

US Count Votes
National Election Data Archive

**The 2004 Presidential Election:
Exit Poll Error or Vote Miscount?**

September 8, 2005

Ron Baiman, Ph.D – Institute of Government and Public Affairs, University of Illinois at Chicago
David Dodge - BSMME Engineer, Composite Developments, Inc., Technology and Innovation
Kathy Dopp -MS in mathematics - USCountVotes, President

Reviewed via USCountVotes' email discussion list for statisticians, mathematicians and pollsters.

Press Contact:

Kathy Dopp, National Election Data Archive, President kathy@uscountvotes.org

This paper can be found on the Internet at:

http://uscountvotes.org/ucvAnalysis/US/exit-polls/USCV_exit_poll_analysis.pdf

Introduction

The exit polling firm Edison/Mitofsky¹ (E/M), and USCV, agree that the historically unprecedented discrepancy between the exit polls and the reported vote count for the 2004 U.S. Presidential election cannot be a result of random sampling error.² This leaves either exit poll error or vote miscount as the only two possible explanations for the exit poll discrepancy.

E/M has claimed that the exit poll discrepancy is exclusively a result of “within precinct error” (WPE)³, and that the entire WPE observed in 2004 could be explained by a hypothetical exit poll completion rate of 56% among Kerry voters and 50% among Bush voters (herein referred to as “the E/M hypothetical”).⁴ The E/M hypothetical was widely interpreted by the media and by USCV as a claim that the 2004 exit poll discrepancy was caused by a pervasive, and on average uniform, shortfall in Bush voter exit poll response relative to Kerry voter exit poll response that was dubbed the “reluctant Bush response” (rBr) hypothesis.⁵

A recent clarification by E/M indicates that the “E/M hypothetical” should be interpreted as referring to hypothetical *average* (rather than *constant* average) partisan exit poll response rates.⁶ In this interpretation, average precinct partisan response rates may *vary widely* by reported precinct vote shares, yet all of the reported WPE could be explained by partisan response rates whose average across the sample is $K=0.56$ and $B=.50$.⁷ This interpretation of the “E/M hypothetical” does *not* depend on the “rBr hypothesis” of *constant* average partisan response rates which was shown by USCV to be inconsistent with the pattern of the exit poll discrepancy.⁸

However, our analysis below shows that even if the “E/M hypothetical,” is interpreted as referring to average, rather than constant average, partisan exit poll response rates, it is *inconsistent* with the reported WPE data. There is no configuration of partisan response rates, however varying across

¹ "Evaluation of the Edison/Mitofsky Election System 2004" January 19, 2005

<http://www.exit-poll.net/election-night/EvaluationJan192005.pdf>

² See “Analysis of the 2004 Presidential Election Exit Poll Discrepancies,” USCV, updated April 12, 2005. E/M state that the average “within precinct error” (WPE) of the exit polls in 2005 of -6.5% was the largest since 1988 and was 30% higher than the next largest mean WPE of -5.0% in 1992, op. cit. January 19 E/M report, p. 34.

³ WPE is defined by E/M is "an average of the difference between the percentage margin between the leading candidates in the exit poll and the actual vote for all sample precincts in a state."

⁴ Op. cit., January 19 E/M report, p. 28 and 31.

⁵ It should be noted that “rBr” does not necessarily imply that Bush voters were “psychologically” more adverse than Kerry voters to completing exit polls. The partisan exit poll response gap could for example be linked to the characteristics and methods of the exit *pollsters*, or it could be a function of the external circumstances in which exit polling was conducted. However, the initial interpretation of “rBr” as stipulating a pervasive and, *on average, uniform*, bias in exit poll response, seemed to indicate that Bush *voters* as a group (regardless of the characteristics of the exit pollsters and any other possible factors) had a lower completion rate than Kerry *voters*. This seemed to suggest that the explanation had to do with, on average, differential behavior of Bush and Kerry partisans that was unrelated to any other factor, i.e. a “psychological” or “group behavioral” explanation.

⁶ In a (5/26/2005) communication to Ron Baiman, Warren Mitofsky states that: “**There is no constant mean bias conjecture on our part. This is wholly USCV's invention.**” Mitofsky's statement would appear to vindicate USCV's position that the “constant mean bias” or rBr hypothesis cannot explain the exit poll discrepancy.

⁷ Average K and B levels will be equal to *weighted average* K and B for partisan precinct categories, where the weights are the relative sample sizes of the precinct categories - see Table 6 in Appendix F.

⁸ See USCV April 12 report, op. cit. Elizabeth Liddle recently published (<http://www.geocities.com/lizzielid/WPEpaper.pdf>) a simulation-based analysis suggesting that a new “unconfounded” index shows that a constant mean response bias *is* consistent with Edison/Mitofsky's reported exit poll discrepancies. However, as Liddle's analysis is based on the *same variables* investigated earlier by USCV (see Appendix A, Liddle's index equals $\text{LN}(K/B)$), its conclusions cannot logically be upheld, if USCV's analysis is correct. The recent statement by Mitofsky (op. cit.) appears to support USCV's analysis.

precinct partisanship categories, with overall averages of $K=0.56$ and $B=0.5$ that can produce the reported values of the actual E/M exit poll data for mean WPE and median WPE.⁹ Thus, neither a "reluctant Bush responder" (rBr) hypothesis interpretation nor an overall average interpretation of the "E/M hypothetical" is consistent with the WPE pattern shown by the Edison/Mitofsky exit polling data.

A *larger* overall average partisan exit poll response gap than that proposed by E/M *could* account for the observed WPE, but the large differences in mean and median partisan response bias¹⁰ necessary to produce the actual observed WPE levels across partisan precinct categories would need to be explained. In particular the *very large* mean and median WPE for precincts with over 80% reported Bush vote, and the *almost zero* mean and median WPE for precincts with over 80% reported Kerry vote, requires an explanation.

Ten months after the election, no plausible explanation of the 2004 exit poll discrepancy, based on exit polling error, has been provided by E/M. The precinct level exit polling and official vote count data that would enable independent investigators to analyze the exit poll discrepancy has not been provided to the public. Perhaps an exit poll explanation for the discrepancy does exist. However, a cloud of suspicion is cast on the 2004 presidential election results because the possibility that a "vote miscounts" explanation is required to generate the reported exit poll discrepancies is still open. It is a matter of the utmost national importance that detailed precinct level exit polling and election data that would allow for investigation by independent analysts, such as USCV, be publicly released.¹¹

Analysis of the Aggregate Exit Poll Data¹⁴

The Edison/Mitofsky report of January 19th, only provides the following summary statistics:

Table 1: Partisanship Precinct Data given in the Edison/Mitofsky Report (pp. 36, 37)							
Partisanship of Precinct by Election Results		Number of Precincts	mean WPE exit poll discrepancy	median WPE exit poll discrepancy	Combined Response Rate	Refusal Rate	Miss Rate
80< Kerry <=100%	0< Bush <=20%	90	0.3%	-0.4%	53%	35%	12%
60< Kerry <=80%	20< Bush <=40%	165	-5.9%	-5.5%	55%	33%	12%
40< Kerry <=60%	40< Bush <=60%	540	-8.5%	-8.3%	52%	37%	11%
20< Kerry <=40%	60< Bush <=80%	415	-6.1%	-6.1%	55%	35%	10%
0< Kerry <=20%	80< Bush <=100%	40	-10.0%	-5.8%	56%	33%	11%

⁹ See Appendix A for derivations of partisan exit poll response rates K and B. See Appendices B and Table 6 in Appendix F for a proof of this statement. "Precinct partisanship categories" refers to precincts grouped together by the percentage of official votes that Kerry and Bush received in them.

¹⁰ "Bias" equals K/B and will sometimes also be labeled "Alpha" in this paper, following Liddle's notation, op. cit.

¹¹ The claim that E/M cannot release this data because of concerns over "respondent confidentiality" is belied by the fact that a "blurred" version of this data for the state of Ohio has been released to the Election Science Institute. See discussion below of apparent misstatements and inconsistencies in the Kyle, Samuleson, Scheuren, and Vicinanza report on the Ohio exit polls based on this data (see: http://www.votewatch.us/reports/view_reports). This report appears to support a vote miscount hypothesis rather than refute it as is claimed by the authors. In any case it is of limited value as it does not include information on the factors that influence WPE necessary for a substantive statistical analysis of the exit poll discrepancies. It should be noted that since this additional data on (anonymous) exit pollsters and polling conditions has no relation to exit poll *respondents*, its release would not compromise confidentiality in any way.

¹⁴ See Appendix F

At the April 2005 AAPOR conference, Warren Mitofsky released a scatter-plot of precincts' WPE which show many outlier precincts where large measurement errors or vote miscounts may have occurred favoring Kerry as well as Bush (reproduced below). However, based on the reported correlation coefficient, E/M's scatter-plot indicates that on average exit poll discrepancy favored Kerry (see coefficient reproduced below).

Individual precinct discrepancies may be the result of measurement error linked to the numerous exit polling factors affecting WPE that were listed in the E/M report, and to other possible factors such as partisanship of the pollster that are not listed in the report.¹⁵ If these factors are pervasive, and more or less randomly distributed across the precincts from which exit polls are taken, one would expect their impact to be either a) on average "neutral", so that the resulting mean bias would be zero; or b) on average, "uniform", so that the exit poll response bias would be pervasive, leading to the mean uniform rBr hypothesis. The net *mean and median* bias in exit polling discrepancy in favor of Kerry needs to be explained statistically.

One question that needs to be answered is "Why do the factors that produce pro-Kerry discrepancies *not* appear to generate a "neutral" or "mean uniform" exit polling bias across precinct partisanship?" Are the factors that produce pro-Kerry discrepancies sufficiently over-represented in precincts with reported Bush election shares of 80% or more to produce the observed very large (absolute) mean and median WPE in these in "high Bush precincts"? Conversely, are these factors *balanced* against factors that produce pro-Bush discrepancies in "high Kerry precincts" so that observed mean and median WPE is almost zero in precincts with reported Kerry election shares of 80% or more?

