



Papers in this series are intended for eventual publication.

Copyrights remain with the authors and/or their publishers. Reproduction, posting to web pages, electronic bulletin boards or other electronic archives is prohibited without express consent of the copyright holders.

Comments or questions directed to the author(s) are most appreciated and will be acknowledged accordingly.

---

**Graduate Division, School of Arts & Sciences  
Center for Organizational Dynamics  
Working Paper #04-10      November 21, 2004**

---

Note on this revision: The thesis of the initial drafts of this paper that were widely circulated on the Internet, was that exit poll discrepancy could not have been due to chance or random error. My purpose was to raise as a legitimate question, "What caused it?" In this version, I begin to try to answer that question.

I'm releasing this paper in two parts, so as to replace the early drafts as soon as possible. Part I is to clean up the previous draft. Part II will be new analysis.

Reading note: To help non-academic readers, I have put in much more explanatory material than typical —, especially in the statistical analysis sections (also a glossary in Appendix B). My apologies to those for whom these explanations are superfluous. Just skip over those sections.

---

---

## **The Unexplained Exit Poll Discrepancy: Part I**

---

**Steven F. Freeman, PhD**

[stfreema@sas.upenn.edu](mailto:stfreema@sas.upenn.edu)

Most Americans who had listened to radio or surfed the Internet on Election Day this year, sat down to watch election night coverage expecting that John Kerry had been elected President. Exit polls showed him ahead in nearly every battleground state, in many cases by sizable margins. Although pre-election day polls indicated the race dead even or Bush slightly ahead, two factors seemed to explain Kerry's edge: turnout was extraordinary high, which is generally good for Democrats,<sup>1</sup> and as in every US Presidential election with an incumbent over the past quarter-century,<sup>2</sup> undecided voters broke heavily toward the challenger.<sup>3</sup>

---

1 Jack Citrin, Eric Schickler & John Sides, "What If everyone voted? Simulating the impact of increased turnout in senate elections" *American Journal of Political Science*, 2003, 47 (1) 75-90: Nonvoters are generally more Democratic than voters. Democratic party candidates generally benefit from higher turnout because the increase comes disproportionately from voters in socio-economic groups that traditionally vote Democratic.

2 Guy Molyneux, "The Big Five-Oh", *The American Prospect Online*, Oct 1, 2004: There have been four incumbent presidential elections in the past quarter-century. On average, the incumbent comes in half a point below his final poll result; challengers exceed their final poll result by an average of 4 points.

3 Even the final "corrected" exit poll data presented on the CNN website — more on such corrected data later — indicate that those who decided in the last three days chose Kerry over Bush 55% - 42%.  
<<http://www.cnn.com/ELECTION/2004/pages/results/states/US/P/00/epolls.0.html>> (Thursday Nov 14, 2004)

But then, in key state after key state, counts showed very different numbers than the polls predicted; and the differentials were all in the same direction. The first shaded column in Table 1 shows the differential between the major candidates' predicted (exit poll) percentages of the vote; the next shaded column shows the differential between their *tallied* percentages of the vote. The final shaded column reveals the "shift." In ten of the eleven consensus battleground states,<sup>4</sup> the tallied margin differs from the predicted margin, and in every one, the shift favors Bush.

**Table 1<sup>5</sup>: Predicted vs. tallied percentages in battleground states**

	Sample size	Bush predicted	Kerry predicted	Predicted differential	Bush tallied	Kerry tallied	Tallied differential	Tallied vs. predicted
Colorado	2515	49.9%	48.1%	Bush 1.8	52.0%	46.8%	Bush 5.2	Bush 3.4
Florida	2846	49.8%	49.7%	Bush 0.1	52.1%	47.1%	Bush 5.0	Bush 4.9
Iowa	2502	48.4%	49.7%	Kerry 1.3	50.1%	49.2%	Bush 0.9	Bush 2.2
Michigan	2452	46.5%	51.5%	Kerry 5.0	47.8%	51.2%	Kerry 3.4	Bush 1.6
Minnesota	2178	44.5%	53.5%	Kerry 9.0	47.6%	51.1%	Kerry 3.5	Bush 5.5
Nevada	2116	47.9%	49.2%	Kerry 1.3	50.5%	47.9%	Bush 2.6	Bush 3.9
New Hampshire	1849	44.1%	54.9%	Kerry 10.8	49.0%	50.3%	Kerry 1.3	Bush 9.5
New Mexico	1951	47.5%	50.1%	Kerry 2.6	50.0%	48.9%	Bush 1.1	Bush 3.7
Ohio	1963	47.9%	52.1%	Kerry 4.2	51.0%	48.5%	Bush 2.5	Bush 6.7
Pennsylvania	1930	45.4%	54.1%	Kerry 8.7	48.6%	50.8%	Kerry 2.2	Bush 6.5
Wisconsin	2223	48.8%	49.2%	Kerry 0.4	49.4%	49.8%	Kerry 0.4	No dif

