20.86%	181	0.65%	66	420
Orland	46369	19956	43.04%	26189	56.48%	224	0.48%	90	515
Palatine	45247	20135	44.50%	24787	54.78%	325	0.72%	94	481
Palos	23909	11251	47.06%	12500	52.28%	158	0.66%	53	451
Proviso	62095	44374	71.46%	17404	28.03%	317	0.51%	162	383
Rich	35208	27022	76.75%	8040	22.84%	146	0.41%	72	489
River Forest	6032	3329	55.19%	2663	44.15%	40	0.66%	15	402
Riverside	8049	4250	52.80%	3736	46.42%	63	0.78%	19	424
Schaumburg	48883	25259	51.67%	23339	47.74%	285	0.58%	106	461
Stickney	12602	7204	57.17%	5316	42.18%	82	0.65%	41	307
Thornton	72253	56831	78.66%	15114	20.92%	308	0.43%	178	406
Wheeling	66108	33085	50.05%	32625	49.35%	398	0.60%	142	466
Worth	62642	32947	52.60%	29248	46.69%	447	0.71%	167	375
Cook County	**1,008,91**	**594,928**	**58.97%**	**407,867**	**40.43%**	**6,115**	**0.61%**	**2,402**	**420**

Appendix C

Chance of Finding a Corrupted Precinct in Chicago with 2% Audit

Total Number of Precincts	Hypothetical Number of Corrupted Precincts	Percent Corrupted Precincts	Audited Precincts for 2% Audit	Expected Value	Standard Deviation	Chance of Finding a Corrupted Precinct
2709	1355	50.00%	54	27.09	3.68	100.00%
2709	1219	45.00%	54	24.38	3.66	100.00%
2709	1084	40.00%	54	21.67	3.61	100.00%
2709	948	35.00%	54	18.96	3.51	100.00%
2709	813	30.00%	54	16.25	3.37	100.00%
2709	677	25.00%	54	13.55	3.19	100.00%
2709	542	20.00%	54	10.84	2.94	100.00%
2709	406	15.00%	54	8.13	2.63	99.99%
2709	271	10.00%	54	5.42	2.21	99.67%
2709	135	5.00%	54	2.71	1.60	93.85%
2709	108	4.00%	54	2.17	1.44	89.13%
2709	81	3.00%	54	1.63	1.26	80.91%
2709	27	1.00%	54	0.54	0.73	42.09%
2709	14	0.50%	54	0.27	0.52	23.08%
2709	3	0.10%	54	0.05	0.23	3.95%
2709	1	0.05%	54	0.03	0.16	1.99%

Chance of Finding a Corrupted Precinct in Chicago with 5% Audit

Total Number of Precincts	Hypothetical Number of Corrupted Precincts	Percent Corrupted Precincts	Audited Precincts for 5% Audit	Expected Value	Standard Deviation	Chance of Finding a Corrupted Precinct
2709	1355	50.00%	135	67.73	5.82	100.00%
2709	1219	45.00%	135	60.95	5.79	100.00%
2709	1084	40.00%	135	54.18	5.70	100.00%
2709	948	35.00%	135	47.41	5.55	100.00%
2709	813	30.00%	135	40.64	5.33	100.00%
2709	677	25.00%	135	33.86	5.04	100.00%
2709	542	20.00%	135	27.09	4.66	100.00%
2709	406	15.00%	135	20.32	4.16	100.00%
2709	271	10.00%	135	13.55	3.49	100.00%
2709	135	5.00%	135	6.77	2.54	99.92%
2709	108	4.00%	135	5.42	2.28	99.64%
2709	81	3.00%	135	4.06	1.99	98.51%
2709	27	1.00%	135	1.35	1.16	75.02%
2709	14	0.50%	135	0.68	0.82	48.63%
2709	3	0.10%	135	0.14	0.37	9.72%
2709	1	0.05%	135	0.07	0.26	4.98%