The standard statistical technique used to answer these kinds of questions is multivariate analysis. Ten months after the election this kind of substantive analysis that would at a high probability, uphold, or reject, an "exit poll error" explanation for the 2004 exit polling discrepancies has still not been provided.¹⁶

Moreover, such a multiple regression analysis would have to take into account the fact that the exit poll data are based on (attempted) random samples of exit polled precincts. Errors in partisan response rates thus need to be related to *modeling* variance similar to that of the exit poll data, and not to the *data* variance that is presumably affected by numerous non-sampling error factors. It is precisely the bias, if there is any, of these *non-sampling errors* that needs to be explained.

Appendix G provides a methodology for estimating the model variances for exit poll partisan response rates. Factors that affect partisan response should explain, up to a reasonable level of "model variance", differences in mean and median exit poll response bias, for statistically significant samples of 30 or more precincts. A regression analysis that attributes these differences to "data variance" cannot be considered to be an adequate explanation of the exit poll discrepancies.¹⁷

¹⁵ January 19, 2005, E/M report, p. 35-46, op. cit.

¹⁶ When Ron Baiman asked Warren Mitofsky about this at the May 2005 AAPOR (American Association for Public Opinion Research) convention, Mitofsky stated that this kind of analysis had been done. He did not answer the questions of why this E/M multivariate analysis had not been publicly released, and why the data that would allow independent regression analysis that could replicate and confirm the E/M analysis, had not been provided.

¹⁷ This is an important point of disagreement with Elizabeth Liddle who has claimed, in private communication with Ron Baiman, that an exit poll "explanation" up the level of "data variance" is adequate. In our view this claim is analogous to a claim that the state level exit poll discrepancies are also "insignificant" as they can be accounted for by the very large *data* variance of the exit polls. But, as has been noted, both E/M and USCV agree that these average exit poll discrepancies are highly significant because they fall outside of reasonable confidence intervals determined by *model*

Without precinct level data, the only data available is the aggregate tabulation data provided in the E/M report.¹⁸ Aggregate means and medians are generally good descriptors of reasonably well behaved (approximately normally distributed) aggregate data like exit poll data. Simple parameters calculated from means and medians of aggregate data can be regarded as fairly good indicators of the mean and median values of the same parameters calculated from individual data points.

An analysis of the aggregated Edison/Mitofsky data (see Appendix F) suggests that it is highly improbable that the exit poll outcomes described in the E/M report can be a result of either “mean uniform”, or “overall average”, $K=0.56$ and $B=0.50$ partisan exit poll response rates.

In Appendix G upper bounds on the standard deviations of K and B (estimated in Table 1 of Appendix F) are calculated. Given that these standard deviation upper bounds are based on an assumption that every precinct has only 20 respondents, they presumably are *considerably larger* than the *actual* random sampling standard deviations of K and B.¹⁹ Thus one would not necessarily expect these upper bounds on standard deviations to pick up any significant differences.

However, Tables 1-4 of Appendix G show that based on these “maximal standard deviations”, K values differ significantly from each other across partisan precinct categories. Moreover, B values also differ but much less so (Tables 1-4). This suggests that the *actual* sampling error standard deviations will show *highly significant* differences between K values, and some B values, across partisan precinct categories.²⁰ These results also suggest that if the exit polling discrepancies are due to partisan response “bias”(=K/B), more of this bias must come from, large, statistically significant, changes in K rather than from changes in B. Exit polling “bias factors” must have a greater effect on K than on B.

Since partisan exit poll response rates K and B are random samples, their mean and median values should not be significantly different from R, the overall precinct exit poll response rate. Table 5 of Appendix G suggests that the only partisan precinct category for which mean and median K and B are insignificantly different from R is the high Kerry ($k \geq 0.80$) precinct category. The high Kerry precinct category is thus the only precinct category that is consistent with pure random sampling error. This suggests that for these precincts response rate exit polling “bias factors”, and/or vote miscount errors, net out to zero, so that the net mean and median discrepancies are within the limits of pure random sampling error.

(and not data) variance. Both the exit poll data and the response rate data are the result of (attempted) random selection and not of arbitrary empirical data measurement.

¹⁸ Though 2004 precinct level exit polling data has been made publicly available, the weights for these exit poll data have been adjusted to match the *state level* reported election outcomes (see: <http://webapp.icpsr.umich.edu/cocoon/ICPSR-STUDY/04181.xml>). Moreover, corresponding *precinct level* reported election outcomes and the data on the exit polling factors that influence WPE have not been provided. In short these data are almost useless for investigating the exit poll discrepancy.

¹⁹ In calculating the upper bounds we know that most precincts have more than 20 respondents but not *how many* more.

²⁰ These are significant differences based on “model” *random sampling* variance. It is clear from the Mitofsky scatter plot (see below) that the *data* variance is much larger. However, claims of no significant differences because of large *data* variance (for example by running an ANOVA test) fail to take into account the fact that these data are (attempted) *random samples* and *not* arbitrary measurements. This suggests that most of the *data variance* in K and B is not *random* sampling variance, but rather *results from factors that bias exit polling response* toward one candidate or another. These are thus *significant differences* that cannot be attributed to random sampling error. As has been noted in footnote 17, this needs to be *explained* by the *bias factors*, if an exit polling explanation exists for them.

The analysis of the aggregated Edison/Mitofsky data performed in Appendix G suggests that:

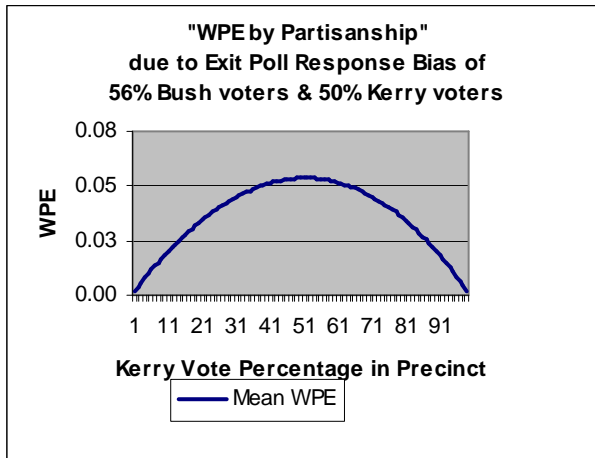
- a) Partisan response rates K and B vary significantly across partisan precinct categories in ways that cannot be explained by pure (or impure) random sampling error.
- b) Variations across partisan precinct categories in the response rate of Kerry voters (K) are larger than those of Bush voters (B).
- c) High Kerry precincts ($k > 0.80$) are the only partisan precinct category for which exit polling response rates K and B are insignificantly different from what they would be for pure random sampling ($K=B=R$).

The Signature of Exit Poll Response Bias

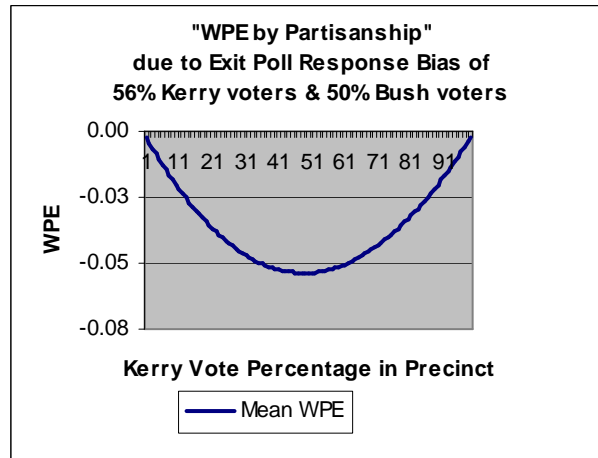
What happens to exit poll within precinct error (WPE) patterns when there is a voter response bias?

Algebraic properties of within precinct error (WPE)²¹ produced by exit poll response bias cause maximum WPE amounts where Bush and Kerry reported votes are closest to 50/50. WPE goes to zero in both Bush and Kerry vote strongholds. The curve in Graph 1 below shows the WPE pattern that would result if Kerry and Bush voters had mean exit poll response rates of 50% and 56% respectively, and the curve in Graph 2 shows the expected curve if Kerry and Bush voters' mean exit poll response rates were 56% and 50% respectively. Positive WPEs (in Graph 1 below) result if Bush voter response rates are higher, and negative WPEs result if Kerry voters' response rates are higher (in Graph 2 below).²²

Graph 1



Graph 2



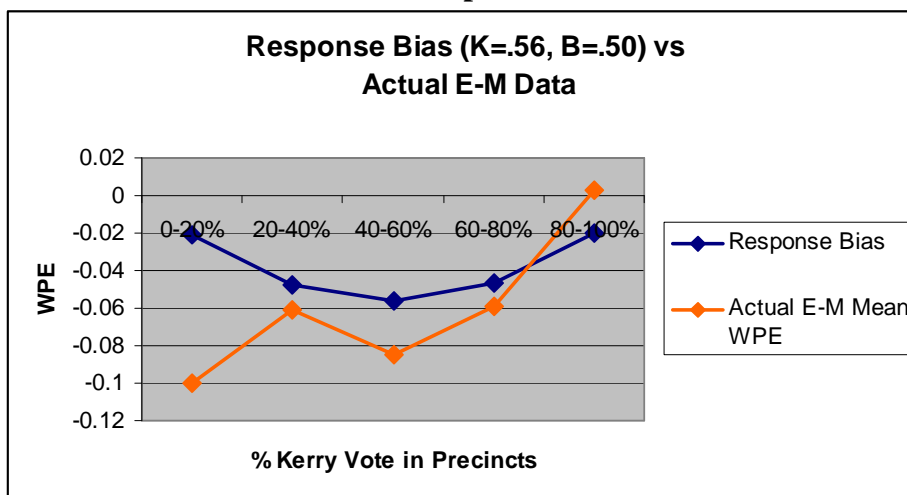
Graph 3 below compares (orange line) the actual data reported by Edison/Mitofsky versus (blue line) the simulated WPE patterns that would result from exit poll response bias of Bush and Kerry voters with means 50% and 56% that was hypothesized by Edison/Mitofsky and nicknamed the reluctant Bush responder (rBr) theory.