The media has largely ignored this discrepancy (although the Blogosphere has been abuzz), suggesting that the polls were either flawed, within normal sampling error, a statistical anomaly, or could otherwise be easily explained away. In Part I of this paper, I examine the validity of exit polls, the likelihood of sampling error, and the possibility of statistical anomaly and show that the exit poll discrepancy could not have been due to chance or random error. In Part II, I explore the explanations for what did cause it, and speculate on the big question on readers' minds, what this data can tell us about whether the count was correct.

## Exit Poll Data

The data I use for this paper are those posted on the CNN website Election Night. CNN had the data by virtue of membership in the National Election Pool (NEP), a consortium of news

<sup>4</sup> These eleven states are classified as battleground states based on being on at least two of three prominent lists: *Zogby*, *MSNBC*, and the *Washington Post*. (These eleven did in fact turn out to be the most competitive states in the election along with Oregon where Kerry won by 4%; in no other state was the winning margin within 7%.)

<sup>5</sup> Source: CNN website. Wednesday November 3, 2004 12:21 am.

organizations that had pooled resources to conduct a large-scale exit poll (as was done in the 2000 election). NEP, in turn, had contracted two respected firms, Edison Media Research and Mitofsky International,<sup>6</sup> to conduct the polls.

### Calibrated and Uncalibrated Exit Poll Data<sup>7</sup>

Part of the reason the issue went away for the media – and has become fodder for conspiracy theorists on the web – is secrecy and confusion about the data and what exactly is being characterized as the exit poll. If you go to the CNN website or any other website on which 2004 exit poll data is available, you'll see numbers very different from those that were released on Election Day. That's because the survey results originally collected and presented to subscribers were subsequently “corrected” to conform to official tallies.

The pollsters explain this as a natural process: the “uncalibrated” data were preliminary; once the counts come in, they recalibrate their original data on the assumptions that the count is correct, and that any discrepancies must have been due to imbalanced representation in their samples or some other polling error. The pollsters have taken great pains to argue that their polls were not designed to verify election results,<sup>8</sup> but rather to provide election coverage support to subscribers: as one piece of data (among many) that networks could use to “call” states and to explain voting patterns, i.e., who voted for whom, and why people voted as they did.

Whatever the merits of calibrating exit poll data, it confuses the issue of *why* the (uncalibrated) polls were so far off and why in the same direction. Although this calibration process may seem perfectly natural to NEP, it confuses nearly everyone else, even sophisticated analysts intimately involved in voting issues. The MIT-Caltech Voting Project, for example, issued a report concluding that exit poll data were consistent with state tallies and that there were no discrepancies based on voting method, including electronic voting systems. But they used this adjusted data to exonerate the process! In other words, they used data in which the count is

---

<sup>6</sup> Warren Mitofsky, the founder of Mitofsky International is credited with having invented the exit poll. David W. Moore, Senior Gallup Poll Editor, “New Exit Poll Consortium Vindication for Exit Poll Inventor” 10/11/2003

<sup>7</sup> To try to avoid reader confusion, I've included a glossary in Appendix B.

<sup>8</sup> Martin Plissner, “In Defense of Exit Polls: You just don't know how to use them” *Slate* Thursday, Nov. 4, 2004

assumed correct to prove that the count is correct. And, sadly, this report is being used to dismiss allegations that anything might be awry.<sup>9</sup>

It's an awful mistake, but it is understandable how they could make it. Few of us realized that this data is corrected. Neither the CNN website, nor any other site of which I'm aware, gives any indication that the data were anything other than what nearly all of us imagine exit poll data to be – data based (solely) on subjects surveyed leaving the polling place.