Note: Data used:

Chicago has	2709	precincts
State of Illinois has	11700	precincts
Chicago has	23.15%	of Illinois precincts
Illinois has	102	counties
2% sample of IL precincts is	234	precincts
5% sample of IL precincts is	585	precincts
Precincts selected randomly for 2%	132	precincts
Precincts selected randomly for 5%	483	precincts
CHI random 2% sample will pull	54.18	precincts
Rounded to	54	precincts
Average number of voters per preci	381	voters
Number of voters sampled in Cook (	20574	voters
And random 5% sample will pull	135.5	precincts
Rounded to	135	precincts
Average number of voters per preci	381	voters
Number of voters sampled in Cook (	51435	voters
Number of voters in Chicago	1,032,878	voters (2004 -President)

Appendix D

Chance of Finding a Corrupted Precinct in Cook County with 2% Audit						
Total Number of Precincts	**Hypothetical Number of Corrupted Precincts**	**Percent Corrupted Precincts**	**Audited Precincts for 2% Audit**	**Expected Value**	**Standard Deviation**	**Chance of Finding a Corrupted Precinct**
2402	1201	50.00%	48	24.02	3.47	100.00%
2402	1081	45.00%	48	21.62	3.45	100.00%
2402	961	40.00%	48	19.22	3.40	100.00%
2402	841	35.00%	48	16.81	3.31	100.00%
2402	721	30.00%	48	14.41	3.18	100.00%
2402	601	25.00%	48	12.01	3.00	100.00%
2402	480	20.00%	48	9.61	2.77	100.00%
2402	360	15.00%	48	7.21	2.47	99.96%
2402	240	10.00%	48	4.80	2.08	99.39%
2402	120	5.00%	48	2.40	1.51	91.67%
2402	96	4.00%	48	1.92	1.36	86.16%
2402	72	3.00%	48	1.44	1.18	77.13%
2402	24	1.00%	48	0.48	0.69	38.54%
2402	12	0.50%	48	0.24	0.49	21.56%
2402	2	0.10%	48	0.05	0.22	3.96%
2402	1	0.05%	48	0.02	0.15	2.00%

Chance of Finding a Corrupted Precinct in Cook County with 5% Audit						
Total Number of Precincts	**Hypothetical Number of Corrupted Precincts**	**Percent Corrupted Precincts**	**Audited Precincts for 5% Audit**	**Expected Value**	**Standard Deviation**	**Chance of Finding a Corrupted Precinct**
2402	1201	50.00%	120	60.05	5.48	100.00%
2402	1081	45.00%	120	54.05	5.45	100.00%
2402	961	40.00%	120	48.04	5.37	100.00%
2402	841	35.00%	120	42.04	5.23	100.00%
2402	721	30.00%	120	36.03	5.02	100.00%
2402	601	25.00%	120	30.03	4.75	100.00%
2402	480	20.00%	120	24.02	4.38	100.00%
2402	360	15.00%	120	18.02	3.91	100.00%
2402	240	10.00%	120	12.01	3.29	100.00%
2402	120	5.00%	120	6.01	2.39	99.82%
2402	96	4.00%	120	4.80	2.15	99.34%
2402	72	3.00%	120	3.60	1.87	97.64%
2402	24	1.00%	120	1.20	1.09	70.95%
2402	12	0.50%	120	0.60	0.77	46.01%
2402	2	0.10%	120	0.12	0.35	9.74%
2402	1	0.05%	120	0.06	0.24	5.00%

Note: Data used:

Cook County has	2402	precincts
State of Illinois has	11700	precincts
So Cook County has	20.53%	of Illinois precincts
Illinois has	102	counties
2% sample of IL precincts is	234	precincts
5% sample of IL precincts is	585	precincts
Precincts selected randomly for 2%	132	precincts
Precincts selected randomly for 5%	483	precincts
So random 2% sample will pull	48.04	precincts
Rounded to	48	precincts
Average number of voters per preci	420	voters
Number of voters sampled in Cook (	20160	voters
And random 5% sample will pull	120.1	precincts
Rounded to	120	precincts
Average number of voters per preci	420	voters
Number of voters sampled in Cook (	50400	voters
Number of voters in Cook County	1,008,910	voters

How Can Independent Paper Audits Detect and Correct Vote Miscounts?