Notice how the patterns differ on the leftmost side in high Bush vote precincts where WPE is much higher than expected.

²¹ See Appendix B. WPE is a measure of the discrepancies between exit poll and election results within a given precinct.

²² Fixing the ratio, or K/B “bias” where $K > B$, in the case of $K=56\%$ and $B=50\%$, generates an asymmetry that slightly increases the (absolute) WPE in high Bush precincts relative to high Kerry precincts. However, this asymmetry, that was first pointed out by Liddle (see footnote 6, op. cit.), results in a *very small* increase in (absolute) WPE that does *not come close* to off-setting the decline in (absolute) WPE due to the over-all “inverted U” shape that results from constant partisan response bias - see Table 5 Appendix F.

Graph 3



What Happened in Precincts where Bush Vote was Over 80%?

In USCV's March 31st paper we noted that the E/M reported mean and median WPE for precincts where the Bush vote was greater than 80% was -10% and -5.8% respectively. This implies that half of these 40 high Bush vote precincts had very large WPEs of at least -14.2%.²⁶

The total number of these high-Bush vote precincts sampled is small – only 40 out of 1250 precincts in Edison/Mitofsky's sample. If votes were shifted from Kerry to Bush in the 415 precincts in the 60% to 80% Bush vote precincts, then some of them, with high mean WPEs would "shift" to the 80% to 100% group. Hence, due to its small number of precincts, mean WPE in the 80% to 100% Bush vote precincts is sensitive to vote miscounts which benefit Bush because greater increases in WPE would occur when precincts with high WPE are shifted into this category. Miscounted votes favoring Bush would increase WPE, and then shift precincts with high WPEs into the high Bush vote category from the 60 to 80% Bush vote categories.

This vote miscount hypothetical could explain some share of the large absolute WPE in high Bush precincts. On the other hand, it is not possible to provide a statistically plausible exit polling measurement error explanation for the large absolute *mean and median* WPE in these precincts, or the almost zero WPE (significantly smaller than would occur with a constant 0.56/0.50 bias even with the “inverted U” pattern – see Appendix F) in high Kerry precincts, without detailed precinct level data.²⁷

²⁶ USCV, April 2005, op. cit., p. 14. http://electionarchive.org/ucvAnalysis/US/Exit_Polls_2004_Edison-Mitofsky.pdf

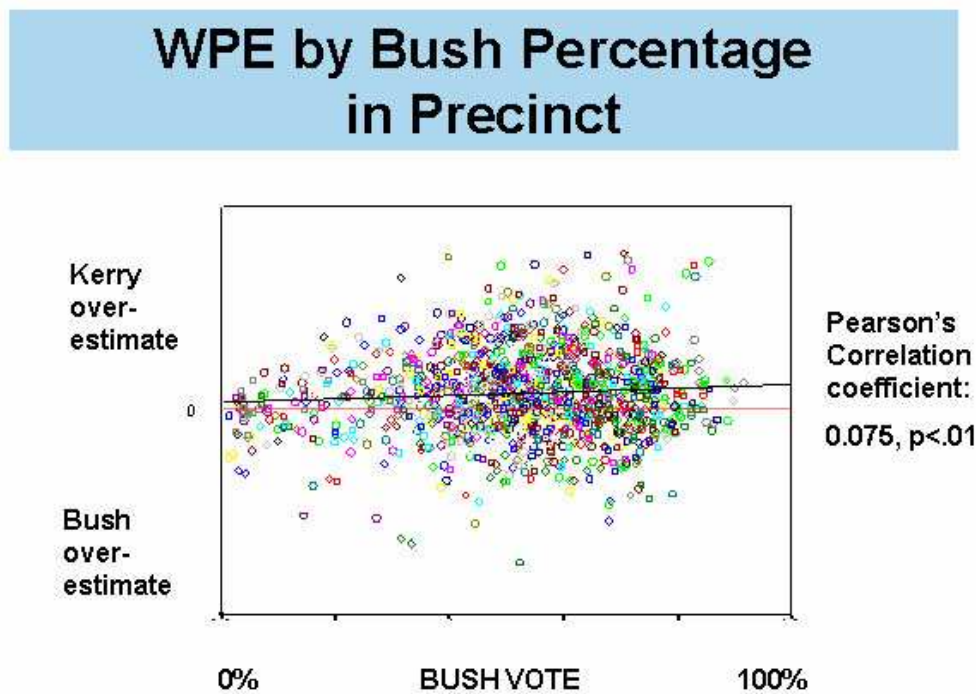
²⁷ Some have argued, from looking at the scatter plot below, that the large WPE in high Bush precincts may be a result of only four outliers. Aside from the fact that “four outliers” constitute a full 10% of the sample of only 40 high Bush precincts and thus presumably represent the outcomes of about 10% of the much large number of these precincts in the population, calculations from *medians* which are presumably not highly affected by outliers produce qualitatively similar results (see Table 1 Appendix F). Moreover the outlier explanation does not explain the zero mean and median WPE's in high Kerry precincts. In short removing outliers will not resolve the puzzle.

The patterns of high discrepancies in 80%+ Bush vote precincts and the almost zero discrepancies in 80%+ Kerry precincts, and large mean and median WPE levels of over 5.5% in the more competitive precincts, has yet to be explained.

Outlier Precincts in Mitofsky's Scatter-Plot

Graph 4 below reproduces a scatter plot of WPE by the percentage of Bush vote in precincts was released by Mitofsky at the May, 2005 AAPOR conference.²⁸ The scatter plot reveals many outlier precincts with extremely high WPE indicating exit poll error favoring Kerry or vote miscounts favoring Bush as well as exit poll error favoring Bush or vote miscounts favoring Kerry.²⁹ However, as noted above net, or average, exit poll error favored Kerry.³⁰

Graph 4



²⁸ Another chart release by Mitofsky at the AAPOR conference showed a zero correlation between reported precinct partisanship and Little's WPE_index (which is the same as $\text{LN}(K/B)$ – see Appendix A). This was widely interpreted at the time as “proving” the E-M hypothetical of a constant mean bias. Aside from the fact, noted above, that E-M has effectively retracted this “rBr hypothesis”, it should be clear that a flat line can be drawn through any number of non-linear “inverted U” or other patterns. Since the number of partisan precincts in the sample is so small, a slight asymmetry in the “inverted U” could still easily generate a “flat” line. In short, the zero linear correlation is a mathematical finding with little operational relevance to the issue of the statistical significance of “bias” variation across partisan precinct categories that include statistically significant (30 or more) precincts.

²⁹ Mitofsky released this scatter plot at the May AAPOR conference.

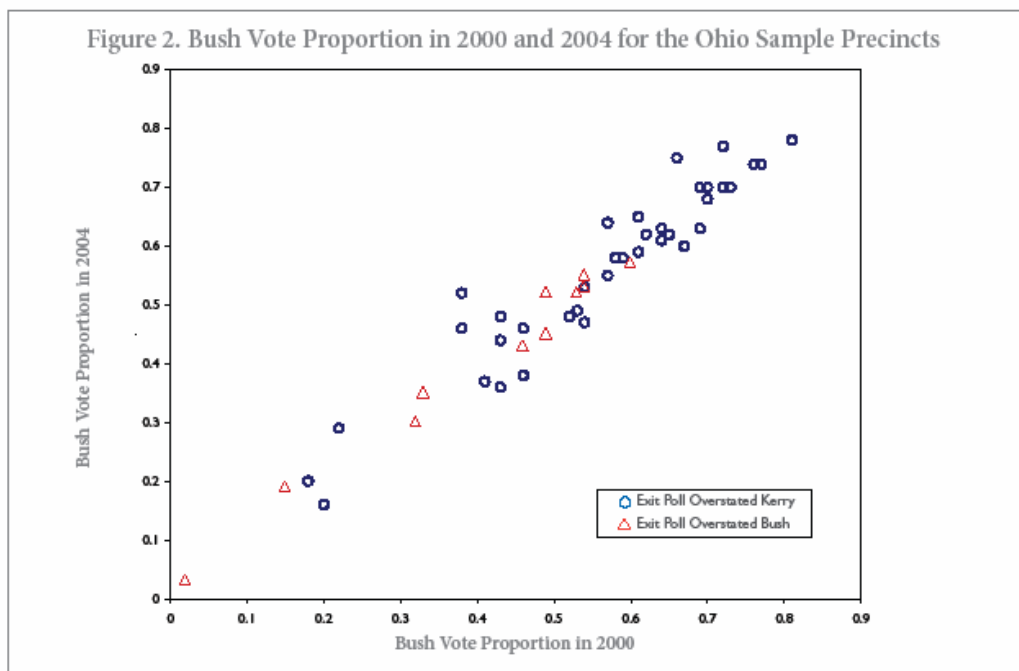
³⁰ Individual precinct discrepancies could, and should, if they are very large and statistically unexplainable, be the subject of follow-up “on the ground” investigations.

Ohio Exit Polls Show Similar Unexplained Patterns of Discrepancies

A brief on an Election Sciences Institute paper by Kyle et. al. claims that ..”the data do not support accusations of election fraud in the Ohio Presidential election of 2004.”³¹ However, the data used to support this claim show unexplained exit poll discrepancy patterns that mirror those reported by E/M from the national exit poll data.

Graph 5 below reproduces Figure 2 of the Kyle et. al. data.

Graph 5



The chart above plots the proportion of the vote Bush received in each of the sample precincts in 2004 (vertical axis) and the proportion he received in the same precincts in 2000 (horizontal axis). The different symbols indicate whether the 2004 exit poll result for the precinct overestimated the proportion for Bush (△) or for Kerry (○). What is evident

from the plot is that all the precincts voted very similarly in 2000 and 2004, and the direction of error in the exit poll estimates does not appear to be strongly related to how large a proportion of the vote Bush received in the precinct.

ESI For further information contact info@electionscience.org.

The text below the Figure 2 states that the:

“...the direction of error in the exit poll estimates does not appear to be strongly related to how large a proportion of the vote Bush received in the precinct.”