### Data Used in This Report

For this report, I use data that apparently are based solely on subjects surveyed leaving the polling place. These data were reportedly not meant to be released directly to the public,<sup>10</sup> and were reportedly available to late evening Election Night viewers only because a computer glitch prevented NEP from making updates sometime around 8:30 p.m. that night.<sup>11</sup> They were collected by Jonathon Simon, a former political survey research analyst, and are corroborated by saved screen shots (see Figure 1). I happened to have ten exit poll internet pages stored in my computer memory, and in each case, his figures are identical to mine. The numbers are also roughly consistent with those released elsewhere (Appendix C shows *Slate* numbers at 7:28 EST).

To derive the “predicted values” used in Tables 1 & 4, I combine the male and female vote, weighted for their percentage of the electorate. Ohio exit poll data (Figure 1) indicates that 51% of men and 53% of women voted for Kerry. Since the electorate is 47% male/53% female, Kerry's overall share of the exit poll was calculated as  $(51\% \times 47\%) + (53\% \times 53\%)$  or 52.1%.<sup>12</sup> Doing the same for calculations for other battleground states and comparing these numbers with final tallies (*New York Times*, Nov. 7), I completed the columns in Tables 1 & 4.

---

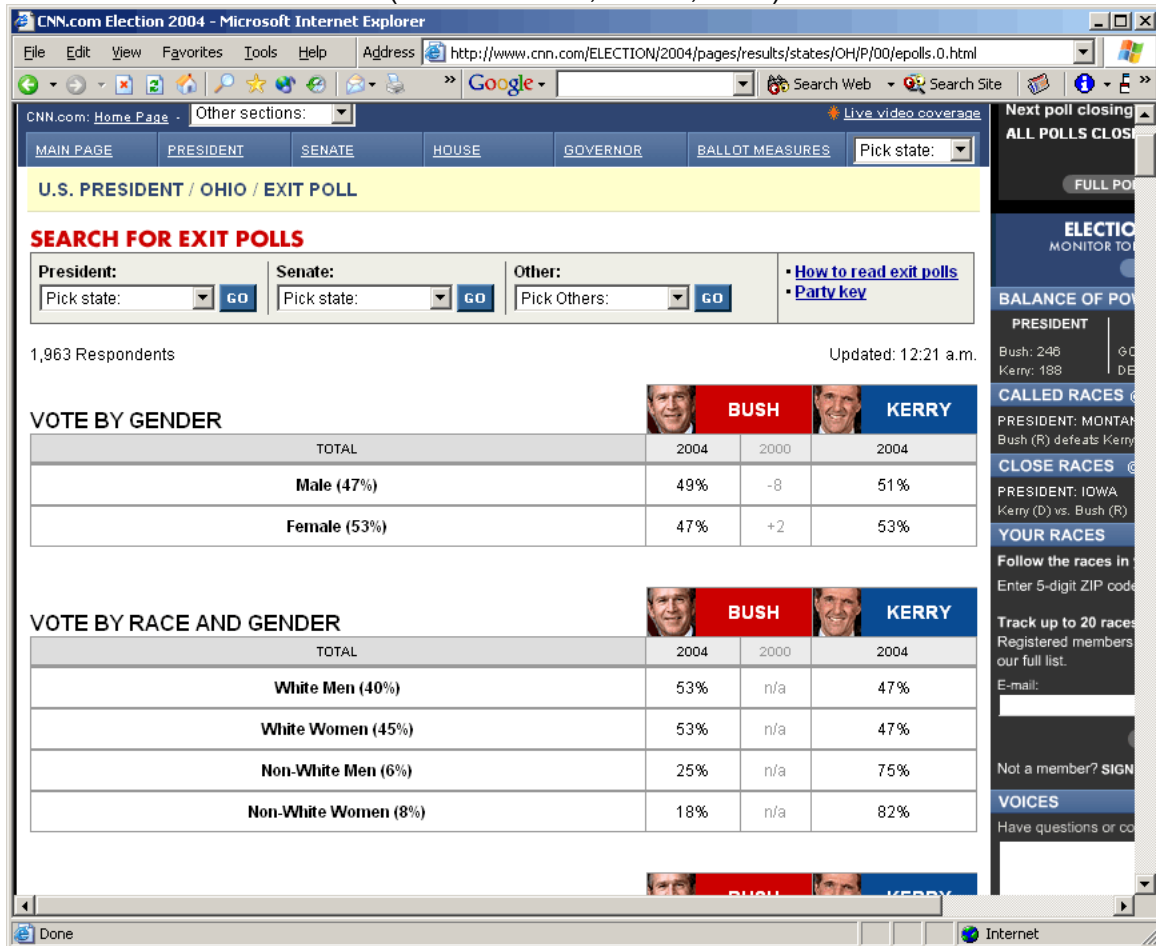
<sup>9</sup> Tom Zeller, Jr. "Vote Fraud Theories, Spread by Blogs, Are Quickly Buried," *New York Times* (Front page); John Schwartz, "Mostly Good Reviews for Electronic Voting," *New York Times*; Keith Olbermann MSNBC Countdown. All three on November 12, 2004.