Routine independent audits of vote counts in all elections in a small percentage of randomly selected precincts[1] have a surprisingly good chance of detecting miscounts.

Why Independently Audit Vote Counts?

For over a decade, states have counted a majority of votes electronically, without performing routine independent audits to ensure accuracy or protect from inside embezzlement of votes. America can and must do better.

It is not enough to require voter verifiable paper records of ballots. The paper records must be easily and "independently" auditable and routinely audited by persons other than the voting machine vendor or other insiders within the election system.

How to Independently Audit Vote Counts?

In particular, if the voter-verifiable paper records are counted in a small percentage of randomly selected precincts immediately when polls close, before removing ballots from the precinct, there is a high probability of detecting existing errors in vote counts.[2] If discrepancies are found, a county-wide recount can be triggered. Additional funding may need to be allocated in order to routinely perform independent audits of vote counts.

What is the Probability of Finding a Miscounted Precinct with 2% or 5% Audits?

To calculate the probability of detecting at least one miscounted precinct in any county or township, you have to know:
1. How many precincts are there in the county?
2. What percentage of precincts do we hypothetically assume have been miscounted?
3. What percentage of precincts will be randomly selected for paper audits?

In the tables on the next page, probabilities for finding at least one corrupted precinct are calculated, based on 1000 precincts with varying hypothetical percentages of corrupted precincts, using 2% and 5% independent audits.[3]

[1] "Randomly selected" is a specifically designed process. An expert needs to be hired to design a truly random process for selecting precincts. Having election officials select the precincts does not qualify as "random" selection.
[2] A random audit, in combination with a mathematical analysis of detailed election results, would ensure election integrity because if we assume that the proportion of precincts with vote miscounts is small, then the amount of error in the miscounted precincts would be high enough to be detectable via mathematical analysis of election results. i.e. The analysis of election results would be likely to catch errors that were under the design threshold of the paper audit.
[3] Details of the calculation are in Appendix A, available in the on-line version. USCountVotes has made a spreadsheet available to find the probability distribution for detecting miscounted precincts for a particular county's number of precincts or audit rate. USCV's AuditCalculator.xls spreadsheet is available at http://ElectionArchive.org/ucvAnalysis/US/paper-audits/

Table 1: Chance of Finding a Corrupted Precinct with a 2% Audit of 1000 Precincts

Total Number of Precincts	Hypothetical Number of Corrupted Precincts	Percent Corrupted Precincts	Number of Audited Precincts for 2% Audit	Expected Value	Standard Deviation	Chance of Finding a Corrupted Precinct
1000	500	50.00%	20	10.00	2.24	100.00%
1000	450	45.00%	20	9.00	2.22	100.00%
1000	400	40.00%	20	8.00	2.19	100.00%
1000	350	35.00%	20	7.00	2.13	99.98%
1000	300	30.00%	20	6.00	2.05	99.93%
1000	250	25.00%	20	5.00	1.94	99.70%
1000	200	20.00%	20	4.00	1.79	98.90%
1000	150	15.00%	20	3.00	1.60	96.25%
1000	100	10.00%	20	2.00	1.34	88.10%
1000	50	5.00%	20	1.00	0.97	64.51%
1000	40	4.00%	20	0.80	0.88	56.15%
1000	30	3.00%	20	0.60	0.76	45.94%
1000	10	1.00%	20	0.20	0.44	18.37%
1000	5	0.50%	20	0.10	0.32	9.63%
1000	1	0.10%	20	0.02	0.14	2.00%
1000	0.5	0.05%	20	0.01	0.10	0.00%

Table 1 shows that a 2% audit has a probability of over 95% of finding a corrupted precinct if at least 15% or 150 out of 1000 precincts are corrupted.