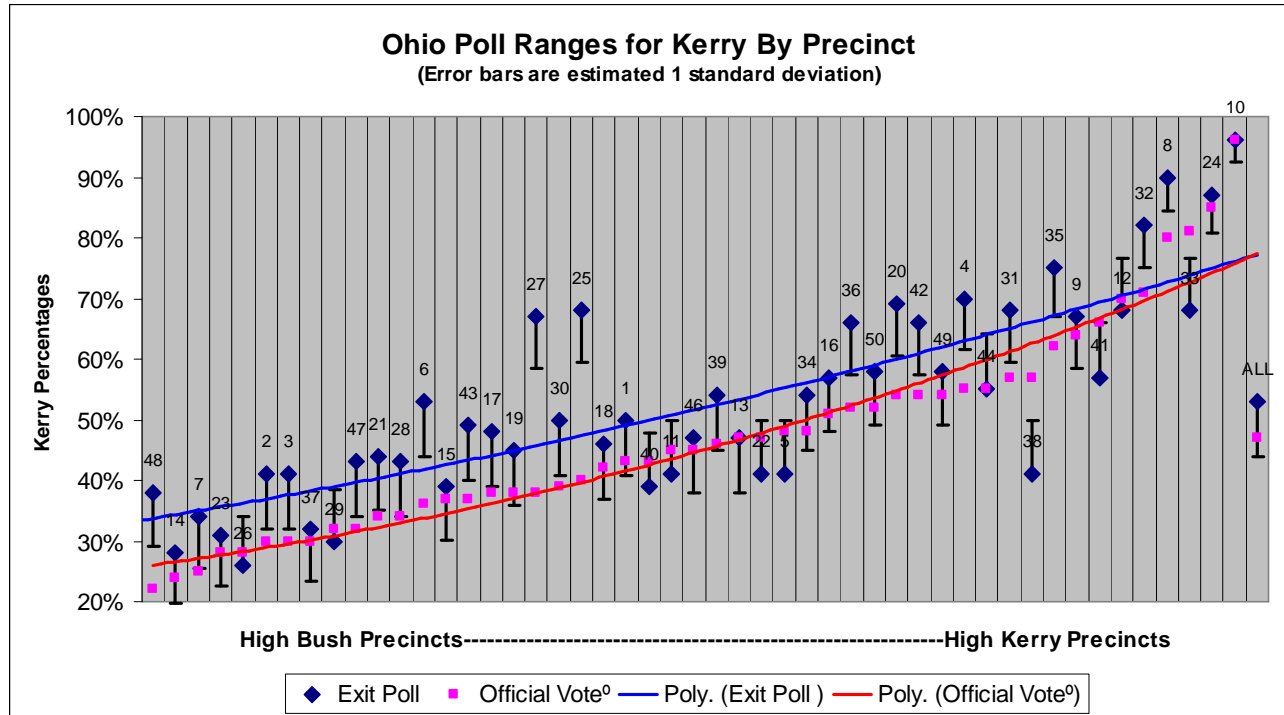
However this is clearly not the case as for precincts with approximately 60% or greater Bush vote in 2004 and 2000, the Figure shows that 2004 exit polls are *exclusively* pro-Kerry discrepancies (only

³¹ See http://www.votewatch.us/reports/view_reports , “ESI Brief - Analysis of the 2004 Ohio Exit Polls and Election Results”, p. 1.

circles appear in the upper right hand corner of the Figure).³² This mirrors the unexplained high pro-Kerry discrepancies for high Bush precincts in the national exit poll data (see Table A above).³³

Finally, we have used the data from Table 1 of the Kyle et. al. report to create Graph 7 below:

Graph 7



Estimates of standard deviation in Graph 7 assume precinct sample sizes of 30 exit poll respondents.³⁴ The vertical bars in the graph display one-sided confidence intervals of one standard deviation (one-tail probabilities of about 84%) based on this conservative standard deviation (in most cases) overestimate - see Appendix H for detailed derivations and methodology.

This graph shows a strong relationship between precinct partisanship and exit poll discrepancy similar to that shown in this report for the national exit poll discrepancies. As the Kerry reported vote share increases, exit poll discrepancy trends to zero. It also shows that confidence intervals for the Kerry exit poll estimates fall *above* of the official precinct level reported Kerry vote shares for 17 of the 49 precincts, indicating Kerry exit poll discrepancies that significantly overestimate

³² In another apparent data inconsistency (see footnote 35 below) in the Kyle et. al. ESI report, the data shown in Table 1 of the report shows 2 pro-Bush discrepancies among high Bush precincts (of about 68% and 72%) out of a total of 11 pro-Bush discrepancies. However, even if Table 1, and not Figure 2, is accurate, this represents a highly disproportionate number pro-Kerry discrepancies among high-Bush precincts.

³³ The other point that report presumably is attempting to show with this graph: that the correlation between Bush vote shares in 2004 and 2000 does not suggest irregularities in 2004, is, of course, conditional on their having been a free and fair election in Ohio in 2000. However, if there is a correlation between pro-Kerry exit poll discrepancy and 2004 *and* 2000 Bush precinct partisanship, this may not be a good assumption to make.

³⁴ The average exit poll response for Ohio was 2042/49=41.7 – see www.exit-poll.net/election-night/MethodsStatementStateGeneric.pdf.

reported official Kerry vote shares in over 34% of the exit polled precincts.³⁵ The confidence interval for the Kerry exit poll vote share falls *below* the reported vote share in only two cases (for precincts 33 and 38 on the far right of the graph), or 4% of the 49 precincts.³⁶

Estimates of the probabilities that exit poll non-responders split in a way that will generate the Kerry reported election results in Ohio given the exit poll results (see “One Tail P-Value of X with 30 Non-Respondents” column in Table 1 in Appendix H) show that a full 24 (out of 49) precincts had reported Kerry election results that had a chance of less than 5% of occurring. Only 4 precincts gave Bush election share results that had only a 5% chance of occurring.

In summary, the Kyle et. al. ESI report on the Ohio data, like the E/M report on the national data, appears to show unexplained, and statistically significant, exit poll discrepancy patterns. Moreover, as we have already pointed out, even if the Kyle et. al. report did not contain these apparent inconsistencies and misstatements, its conclusions would be of limited value as it does not provide an analysis that links factors that influence WPE to these precinct level exit poll discrepancies, apparently because of a lack of data necessary to perform this analysis. However, this kind of analysis (that is based on “model” and not “data” variance – see above) is essential to providing a statistically substantive exit-poll error based explanation of the discrepancy. To the extent that this kind of substantive statistical explanation for these exit poll discrepancies is not forthcoming, the accuracy of the reported election result is in doubt.

Conclusions

We reiterate that more than ten months after the election, no plausible explanation of the 2004 exit poll discrepancy, based on exit polling error, has been provided by E/M, or by the ESI (the only other group of analysts who have been given access to partial precinct level data – and this only for Ohio). Moreover, the national precinct level exit polling and official vote count data that would enable independent investigators to analyze the exit poll discrepancy, and to confirm or reject explanations has also not been provided to the public.

This lack of an explanation and of the data that could provide an explanation of the 2004 exit poll discrepancy in one of the world’s oldest democracies is unacceptable. It is a matter of the utmost national importance that detailed precinct level exit polling and election data that would allow for investigation by independent analysts, such as USCV, be released.

We remain concerned that the 2004 presidential exit poll data is consistent with a pattern that would be produced by significant vote miscounts, primarily favoring Bush. We believe that U.S. election systems are vulnerable to undetected vote embezzlement and innocent miscounts.

Much work must be done to secure American democracy. USCV's continuing investigation of the 2004 election, including the construction of a national election data archive will contribute to that work. We welcome the efforts of all who share the goal of protecting and strengthening U.S. democratic processes, despite inevitable disagreements.

³⁵ If we assume that our (probable overestimates of the) standard deviations are correct, precinct level exit polls should fall outside this one-sided confidence interval in about 16% of the precincts.

³⁶ This again indicates a significant pro-Kerry exit poll discrepancy as there should be an underestimate of the Kerry reported vote share that falls below (on the other side of confidence interval) in about 16%, and not 4%, of the precincts.

The persistence of credible hypotheses of vote embezzlement and unexplained irregularities in the exit poll data, seven months after the election, underscores the fragility of U.S. elections system. Our country can and must do better.

Recommendations

Few procedures are more important to the strength and health of U.S. democracy than credible and transparent, fair and free, elections. We appeal to our fellow citizens, the media, our courts, and our elected officials, to implement:

- a thorough non-partisan investigation of the 2004 presidential election
- full funding of the National Election Data Archive precinct level database
- election equipment that permits access by non-specialist citizen election judges to recount voter verified paper ballots
- routine 3%, randomly selected, independent audits of vote counts in all elections
- transparent and publicly accessible exit polling
- election administration by non-partisan public civil servants
- non-proprietary open-source coding for all computerized election equipment
- no wired or wireless network connections to any vote casting or counting equipment

Vote counts in America need to be routinely and independently audited. It is not enough to require voter verified paper records of ballots. These paper records must be easily and "independently" auditable by persons other than the voting machine vendor, preferably without having to hire computer technicians, paper roll advancers, bar code readers, and laptops, as is true with many voting systems on the market today.

In particular, 3% of randomly selected precincts can be recounted, using the paper record, immediately when polls close, in the precinct, before removing ballots from the precinct. If discrepancies are found, a county-wide recount can be automatically triggered. Additional funding may need to be allocated to state and county election offices to routinely perform independent audits of vote counts.

In order to monitor the accuracy of vote counting systems, all state and county election offices should set up election data reporting systems to quickly and easily make publicly available, their precinct-level vote totals, broken out by vote type (i.e. election day, absentee, overseas, provisional, early voting, etc.) If vote counts are not reported down to this detailed level, then padded votes in one vote type can easily "cancel out" under-votes in another type. In other words votes can be subtracted from one candidate in one vote type, while being added for another candidate in another vote type, yet these two problems, when added together, may look perfectly normal.

Edison/Mitofsky can materially improve collective understanding of the exit polls -- and whether they are evidence of vote fraud -- by a full release of the data with precinct identifiers and by conducting further tests on the exit poll data, including analysis of variables for voting method and size of place [urban vs. non-urban] to help resolve this issue.