<sup>10</sup> Martin Plissner, "In Defense of Exit Polls: You just don't know how to use them" *Slate* Thursday, Nov. 4, 2004

<sup>11</sup> Richard Morin, "New Woes Surface in Use of Estimates," *Washington Post*, Thurs, Nov. 4, 2004; Page A29

<sup>12</sup> Among the limitations of the CNN exit poll data is the lack of significant digits. Rounding errors mean that exit poll numbers for individual state analyses could be off by up to .5. This is unlikely because it comes from two groups, male and female, and it's unlikely that they are both rounded that much in the same direction. Regardless, the strength of the finding is such that even if all numbers had been rounded the full .5 in an unfavorable direction, the basic finding would still hold.

**Figure 1. CNN web page with apparently “uncorrected” exit poll data (12:21 am Wed, Nov. 3, 2004)**



**Is the Data Valid?**

Ruy Teixeira and others have rejected these data as unweighted, meaning that they have not been adjusted to appropriately weight demographic groups pollsters knowingly under- or over-sampled,<sup>13</sup> but that would seem very unlikely. NEP’s predecessor, Voter News Service, warns in bold letters in its 2000 Methodology statement never to use unweighted data (see Appendix A). It makes no sense *ever* to distribute unweighted data to anyone. Pollsters want to get it right. Their customers are depending on it. Broadcasters want to be alerted to probable outcomes, so as to plan their coverage accordingly (e.g., pre-writing stories so they could be completed shortly after

<sup>13</sup> Ruy Teixeira: <http://www.emergingdemocraticmajorityweblog.com/donkeyrising/>. Pollsters typically oversample minorities so that they have a sufficient sample size of important demographic groups, but then they negatively weight respondents in these groups to adjust for their actual percentage of the electorate.

poll-closings, assigning top reporters to winners' headquarters, being prepared for when concession or victory speeches might be forthcoming, etc...). In this case, subscribers were very surprised. Editors and network managers had to scramble, and journalists complained that they had to rewrite their lead stories.

It's likewise possible that the data were already partially calibrated to tallies, but given the news story and the abrupt change at 1:30 am, that too seems unlikely. If, in fact, the data already had been partially calibrated, however, it would mean that the pure exit poll numbers favored Kerry to an even greater extent.

In summary, I'd rather have NEP data, but no one is going to see those until well into 2005 (if then). That said, I believe this CNN data are good, and can be used to generate some highly suggestive findings.

### On (Uncorrected) Exit Polls

There are many challenges when conducting an exit poll, several of which potentially might have caused errors that would have resulted in Election Day discrepancies. I'll discuss these at length in a later section of this report (along with a discussion of potential count errors), but at this point, I want to validate curiosity and concern with this issue.

In general, we have reason to believe that exit polls, by which I mean *uncorrected* exit polls, are reasonably accurate survey instruments. Exit polls are surveys taken of representative respondents from the overall voting population. Both the logic behind them and experience suggest that these surveys should be able to predict overall results within statistical limits. It's relatively easy to get a representative sample, and there is no problem with figuring out who is actually going to vote or how they will vote.

In Germany, the minute the polls close, polling agencies release prognoses that have proven highly reliable. In the three most recent national elections, poll percentages diverged from official counts by only **0.26%** on average (Table 2). If we look at the coalitions, so as to make it more

comparable to the American two-party system, we get a similar result: **0.27%** average differential from tallied results. Polls have been almost as accurate for the German vote in the European Parliament Elections (Table 3), averaging **0.44%** differential from tallied results over the past three elections.