Table 2: Chance of Finding a Corrupted Precinct with a 5% Audit of 1000 Precincts

Total Number of Precincts	Hypothetical Number of Corrupted Precincts	Percent Corrupted Precincts	Number of Audited Precincts for 5% Audit	Expected Value	Standard Deviation	Chance of Finding a Corrupted Precinct
1000	500	50.00%	50	25.00	3.54	100.00%
1000	450	45.00%	50	22.50	3.52	100.00%
1000	400	40.00%	50	20.00	3.46	100.00%
1000	350	35.00%	50	17.50	3.37	100.00%
1000	300	30.00%	50	15.00	3.24	100.00%
1000	250	25.00%	50	12.50	3.06	100.00%
1000	200	20.00%	50	10.00	2.83	100.00%
1000	150	15.00%	50	7.50	2.52	99.98%
1000	100	10.00%	50	5.00	2.12	99.55%
1000	50	5.00%	50	2.50	1.54	92.80%
1000	40	4.00%	50	2.00	1.39	87.68%
1000	30	3.00%	50	1.50	1.21	79.03%
1000	10	1.00%	50	0.50	0.70	40.27%
1000	5	0.50%	50	0.25	0.50	22.66%
1000	1	0.10%	50	0.05	0.22	5.00%
1000	0.5	0.05%	50	0.03	0.16	0.00%

Table 2 shows that a 5% audit has a probability of over 90% (the lower end of customary “statistical significance”) of finding a corrupted precinct if at least 5% or 50 out of the 1000 are corrupted.

Appendix A:

The probability estimates are based on a "Hypergeometric" distribution which determines the probability of finding:

a) **x** (target corrupted precincts) We let x = 0 to find the probability of detecting no corrupted precincts.
b) in an overall sample of **n** (Number of Audited Precincts) which in this case is 20 (for 2% audit) or 50 (for 5% audit),
c) when there are **X** (Hypothetical Number of Corrupted Precincts) which in this case ranges from 50% to 0.05% of 1000,
d) out of **N** (Total Number of Precincts) which in this case is 1000.

This distribution is calculated using the Excel Function:

$$\text{HYPGEOMDIST}(x, n, X, N) = \frac{\binom{X}{x}\binom{N-X}{n-x}}{\binom{N}{n}}$$

The hypergeometric function assumes all individual "picks" are random but adjusts this random probability for each pick. The first row in Table 1 for example, assumes a 50% probability that the first of 20 picks will be corrupted, a 499/999 probability that the second precinct chosen will be corrupted, and so on for all 20 picks.

The probability that one or more of the 20 precincts will be a corrupted precinct is 1 or 100% minus the probability that none of the 20 precincts will be corrupted. So, the probability that *at least one* of the 20 is corrupted which equals:

$$1 - \text{HYPGEOMDIST}(0, n, X, N) = 1 - \frac{\binom{X}{0}\binom{N-X}{n}}{\binom{N}{n}}$$

For any county, the probability distribution will depend on the exact values of n, X, and N, so that it is best to recreate tables that are specific to the number of precincts of the county for which the audit probability estimates are being made.

A free spreadsheet calculator for determining the probabilities of detecting one or more miscounted precincts is available on the USCV web site:
http://ElectionArchive.org/ucvAnalysis/US/paper-audits/

Prepared by
National Election Data Archive Project (NEDA) - ElectionArchive.org
501-3c status pending
Contact: Kathy Dopp kathy@uscountvotes.org 435-608-1382

The National Election Data Archive is a scientific research project sponsored by USCountVotes whose mission is to objectively investigate the accuracy of elections in America through the creation and analysis of a database containing precinct-level election data for the entire United States.