We emphasize that USCV's objective is to ensure that, if vote miscounts favoring either party occurred, it does not occur in the future.

The Future: How would a National Election Data Archive Protect Democracy?

If, for decades, we had never independently audited our financial institutions, we would expect to see ubiquitous insider embezzlement of monies. For decades now, we have counted the vast majority of U.S. votes via mechanical or electronic methods, yet there have never been any routine independent audits of vote counts.

US Count Votes is seeking funding to create the first-ever nation-wide database of precinct-level and vote-type election results in order to statistically audit U.S. vote counts to detect patterns that suggest the embezzlement of votes. To obtain all the needed election data in all its diverse forms from the over 33,000 separate election offices in America is a huge project. Full-time programming staff, statisticians, and administrative staff are needed. For somewhat less than one million dollars, the National Election Data Archive could assist all candidates of any party to determine whether or not their elections were accurately counted, and produce court-worthy evidence that is needed to obtain recounts, investigations, or possibly even re-elections.

The "National Election Data Archive" project is particularly important, given the fact that private exit pollsters could, in the future, elect to adjust exit poll data to conform to actual official election results and neglect to publicly release any "unadjusted" exit poll data.

The development of a "National Election Data Archive" would provide the public with all the data it needs to analyze vote counts within days of the November 2006 election. The technical implementation of well-developed and sound plans for such a system needs to begin very soon, in order to ensure by January 2007 and thereafter, that the candidates actually selected by the voters, are sworn into office. Our hope is that through careful analysis, we can develop the capacity to identify future vote count errors, whether fraudulent or inadvertent, in time to challenge the outcomes.

Appendix A: Voter Response Rate Calculations

Calculated Kerry and Bush voters' response rates required to reconcile Edison/Mitofsky's Precinct Partisanship data as given in Table 1.

We assume that there are no significant differences in precinct size between the various precinct groupings by partisanship. For any assumed percentage of Bush and Kerry voters within any partisanship precinct group, there exist equations where the unknowns are "the response rate for Bush voters" and "the response rate of Kerry voters" that have a single solution.

For Each Partisan Precinct Grouping we let

- N be the number of votes cast in each precinct grouping
- n be the number of voters in the exit poll sample where $n = qN$
- q be the overall sampling proportion for the precinct grouping
- k be the proportion of Kerry votes in the precinct grouping
- b be the proportion of Bush votes in the precinct grouping
- K be the proportion of Kerry voters in the sample who answer the exit poll.
(Kerry voter response rate)
- B be the proportion of Bush voters in the sample who answer the exit poll.
(Bush voter response rate)
- R be the overall response rate within each sample i.e. the proportion of all voters in the sample who answered the exit poll surveys
- E be the mean WPE error for that precinct grouping
- m be the miss rate
- $M = k - b$ be the margin difference in Bush and Kerry percentage votes
- $w = K - B$ be the differential response rate of Kerry and Bush voters
- $\alpha = \frac{K}{B}$ be the ratio of Kerry response rate to Bush response rate

First, we calculate Bush and Kerry response rates, K and B , as a proportion of the *sample*:

- kn is the number of Kerry voters in the sample
- bn is the number of Bush voters in the sample
- kN is the number of Kerry votes/voters in the precinct grouping
- bN is the number of Bush votes/voters in the precinct grouping

knK is the number of Kerry voters in the sample who responded to exit polls

bnB is the number of Bush voters in the sample who responded to the exit polls

Rn is the total number of voters who completed the exit poll in the precinct grouping

$\frac{knK}{Rn}$ and $\frac{bnB}{Rn}$ are the ratios of Kerry and Bush voters who responded to exit polls

$k - 0.5E$ is the ratio of Kerry voters who responded to exit polls given the WPE discrepancy

$b + 0.5E$ is the ratio of Bush voters who responded to exit polls given the WPE discrepancy

Note that $k + b = 1$ and $kK + bB = R$

So that,

$$\frac{knK}{Rn} = k - 0.5E \quad \text{and} \quad \frac{bnB}{Rn} = b + 0.5E$$

Solving for K and B we obtain:

$$\text{Equation 1.} \quad K = \frac{(k - 0.5E)R}{k} \quad \text{and} \quad B = \frac{(b + 0.5E)R}{b}$$

Appendix B: WPE and Differential Partisan Response

WPE is a poor measure of “differential response by party” since its magnitude is affected by the partisan composition of the precinct (k or b) and by the overall response rate (R), in addition to the relative response to exit pollsters by members of each party.³⁷ This can be seen by inspecting Tables 2-4 above. This is because, in addition to differential response by party, overall response rates and Bush/Kerry vote rates affect WPE

This can be seen by setting $K = r - .5w$ and $B = r + .5w$, where $w = B - K$ is "differential response by party" and r is “mean response by party”, and substituting these into the solutions for K and B in Equation 1. Appendix A, to get the following two equation system for r and w :

$$\frac{(k - 0.5E)R}{k} = r - 0.5w \quad \text{and} \quad \frac{(b + 0.5E)R}{b} = r + 0.5w$$

The solution of this system for w is:

$$w = \left(\frac{R}{b}\right)\left(\frac{E}{2} + b\right) + \left(\frac{R}{k}\right)\left(\frac{E}{2} - k\right)$$

so that when $b + k = 1$ as we assume in this report (neglecting the response of “independent voters” who made up 1% or so of the national electorate) we get:

$$\text{Equation 3.} \quad w = \left(\frac{RE}{2}\right)\left(\frac{1}{b(1-b)}\right) = \left(\frac{RE}{2}\right)\left(\frac{1}{k(1-k)}\right) \quad \text{or conversely}$$

$$\text{Equation 4.} \quad E = \left(\frac{2w}{R}\right)b(1-b) = \left(\frac{2w}{R}\right)k(1-k)$$

From equation 4, we see that E , or WPE, increases in size as the difference between the Bush and Kerry response rates, w , increases, and is largest when k or b goes to 0.5, and is smaller when the overall response rate, R , increases. Competitive precincts (when $b \rightarrow 0.5$ $k \rightarrow 0.5$) and precincts with lower overall response rates R , will have higher absolute WPE.

The difference in response rates between Kerry and Bush voters, w , will *not* be equal to E unless $\left(\frac{2}{R}\right)k(1-k) = \left(\frac{2}{R}\right)b(1-b) = 1$ for a perfectly competitive precinct ($b = k = 0.5$) $E = \left(\frac{0.5}{R}\right)w$, its maximum value. If, in addition, $R = 0.5$, then $E = w$.

This suggests that the WPEs listed in Tables 2-4 of our report substantially understate the differences between Bush and Kerry response rates, especially for high Bush precincts. As we have shown, w , the difference of Bush and Kerry voter exit poll response rates, has to be implausibly large in all cases, if WPEs are to be explained.

If there were a pervasive and more or less constant bias in exit polling because of a differential response by party, WPE should be greatest in absolute value for more balanced precincts and fall towards zero as precincts become more partisan. The data presented on p. 36, 37 of the E/M report

³⁷ We thank Elizabeth Liddle, of the University of Nottingham, U.K., for calling our attention to the effect of precinct partisanship on the relationship between differential partisan response and WPE. We take full responsibility for the derivations and conclusions that we have arrived at from analyzing this pattern in this Appendix.

and displayed in Table 1 of our report above, show that this is the pattern *for all except the most highly partisan Bush precincts for which WPE dramatically increases in size to -10.0%. The difference in Bush and Kerry exit poll response rates necessary to generate this level of WPE in these precincts ranges from 40% (Table 2) to an absolute minimum of 20.5% (Table 4).* This pattern of higher WPE in Bush strongholds could be caused by either "Bush Strongholds have More Vote-Corruption" (Bsvcc) or if enough votes were shifted from Kerry to Bush that it caused precincts from lower categories with higher WPE to shift into precinct categories with higher proportion of Bush votes.

Signed WPE versus Absolute WPE in Partisan Precincts

The mean absolute value WPE (unsigned WPE) for high Bush vote precincts *declines* to 12.4 % relative to lesser Bush vote precincts (having WPEs of 13.2% and 13.4%) and more balanced precincts (15.2%). Only highly partisan Kerry precincts have a lower mean absolute value WPE of 8.8%.³⁸

Comparing this to the mean WPEs in Table 2 shows that:

a) High Kerry vote precincts had large absolute value WPE's (totaling 8.8%) but these included both pro-Kerry and pro-Bush discrepancies that off-set each other so that the average (signed) WPE was only 0.3%. Pro-Bush bias was almost offset by pro-Kerry bias in these precincts, as one would expect for random sampling bias and random measurement error.

b) Similar, but less balanced, patterns occurred in more evenly divided precincts where (signed) WPE, though consistently negative, was roughly half the magnitude of mean absolute value WPE (signed and unsigned values being -5.5 and 13.4, -8.3 and 15.2, -6.1 and 13.2, respectively). This suggests that in these precincts about half of pro-Kerry exit poll bias was offset by pro-Bush exit poll bias. While this is not what one would expect from random exit poll bias and measurement error, it at least moves in the expected direction.

c) The dramatic and unexpected increase in (signed) mean WPE in highly Bush precincts of -10.0% is also unexpectedly close to the mean absolute value WPE (12.4%) in these precincts. *This suggests that the jump in (signed) WPE in high Bush vote precincts occurred primarily because (signed) WPE discrepancies in these precincts were, unlike in a) above, and much more so than in b) above, overwhelmingly one-sided negative overstatements of Kerry's vote share.*

These results lend further support to the "Bush Strongholds have more Vote-Corruption" (Bsvcc) hypothesis³⁹ or alternatively this pattern could be produced by vote shifts to Bush in precincts that would normally fall into categories with fewer Bush vote percentages.

We would like Edison/Mitofsky to explain why signed WPE in highly partisan precincts is not lower than in less partisan precincts as would be mathematically expected, and why these patterns are at odds with the more or less random pattern of signed WPE error in highly Kerry precincts.

38 p. 36 of E/M report op. cit to E-M report URL in footnote 1.