**Table 2<sup>14</sup>: Exit poll vs Tallies in German National Elections**

Parties	2002 predicted	2002 tallied	2002 diff	1998 predicted	1998 tallied	1998 diff	1994 predicted	1994 tallied	1994 diff	average dif
<b>SPD</b>	38.0%	38.5%	0.5%	41.0%	40.9%	0.5%	36.5%	36.4%	0.1%	
<b>CDU/CSU</b>	38.0%	38.5%	0.5%	35.0%	35.2%	0.5%	42.0%	41.4%	0.6%	
<b>Green</b>	9.0%	8.6%	0.4%	6.5%	6.7%	0.4%	7.0%	7.3%	0.3%	
<b>FDP</b>	7.5%	7.4%	0.4%	6.5%	6.2%	0.4%	7.0%	6.9%	0.1%	
<b>PDS</b>	4.0%	4.0%	0.4%	5.0%	5.1%	0.4%	4.0%	4.4%	0.4%	
Rest	9.0%	8.6%		6.0%	5.9%		3.5%	3.6%		
<b>Average differential</b>			<b>0.30%</b>			<b>0.18%</b>			<b>0.30%</b>	<b>0.26%</b>
<b>Coalitions</b>										
<b>SPD/Green</b>	47.0%	47.1%	0.1%	47.5%	47.6%	47.5%	43.5%	43.7%	0.2%	
<b>CU/FDPI</b>	45.5%	45.9%	0.4%	41.5%	41.4%	0.1%	49.0%	48.3%	0.7%	
<b>Average differential</b>			<b>0.25%</b>			<b>0.10%</b>			<b>0.45%</b>	<b>0.27%</b>

**Table 3: Exit poll vs Tallies in European Parliament Elections (only German part)**

Parties	2004 predicted	2004 tallied	2004 diff	1999 predicted	1999 tallied	1999 diff	1994 predicted	1998 tallied	1998 diff	average dif
<b>SPD</b>	22.0%	21.5%	0.5%	31.0%	30.7%	0.3%	33.0%	32.2%	0.8%	
<b>CDU/CSU</b>	45.5%	44.5%	1.0%	48.0%	48.7%	0.7%	40.5%	38.8%	1.7%	
<b>Green</b>	11.5%	11.9%	0.4%	7.0%	6.4%	0.6%	10.0%	10.1%	0.1%	
<b>FDP</b>	6.0%	6.1%	0.1%	3.0%	3.0%	0.0%	4.0%	4.1%	0.1%	
<b>PDS</b>	6.0%	6.1%	0.1%	6.0%	5.8%	0.2%	4.5%	4.7%	0.2%	
<b>REP</b>							3.5%	3.9%	0.4%	
Rest	9.0%	9.8%		5.0%	5.4%		4.5%	6.2%		
<b>Average differential</b>			<b>0.42%</b>			<b>0.36%</b>			<b>0.55%</b>	<b>0.44%</b>

In the US, exit polls have also been quite precise. In particular, students at BYU have been conducting Utah exit polls since 1982.<sup>15</sup> They write:

... results are very precise; In the 2003 Salt Lake [City] mayoral race, the KBYU/Utah Colleges Exit Poll predicted 53.8 percent of the vote for Rocky Anderson and 46.2 percent for Frank Pignanelli. In the actual vote, Anderson carried 54 percent of the vote to Pignanelli's 46 percent.

14 Source: Election data: <http://www.bundeswahlleiter.de/> (English: [http://www.bundeswahlleiter.de/wahlen/e/index\\_e.htm](http://www.bundeswahlleiter.de/wahlen/e/index_e.htm))  
Prognoses: [www.forschungsgruppe.de](http://www.forschungsgruppe.de) Polling data also can be found at  
[http://de.wikipedia.org/wiki/Prognosen/Hochrechnungen\\_der\\_Bundestagswahl\\_2002](http://de.wikipedia.org/wiki/Prognosen/Hochrechnungen_der_Bundestagswahl_2002) (gives the predictions of another polling company for the 2002 election; it's predictions are in all cases within 1% of ZDF). I'd like to thank Dr. Andreas M. Wuest, Michael Morrissey, Kurt Gloos and Lars Vinx for their help in compiling this data.

15 <http://exitpoll.byu.edu/2004results.asp>. Aside from an LA Times poll, for which I could not determine whether or not the data was corrected, this was the only other exit poll conducted on the 2004 presidential election as far as I have been able to determine.

True to their word, predictions in this year's contests were quite accurate. In the Utah presidential vote, for example, they predicted Bush 70.8%, Kerry 26.5%. The actual was Bush 71.1%, Kerry 26.4%. Consistently accurate exit poll predictions from student volunteers, including in this presidential election, suggest we should expect accuracy, within statistical limits, from the world's most professional exit polling enterprise.

Exit polls have been widely used to verify elections. When Mexico sought legitimacy as a modernizing democracy in 1994, Carlos Salinas instituted reforms designed to ensure fair elections, and central among these were exit polls.<sup>16</sup> Exit pollsters were hired again for the next Presidential election in 2000,<sup>17</sup> and, not coincidentally, it was the first loss for the Institutional Revolutionary Party (PRI) in its 72 year history.

In Russia, and throughout the former Soviet block, exit polls have been used to verify elections. Last fall, international foundations sponsored an exit poll in the former Soviet Republic of Georgia during a parliamentary election. On Election Day, the pollsters projected a victory for the main opposition party. When the sitting government announced that its own slate of candidates had won, supporters of the opposition stormed the Parliament, and the president, Eduard A. Shevardnadze, resigned under pressure from the United States and Russia.<sup>18</sup>

## **Statistical Analysis of the Three Critical Battleground States: Ruling Out Random Error**

### **Three Critical Battleground States**

The conventional wisdom going into the election was that three critical states would likely determine who would win the Presidential election -- Ohio, Pennsylvania, and Florida. Typical analyst comments:

Since Election 2000, Republicans and Democrats have banked their aspirations on an electoral trinity: Florida, Pennsylvania and Ohio. As the Big Three goes, so goes the nation.

---

<sup>16</sup> Paul B. Carroll, Dianne Solis. "Zedillo's apparently clean win at polls diminishes threat of Mexican unrest." *The Wall Street Journal* August 23, 1994 pA2

<sup>17</sup> Rebeca Rodriguez U.S. political consultants signed to conduct exit poll in Mexico. *Knight Ridder Newspapers*, June 16, 2000

<sup>18</sup> Martin Plissner, Exit Polls to Protect the Vote, *New York Times* 10/17/04



- David Paul Kuhn, CBS News: High-Stakes Battle For The Big 3” Oct. 26, 2004

Conventional wisdom for months, including RCP's, had been that whoever won two of the "big three" Ohio, Pennsylvania and Florida would almost certainly become President.

- Real Clear Politics: [posting 10/28/04](#)

The accepted wisdom is that whoever wins two out of the three states of Ohio, Pennsylvania and Florida will win the election.

- Rob Watson, BBC News: October 28, 2004

The numbers and logic were straightforward. Of the other battleground states: Michigan and Minnesota always leaned Democratic, Colorado and Nevada Republican. Iowa, New Hampshire and New Mexico don't have many electoral votes. Wisconsin has a long tradition as a liberal state, and only 10 electoral votes compared to 20, 21, and 29 for the big three.

The campaigns themselves apparently agreed. Ohio, Pennsylvania, and Florida were the three states the candidates visited most, and in which they spent the most money.<sup>19</sup>

The conventional wisdom proved correct. Bush won two of the three and ascended to electoral victory by virtue of that. In each of these states, however, exit polls differed considerably from recorded tallies (Table 4).

**Table 4: Predicted vs. tallied percentages in the three critical battleground states**

	Sample size	Bush predicted	Kerry predicted	Predicted differential	Bush tallied	Kerry tallied	Tallied differential	Tallied vs. predicted
Florida	2846	49.8%	49.7%	Bush 0.1 <sup>20</sup>	52.1%	47.1%	Bush 5.0	Bush 4.9
Ohio	1963	47.9%	52.1%	Kerry 4.2	51.0%	48.5%	Bush 2.5	Bush 6.7
Pennsylvania	1930	45.4%	54.1%	Kerry 8.7	48.6%	50.8%	Kerry 2.2	Bush 6.5

## A Statistical Anomaly?

A basic question to ask on looking at such a discrepancy is whether it is just a statistical anomaly. It can happen, for example, that a fair coin tossed ten times will land heads each time, but it doesn't happen often (1 out of 256 times). If we witness this, we will at least suspect that the coin might be adulterated, especially if the stakes are high and we are not permitted to inspect the coin carefully.

<sup>19</sup> See stories cited above.