³⁹ Discussed in

http://electionarchive.org/ucvAnalysis/US/Exit_Polls_2004_Edison-Mitofsky.pdf, in our discussion of the "Implausible Patterns of Exit Poll Participation Are Required to Satisfy E/M's data in 80-100% Bush Precincts".

Appendix C: Exit Poll Response Bias Using the Ratio of Kerry and Bush Response Rates

Let $\alpha = \frac{K}{B}$, the ratio of Kerry voters' exit poll response rate to Bush voters' exit poll response rate

Using Equations 1. In Appendix A which gives the response rates for Kerry and Bush voters, K and B , in terms of the WPE, E , the percentages of Bush and Kerry votes, b and k , and the overall response rate R , we can calculate

$$\alpha = \left(\frac{K}{B}\right) = \frac{(k - 0.5E)\left(\frac{R}{k}\right)}{(b + 0.5E)\left(\frac{R}{b}\right)} \quad \text{which reduces to}$$

$$\text{Equation 5.} \quad \alpha = \frac{b(k - 0.5E)}{k(b + 0.5E)}$$

The above equation gives a measure for exit poll partisanship response bias in terms of WPE and the percent of Kerry and Bush votes.⁴⁰ Taking the natural logarithm of Equation 5. produces a bias index that is symmetric around zero. i.e. 2 and 0.5 don't average out to 1, but $\ln 2$ and $\ln 0.5$ do average out to $\ln 1$.

$$\text{Equation 6.} \quad \log_e \alpha = \log \left(\frac{K}{B}\right) = \log_e \left[\frac{b(k - 0.5E)}{k(b + 0.5E)} \right]$$

Appendix D: WPE Error in Terms of Exit Poll Response Bias $\alpha = \frac{K}{B}$

Solving Equation 5. in Appendix C $\alpha = \frac{b(2k - E)}{k(2b + E)}$ for E , gives

$$\text{Equation 7.} \quad \text{WPE} = \frac{2bk(1 - \alpha)}{k\alpha + b}$$

This gives WPE error, E , in terms of the percentages of Bush and Kerry voters, k and b , and the response bias factor α .

⁴⁰ Note: The equations here are equivalent to those used by Liddle in her paper published on April 16, 2005 (see <http://www.geocities.com/lizzielid/WPEpaperARCH.pdf>). USCV disagree with Liddle's conclusions.

Appendix E: WPE Error Caused by Vote Shift

This section uses the same notation and variables defined in the beginning of Appendix A.

WPE = Vote Margin - Exit Poll Margin

Let

k_s be the Kerry vote percentage after the vote has been shifted
 b_s be the Bush vote percentage after the vote has been shifted
 s be the amount that the vote has been shifted

where $k_s + b_s = 1$

Then $k_s = k - 0.5s$ and $b_s = b + 0.5s$

and the Vote Margin is $\frac{(k_s - b_s)}{(k_s + b_s)}$

and from Appendix A. the Exit Poll Margin = $\frac{(kK - bB)}{(kK + bB)}$

Then we have:

$$\text{Equation 8: } \text{WPE}_{\text{voteshift}} = \frac{(k_s - b_s)}{(k_s + b_s)} - \frac{(kK - bB)}{(kK + bB)}$$

Note that the "shifted" values for Kerry and Bush vote percentages are used to calculate the Vote Margin, but the original un-shifted values for Kerry and Bush vote percentages are used to calculate the Exit Poll Margin.

The above formula can be used to determine WPE that would result from vote shifting from Kerry to Bush.

Appendix F: Existing "Exit Poll Error" Explanations for the Exit Poll Discrepancies are Not Consistent with the Aggregate Data.

Statistics such as means and medians are generally regarded as good descriptors of reasonably well behaved (approximately normally distributed) aggregate data like exit poll data. Simple parameters calculated from means and medians of aggregate data are regarded as fairly good indicators of the mean and median values of the same parameters calculated from individual data points.

The following table displays mean and median calculations for a "bias" or "Alpha" parameter (Alpha=K/B in our Appendix A notation) – see Liddle (previous citation). We have also included the values for B, K and w derived in our report (Table 2, Appendix A, Appendix B notation).

As can be seen from this table, it is highly unlikely that the actual mean and median of Alpha, calculated from precinct level data, would be equal across categories of precincts.

Precinct Vote			Mean Based Calculations					Median Based Calculations				
Bush	Kerry	R	Mean WPE	B	K	Alpha	w	Median WPE	B	K	Alpha	w
0.1	0.9	53%	0.3%	53.8%	52.9%	0.98	0.9%	-0.4%	51.9%	53.1%	1.02	-1.2%
0.3	0.7	55%	-5.9%	49.6%	57.3%	1.16	-7.7%	-5.5%	50.0%	57.2%	1.14	-7.2%
0.5	0.5	52%	-8.5%	47.6%	56.4%	1.19	-8.8%	-8.3%	47.7%	56.3%	1.18	-8.6%
0.7	0.3	55%	-6.1%	52.6%	60.6%	1.15	-8.0%	-6.1%	52.6%	60.6%	1.15	-8.0%
0.9	0.1	56%	-10.0%	52.9%	84.0%	1.59	-31.1%	-5.8%	54.2%	72.2%	1.33	-18.0%

The following Table 2 (calculated from means) shows that a fixed "Alpha" will allow for some WPE asymmetry across precincts. However, a level of Alpha (1.15) will generate the roughly 6.5% overall WPE reported by E/M, this "ratio" effect is hardly enough to account for the highly asymmetric differences in WPE by precinct partisanship displayed in Table 1. The "ratio effect" is a purely mathematical result of linking a *ratio* of percentages (Alpha) to a *difference* in percentages (WPE).

Bush	Kerry	Alpha	R	r	w	WPE	B	K
0.1	0.9	1.15	53%	50.20%	-7.0%	-2.4%	46.7%	53.7%
0.3	0.7	1.15	55%	53.51%	-7.5%	-5.7%	49.8%	57.2%
0.5	0.5	1.15	52%	52.00%	-7.3%	-7.0%	48.4%	55.6%
0.7	0.3	1.15	55%	56.58%	-7.9%	-6.0%	52.6%	60.5%
0.9	0.1	1.15	56%	59.31%	-8.3%	-2.7%	55.2%	63.4%

Both Tables 2 and 3 show that the most important effect of either a fixed "Alpha" or a fixed "w" on WPE is to greatly reduce WPE in highly partisan precincts in direct contradiction to the large increase in mean WPE, and small decline in median WPE, in 90% Bush precincts in Table 1.

Table 3, below, calculated from means, shows that a constant “partisan response differential” $w = K - B$ (see Appendix B) does *not* generate asymmetric WPE by precinct partisanship.

Table 3									
Alpha and WPE with $w = -6.5\%$ and Varying R									
Based on $r = R - w(1-k)/w + wk/2$ from Appendix B									
Bush	Kerry	w	R	r	B	K	WPE	Alpha	
0.1	0.9	-6.5%	53%	50.4%	47.2%	53.7%	-2.3%	1.138	
0.3	0.7	-6.5%	55%	53.7%	50.5%	57.0%	-5.5%	1.129	
0.5	0.5	-6.5%	52%	52.0%	48.8%	55.3%	-6.5%	1.133	
0.7	0.3	-6.5%	55%	56.3%	53.1%	59.6%	-5.5%	1.123	
0.9	0.1	-6.5%	56%	58.6%	55.4%	61.9%	-2.3%	1.117	

Finally as the relation between $\text{Alpha} = K/B$ and WPE is one-to-one and independent of R, Alpha can be calculated based on WPE for any possible value of R as is done in Table 4 below.⁴¹ The equation for $\text{Alpha} = K/B$ is obtained from the equations for K and B (Equation 1 in Appendix A). As can be seen R drops out of this equation. Both mean and median values of WPE are used to estimate Alpha. As can be seen in the Table, estimated biases calculated from either mean or median WPE levels vary significantly across precinct partisanship categories suggesting that “rBr” is not a viable hypothesis.

Table 4						
Alpha Generated from WPE						
Based on $\text{Alpha} = ((k-0.5E)/k)/(b+0.5E)/b$						
Bush	Kerry	Alpha from Mean	Mean WPE	Alpha from Median	Median WPE	
0.1	0.9	0.98	0.3%	1.02	-0.4%	
0.3	0.7	1.16	-5.9%	1.14	-5.5%	
0.5	0.5	1.19	-8.5%	1.18	-8.3%	
0.7	0.3	1.15	-6.1%	1.15	-6.1%	
0.9	0.1	1.59	-10.0%	1.33	-5.8%	

The following Table 5 shows mean WPE levels that would be generated by a Alpha of 1.12 corresponding to the “rBr hypothesis” of $K=0.56$ and $B=0.5$ ($K/B=1.12$).

⁴¹ I’m indebted to Prof. Mark Lindeman of Bard College for pointing out and insisting that an earlier version of Table 4 had an error. His (correct) insistence induced me to produce the corrected Tables 4 and 5 below.

Table 5

WPE Generated by Alpha
Based on $E = (kb(1-\alpha))/(0.5(\alpha*k+B))$

Bush	Kerry	Alpha	WPE from Alpha	Mean WPE	Median WPE
0.1	0.9	1.12	-1.9%	0.3%	-0.4%
0.3	0.7	1.12	-4.6%	-5.9%	-5.5%
0.5	0.5	1.12	-5.7%	-8.5%	-8.3%
0.7	0.3	1.12	-4.9%	-6.1%	-6.1%
0.9	0.1	1.12	-2.1%	-10.0%	-5.8%

Table 5 shows how unlikely it is that such a hypothesis could explain: the high absolute WPE in high Bush precincts, the low absolute WPE in high Kerry precincts, and the high absolute WPE levels in more competitive precincts.

These mean and median estimates suggest that in order to explain the exit polling discrepancy alpha would have to be much *lower* in high Kerry precincts, and significantly *higher* in high Bush precincts and in more competitive precincts.