<sup>20</sup> Earlier exit polls, including one released by Slate at 7:28 EST, 28 minutes after the Florida polls closed showed Kerry leading 50% to 49% (Appendix C)

Statistical significance, which means that the discrepancy is such that it is unlikely to occur by chance, depends on four factors – the size of the discrepancy, the sample size, sample characteristics, and the level of significance (just how unlikely does it have to be?).<sup>21</sup> Table 4 provides sample size and discrepancy. For statistical purposes, these samples are quite large. Two thousand or so respondents is roughly the size of most national polls.

Without access to the data and methodology, we cannot model the sample characteristics precisely. But we do know the general procedures by which exit polls are conducted. Appendix A provides the 2000 Presidential election exit poll methodology and a bibliography of articles on the process from that and other elections. Based on these we can make a reasonable approximation.

A random sample of a population can be modeled as a normal distribution curve. Exit polls, however, are not random samples. To avoid prohibitive expense, exit poll samples are clustered, which means that precincts, rather than individuals, are randomly selected. This increases variance and thus the margin of error because of the possibility that precinct voters share similar characteristics that differentiate them from the rest of the state in ways that past voting behavior would not predict.

Pollsters also use a counterbalancing process that decreases variance – stratification. Identifying voters by key characteristics that predict voting behavior (race, sex, age, income, ethnicity, religion, party affiliation, etc...) ensures that the sample is representative of the overall population, either by seeking out subjects with specific demographic characteristics and/or weighting groups depending on their representation in the sample compared with that of the overall voting population. By getting samples in which minorities are over-represented (but subsequently negatively weighted), pollsters can ensure adequate sample sizes of each of these representative subgroups. Knowing exactly how much to weight over- or under-represented population depends on an accurate knowledge of overall demographics of the electorate.

---

<sup>21</sup> (Apologies to those who are well versed in statistical inference. Most readers of this paper are not, so I provide much more explanation than I would for a purely academic reader.)

Historical data, census data, and registration roles, can be used to compliment sampling site counts to try to accurately weight the sample.

An early draft based of this paper, based on an assumption that the effects stratification of could balance the effects of clustering generated a headline grabbing probability of 250,000,000-to-one odds that exit poll deviations from counts could be due to chance or random error. In this analysis, I use more conservative estimates. An analysis of the 1996 exit polls estimated that the cluster sample design adds "a 30 percent increase in the sampling error computed under the assumption of simple random sampling" (Merkle and Edelman, 2000, p. 72). That study is particularly apt because the 1996 state exit polls involved roughly the same number of precincts (1,468) as this year's polls (1,480).<sup>22</sup> In the analysis below, I also conservatively assume no counterbalancing effects due to stratification. Although in principal, pollsters weight over- and under-sampled groups, thereby ensuring a more representative sample than chance alone would dictate, there is no magic formula for exactly what weight to assign a group. The only measure of the demographics of actual voters on Election Day is the exit poll itself.

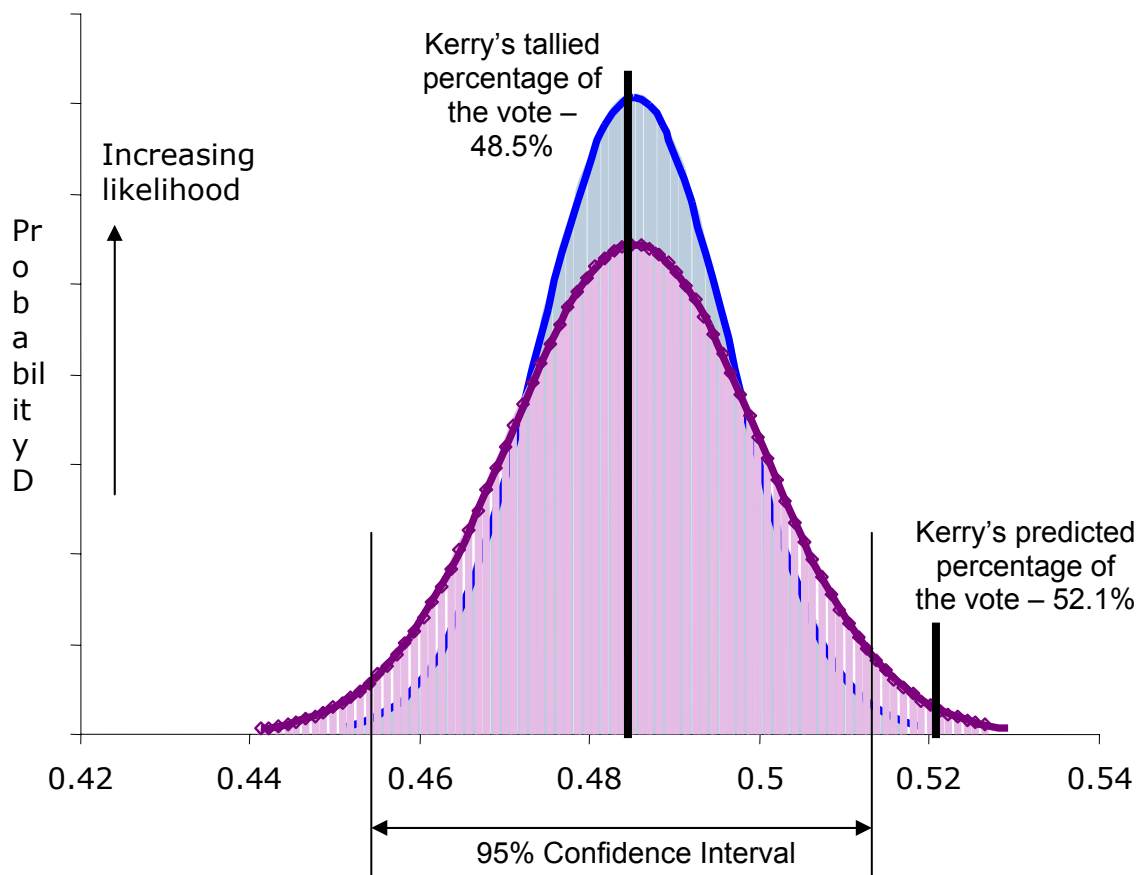
Figure 2 depicts the resulting distribution curve for samples of 1,936 randomly selected respondents from approximately 40 randomly selected precincts in a state in which 48.5% of the vote went for Kerry. The thin blue density curve is that of a simple random sample; the wider purple curve is of a clustered sample with no stratification. The horizontal double arrow below the curve indicates the poll's statistical margin of error, the corresponding 95% confidence interval.<sup>23</sup> If one hundred unbiased samples were drawn from this population, we would expect that in 95 (on average), Kerry would poll between 45.6% and 51.4%. And because half of the 1-in-20 cases that fall outside the interval would be low rather than high, 97.5% of the time we

<sup>22</sup> <http://www.exit-poll.net/faq.html#a7>

<sup>23</sup> To determine the margin of error, calculate the standard error of a random sample using the formula  $\sqrt{\frac{p(1-p)}{N}}$  where p = Kerry percentage of the vote and N is the sample size. (.0113). To adjust for the fact that this is a clustered sample, add 30% (.01466 or 1.47%). Sixty-eight percent of the time, a prediction from a sample this size would be within one standard error, then . Ninety-five percent of the time, it will be within 1.96 standard errors (2.87% in this case).

would expect Kerry to poll no more than 51.4%. It turns out that the likelihood that Kerry would poll 52.1% from a population in which he receives only 48.5% of the vote is less than one-in-one-hundred (.0073).

**Figure 2. Statistical prediction of Kerry's true percentage of the vote in Ohio**



Conducting the same analysis for Florida, we find that Kerry's poll prediction of 49.7% of the vote is likewise outside the 95% confidence interval. Given a population in which he receives only 47.1% of the vote, the chances that he would poll 49.7% out of 2846 respondent in an exit poll with no systematic error is less than two-in-one-hundred (.0164). Kerry's poll numbers are outside the 95% confidence interval as well in the third critical battleground state, Pennsylvania. Although he did carry the state, the likelihood that an exit poll would predict 54.1%, given 50.8% support of the electorate is just slightly more than one-in-one-hundred (.0126).

Assuming independent state polls with no systematic bias, the odds against any two of these statistical anomalies occurring together is between 5,000:1 and 10,000:1. (20-40 times more improbable than ten straight heads from a fair coin) The odds against all three occurring together are 662,000-to-one. As much as we can say in social science that something is impossible, it is impossible that the discrepancies between predicted and actual vote counts in the three critical battleground states of the 2004 election could have been due to chance or random error.