Finally applying the “theorem” from Appendix B showing that the largest WPE for any given partisan response differential or bias (B-K or K/B) will be in the most competitive precinct, we calculate the minimal partisan response rates possible, given reported mean and median WPE and overall response levels by assuming the most “competitive precinct distribution” possible in each of the precinct partisanship categories. Table 6 shows that a *minimal* weighted average bias of K=0.58 to B=0.50 is necessary to generate the reported mean and median WPE and overall R levels.

Table 6

Mean and Median Based Minimal Average B and K

Precinct Vote			Sample	Mean	B from	K from	Median	B from	K from
Bush	Kerry	R	Size	WPE	Mean	Mean	WPE	Median	Median
0.19	0.81	0.53	90	0.3%	53.4%	52.9%	-0.4%	52.4%	53.1%
0.39	0.61	0.55	165	-5.9%	50.8%	57.7%	-5.5%	51.1%	57.5%
0.5	0.5	0.52	540	-8.5%	47.6%	56.4%	-8.3%	47.7%	56.3%
0.61	0.39	0.55	415	-6.1%	52.3%	59.3%	-6.1%	52.3%	59.3%
0.81	0.19	0.56	40	-10.0%	52.5%	70.7%	-5.8%	54.0%	64.5%
			1250						
Weighted Averages:					50.1%	57.7%		50.2%	57.5%

These tables all indicate that it is highly improbable that the exit poll outcomes described in the E/M report are a result of any randomly distributed “bias” that has a uniform mean and median. Rather the data suggest that significantly varying, and mostly non-zero mean and median bias, with an overall average of at least 1.16 (K=0.58 and B=0.50) across partisan precincts is required for any “exit polling error” explanation of the exit poll discrepancy.

Appendix G: Aggregate Calculations Suggest that Changes in Partisan Response Rates Across Partisan Precinct Categories Cannot be Explained by Random Sampling Error

For each precinct, for a perfect random sample $K=B=R$ where R is the precinct overall response rate. So each Kerry or Bush voter has an R chance of completing the exit poll survey. This implies that the sampling distribution of K for each precinct has $\text{mean}=R$ and variance equal to: $R(1-R)/nk$ where nk is the number of Kerry voters in the precinct (and similarly for B). Now we know that R will vary from precinct to precinct (though its average across precincts is from 52% to 56%) so we conservatively set it at 50% to maximize precinct variance. This implies that: $R(1-R) = 0.5 \times 0.5 = 0.25$.

Since precinct samples are independent, the variance of the sum of the precinct specific variances of K in a given partisan category with N precincts will be:

$$\text{Equation 9: } V_{\sum K} = \left(\frac{1}{n_1 k_1} + \frac{1}{n_2 k_2} + \dots + \frac{1}{n_N k_N} \right) 0.25$$

where n_1, \dots, n_N are the sizes of the precincts for example for high Kerry precincts 1 through N , and k_1, \dots, k_N are their respective reported Kerry vote shares.

The variance of the *mean* of K for a given precinct category will therefore be equal to:

$$\text{Equation 10: } V_{\bar{K}} = \frac{\left(\frac{1}{n_1 k_1} + \frac{1}{n_2 k_2} + \dots + \frac{1}{n_N k_N} \right) 0.25}{N^2}$$

so that the standard deviation of \bar{K} is:

$$\text{Equation 11: } S_{\bar{K}} = \frac{\sqrt{\left(\frac{1}{n_1 k_1} + \frac{1}{n_2 k_2} + \dots + \frac{1}{n_N k_N} \right) 0.25}}{N}$$

And similarly for B .

We can derive *maximal standard deviations* from these formulas by using known lower bounds for n_i and k_i .

For example, applying these estimates to K for the 90 high Kerry precincts ($k \geq 0.80$) we get:

$$V_{\sum K} = \left(\frac{1}{n_1 k_1} + \frac{1}{n_2 k_2} + \dots + \frac{1}{n_{90} k_{90}} \right) 0.25 \leq \frac{90 \times 0.25}{20 \times 0.8} = 1.41$$

as the minimal number of respondents in any precinct is 20 (see E-M report p. 34) and the minimal share of Kerry voters in this category is 0.8.

The variance of the *mean* will thus be less or equal to $\frac{1.41}{90^2}$, and the standard deviation of the mean of these proportions will be less or equal than: $\frac{\sqrt{1.41}}{90} = 0.0132$

This maximal standard deviation will be *considerably larger* than the actual standard deviation of K , as $n_i k_i$ is generally greater than $20 \times 0.8 = 16$. However, as we have no way of knowing how much smaller the actual standard deviation is, we conservatively set it at its highest possible level.

Using similar estimates for the other categories of precincts and for the standard deviation of B we derive the following tables. The formulas for maximal standard deviations used in the tables are derived above. The formulas for a test of difference of mean proportions are standard and can be found in any introductory statistics text.

Table 1 shows that even with these maximal sampling error standard deviations, the differences in K values are highly significant, often at the 1% level (see bolded red values). This implies that these differences in K cannot be a result of random sampling error and therefore must be explained by other exit polling factors if an exit polling error explanation for these data exists.

		k=.90	k=.70	k=.50	k=.30	k=.10
	N=	90	165	540	415	40
	K=	0.529	0.573	0.564	0.606	0.84
	S=(N*0.25)/(MinK*20)	1.40625	3.4375	16.875	25.9375	50
	SD(K)=SQRT(S)/N	0.01318	0.01124	0.0076	0.01227	0.17678
	R=	0.53	0.55	0.52	0.55	0.56
1	90 0.529 0.01318 0.53					
2	165 0.573 0.01124 0.55	2-1 0.044 0.557 0.065 0.2495	=categories compared =K 2 - K1 =K0, if K2=K1=K0 =random stddev of K2-K1 =one tail K2>K1 prob of K2-K1 if K2=K1=K0			
3	540 0.564 0.00761 0.52	3-1 0.035 0.559 0.057 0.2679	2-3 0.009 0.566 0.044 0.4191			
4	415 0.606 0.01227 0.55	4-1 0.077 0.592 0.057 0.0889	4-2 0.033 0.597 0.045 0.2324	4-3 0.042 0.582 0.032 0.0960		
5	40 0.84 0.17678 0.56	5-1 0.311 0.625 0.092 0.0004	5-2 0.267 0.625 0.085 0.0009	5-3 0.276 0.583 0.081 0.0003	5-4 0.234 0.627 0.080 0.0017	

Table 2 below shows that maximal sampling standard deviation estimates show no significant differences between B values with the exception of categories 4 and 3. This suggests that most of the difference in bias, or “Alpha” (=K/B), are a result of changes in K with a relatively constant B.

		k=.90	k=.70	k=.50	k=.30	k=.10
	N=	90	165	540	415	40
	B=	0.538	0.496	0.476	0.526	0.529
	S=(Nx0.25)/(MinBx20)	112.500	10.313	16.875	8.646	0.625
	SD(K)=SQRT(S)/N	0.11785	0.01946	0.0076	0.00709	0.01976424
	R=	0.53	0.55	0.52	0.55	0.56
1	90 0.538 0.11785 0.53					
2	165 0.496 0.01946 0.55	2-1 -0.042 0.511 0.066 0.7393	=categories compared =B 2 - B1 =B0, if B2=B1=B0 =random stddev of B2-B1 =one tail B2>B1 prob of B2-B1 if B2=B1=B0			
3	540 0.476 0.00761 0.52	3-1 -0.062 0.48486 0.057 0.8621	2-3 0.02 0.481 0.044 0.3263			
4	415 0.526 0.007 0.55	4-1 -0.012 0.528 0.058 0.5819	4-2 0.03 0.517 0.046 0.2571	4-3 0.05 0.498 0.033 0.0628		
5	40 0.529 0.020 0.56	5-1 -0.009 0.53523 0.09478 0.5378	5-2 0.033 0.50244 0.08812 0.3540	5-3 0.053 0.4797 0.0819 0.2587	5-4 0.003 0.52626 0.08267 0.4855	

Tables 3 and 4 below give exactly the same results, in terms of significant difference, for K and B calculated from medians rather than means. This suggests that the results of Tables 1 and 2 are not dependent on individual outliers but rather are robust characteristics of K and B in these partisan precinct categories.

Note again that the results in all of these tables are based on maximal standard error estimates. Exact sampling error calculations from precinct level data will generate much smaller sampling standard errors and, most likely, more significant differences in K and B values between partisan precinct categories. However, precinct level data is necessary to do these more accurate tests of difference of proportions.

		k=.90	k=.70	k=.50	k=.30	k=.10
	N=	90	165	540	415	40
	K=	0.531	0.572	0.563	0.606	0.722
	$S=(N \times 0.25) / (\text{Min} K \times 20)$	1.40625	3.4375	16.875	25.9375	50
	$SD(K)=\text{SQRT}(S)/N$	0.01318	0.01124	0.0076	0.01227	0.17678
	R=	0.53	0.55	0.52	0.55	0.56
1	90 0.531 0.01318 0.53					
2	165 0.572 0.01124 0.55	2-1 0.041 0.558 0.065 0.2644	=categories compared =K 2 - K1 =K0, if K2=K1=K0 =random stddev of K2-K1 =one tail K2>K1 prob of K2-K1 if K2=K1=K0			
3	540 0.563 0.00761 0.52	3-1 0.032 0.55843 0.057 0.2857	2-3 0.009 0.565 0.044 0.4191			
4	415 0.606 0.01227 0.55	4-1 0.075 0.593 0.057 0.0946	4-2 0.034 0.596 0.045 0.2257	4-3 0.043 0.582 0.032 0.0909		
5	40 0.722 0.17678 0.56	5-1 0.191 0.590 0.093 0.0205	5-2 0.15 0.601 0.086 0.0411	5-3 0.159 0.574 0.081 0.0249	5-4 0.116 0.616 0.081 0.0748	

Table 4: Minimally Significant Differences of Proportions for B Calculated from Median WPE

		k=.90	k=.70	k=.50	k=.30	k=.10
	N=	90	165	540	415	40
	B=	0.519	0.500	0.477	0.526	0.542
	S=(N \times 0.25)/(MinB \times 20)	112.500	10.313	16.875	8.646	0.625
	SD(K)=SQRT(S)/N	0.11785	0.01946	0.0076	0.00709	0.0197642
	R=	0.53	0.55	0.52	0.55	0.56
1	90					
	0.519					
	0.11785					
	0.53					
		2-1	=categories compared			
2	165	-0.019	=B 2 - B1			
	0.500	0.507	=B0, if B2=B1=B0			
	0.01946	0.066	=random stddev of B2-B1			
	0.55	0.6141	=one tail B2>B1 prob of B2-B1 if B2=B1=B0			
		3-1	2-3			
3	540	-0.042	0.023			
	0.477	0.483	0.482			
	0.00761	0.057	0.044			
	0.52	0.7698	0.3024			
		4-1	4-2	4-3		
4	415	0.007	0.026	0.049		
	0.526	0.525	0.519	0.498		
	0.007	0.058	0.046	0.033		
	0.55	0.4520	0.2859	0.0666		
		5-1	5-2	5-3	5-4	
5	40	0.023	0.042	0.065	0.016	
	0.542	0.52608	0.5082	0.4815	0.52741	
	0.020	0.09489	0.08811	0.0819	0.08265	
	0.56	0.4042	0.3168	0.2136	0.4233	

Finally, Table 5 below shows that almost all of the mean and median K and B values deviate significantly (even with a maximal standard deviation) from their true random sampling value of R except for K and B in high Kerry precincts (k = 0.90). Moreover, the significant discrepancies are all in the same direction, K>R and B<R. These results suggest exit poll, or vote miscount, error in all partisan precinct categories except for high Kerry (k>=0.80) precincts.

The only insignificant differences from R in other precincts are for K and B median values for high Bush precincts (b = 0.90), but these significance levels are low (below 0.18) and are likely to become significant with a more realistic (smaller) standard deviation.

Appendix H: Estimation of Precinct Level Standard Deviations from Table 1 of the ESI Report on the Ohio Exit Polls.

If $\frac{K}{K+B}$ is the Kerry precinct level exit poll share and N are non-responders then:

$K_{\min} = \frac{K}{K+B+N}$ is the minimum Kerry exit poll vote share, and:

$K_{\max} = \frac{K+N}{K+B+N}$ is the maximum Kerry vote share,

shown in Table 1 of the Kyle et. al. ESI report.⁴²

So the share of non-responders is:

$$N_s = \frac{N}{K+B+N}.$$

This implies that:

$$Range = K_{\max} - K_{\min} = \frac{K+N}{K+B+N} - \frac{K}{K+B+N} = N_s$$

If a share X of N is allocated to Kerry then the Kerry vote would be: $\frac{K+XN}{K+B+N}$, and the

Bush vote would be: $\frac{B+(1-X)N}{K+B+N}$.

So if $\frac{K+XN}{K+B+N} = \frac{K}{K+B}$ the original exit poll value, we must have: $X = \frac{K}{K+B}$. Non-responders to an exit poll that is a true random sample of the precinct will split this way,

so that, in the absence of sampling and other exit poll error, this should be the “true” value of X.

The reported election results for Kerry will differ from the exit poll results to the extent that non-responders don't split this way.

So for X to generate the reported Kerry vote share k, we need:

⁴² See: http://www.votewatch.us/reports/view_reports.

$$\frac{K + XN}{K + B + N} = k, \text{ so that: } X = \frac{k - \frac{K}{K + B + N}}{\frac{N}{K + B + N}} = \frac{k - K_{\min}}{N_s} = \frac{k - K_{\min}}{\text{Range}}$$

This is the actual value of X that is necessary to generate the reported official election result given the exit poll results.

Based on this expression for X, and assuming non-responder samples of 30⁴³, we generate estimates for the probability of the non-responder split necessary to produce the reported official Kerry vote, given the exit poll result (or the one-tail P-value for X), using the data provided in Table 1 of the Kyle et. al. ESI brief, in Table 1 below:

⁴³ This is a conservative assumption as the average exit poll response for Ohio was 2042/49=41.7 – see www.exit-poll.net/election-night/MethodsStatementStateGeneric.pdf and the response, or completion rate, was between 53% and 56%.

Table 1: Statistical Analysis of Ohio Kerry Precinct Level Exit Poll and Official Election Results

1	2	3	4	5	6	7	8	9	10	11	12	13
Infeasible Outcome	Mitofsky Precinct Number	Official Vote ^a	Exit Poll		Exit Poll minus Official Vote	Non-Respondent Share (X)	Maximal SD for X with 30 Non-Respondents	Signed one SD for X with 30 Non-Respondents	One Tail P-Value of X with 30 Non-Respondents	Outside of 95% One-tail Confidence Interval for Pro-Kerry Discrepancy	Outside of 95% One-tail Confidence Interval for Bush Discrepancy	Mean WPE Values by Quintiles of Official Vote Shares
		Minimum	Original	Maximum	(4)-(2)	((2)-(3))/((5)-(3))	$\text{sqrt}((4) \times (1-(4))/30)$	$(8) \times \text{Sign}(6)$	$\text{normdist}((7),(4),(8),1)$	$(10) < 0.05$	$(10) > 0.95$	Average((6))
	48	0.22	0.11	0.38	0.81	0.16	0.16	0.09	0.09	0.01	x	
	14	0.24	0.18	0.28	0.54	0.04	0.17	0.08	0.08	0.08		
	7	0.25	0.25	0.34	0.53	0.09	0.00	0.09	0.09	0.00	x	
	23	0.28	0.09	0.31	0.79	0.03	0.27	0.08	0.08	0.32		
	26	0.28	0.12	0.26	0.67	-0.02	0.29	0.08	-0.08	0.65		
	2	0.30	0.20	0.41	0.71	0.11	0.20	0.09	0.09	0.01	x	
	3	0.30	0.22	0.41	0.69	0.11	0.17	0.09	0.09	0.00	x	
	37	0.30	0.19	0.32	0.59	0.02	0.28	0.09	0.09	0.30		
	29	0.32	0.18	0.30	0.59	-0.02	0.34	0.08	-0.08	0.69		
a	47	0.32	0.36	0.43	0.53	0.11	-0.24	0.09	0.09	0.00	a	
	21	0.34	0.31	0.44	0.61	0.10	0.10	0.09	0.09	0.00	x	
	28	0.34	0.20	0.43	0.73	0.09	0.26	0.09	0.09	0.03	x	
	6	0.36	0.13	0.53	0.88	0.17	0.31	0.09	0.09	0.01	x	
	15	0.37	0.24	0.39	0.63	0.02	0.33	0.09	0.09	0.26		
	43	0.37	0.30	0.49	0.68	0.12	0.18	0.09	0.09	0.00	x	
	17	0.38	0.12	0.48	0.87	0.10	0.35	0.09	0.09	0.07		
	19	0.38	0.28	0.45	0.66	0.07	0.26	0.09	0.09	0.02	x	
	27	0.38	0.11	0.67	0.95	0.29	0.32	0.09	0.09	0.00	x	
	30	0.39	0.21	0.50	0.79	0.11	0.31	0.09	0.09	0.02	x	
a	25	0.40	0.42	0.68	0.80	0.28	-0.05	0.09	0.09	0.00	a	
	18	0.42	0.18	0.46	0.79	0.04	0.39	0.09	0.09	0.23		0.10
	1	0.43	0.19	0.50	0.81	0.07	0.39	0.09	0.09	0.11		
	40	0.43	0.25	0.39	0.61	-0.04	0.50	0.09	-0.09	0.89		
	11	0.45	0.08	0.41	0.88	-0.04	0.46	0.09	-0.09	0.72		
	46	0.45	0.40	0.47	0.55	0.02	0.33	0.09	0.09	0.07		
	39	0.46	0.32	0.54	0.73	0.08	0.34	0.09	0.09	0.01	x	
	13	0.47	0.24	0.47	0.74	0.00	0.46	0.09	0.09	0.46		
	22	0.47	0.30	0.41	0.58	-0.06	0.61	0.09	-0.09	0.99		
	5	0.48	0.17	0.41	0.76	-0.07	0.53	0.09	-0.09	0.90	x	
	34	0.48	0.19	0.54	0.84	0.06	0.45	0.09	0.09	0.15		
	16	0.51	0.47	0.57	0.64	0.06	0.24	0.09	0.09	0.00	x	
	36	0.52	0.44	0.66	0.77	0.14	0.24	0.09	0.09	0.00	x	
	50	0.52	0.43	0.58	0.69	0.06	0.35	0.09	0.09	0.00	x	
	20	0.54	0.40	0.69	0.82	0.15	0.33	0.08	0.08	0.00	x	
	42	0.54	0.28	0.66	0.85	0.12	0.46	0.09	0.09	0.01	x	
	49	0.54	0.14	0.58	0.90	0.04	0.53	0.09	0.09	0.28		
	4	0.55	0.28	0.70	0.88	0.15	0.45	0.08	0.08	0.00	x	
	44	0.55	0.11	0.55	0.91	0.00	0.55	0.09	-0.09	0.50		
	31	0.57	0.28	0.68	0.87	0.11	0.49	0.09	0.09	0.01	x	
	38	0.57	0.11	0.41	0.84	-0.16	0.63	0.09	-0.09	0.99		
	35	0.62	0.53	0.75	0.82	0.13	0.31	0.08	0.08	0.00	x	
	9	0.64	0.29	0.67	0.86	0.03	0.61	0.09	0.09	0.26		
	41	0.66	0.33	0.57	0.75	-0.09	0.79	0.09	-0.09	0.99	x	
	12	0.70	0.33	0.68	0.85	-0.02	0.71	0.09	-0.09	0.64		
	32	0.71	0.55	0.82	0.88	0.11	0.48	0.07	0.07	0.00	x	
	8	0.80	0.41	0.90	0.95	0.10	0.72	0.05	0.05	0.00	x	
	33	0.81	0.19	0.68	0.91	-0.13	0.86	0.09	-0.09	0.98		
	24	0.85	0.64	0.87	0.90	0.02	0.81	0.06	0.06	0.16	x	
	10	0.96	0.40	0.96	0.99	0.00	0.95	0.04	0.04	0.38		-0.04