



---

Overreporting the Vote in the 1988 Senate Election Study: A Response to Wright

Author(s): Paul Gronke

Reviewed work(s):

Source: *Legislative Studies Quarterly*, Vol. 17, No. 1 (Feb., 1992), pp. 113-129

Published by: [Comparative Legislative Research Center](#)

Stable URL: <http://www.jstor.org/stable/440084>

Accessed: 29/05/2012 01:07

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Comparative Legislative Research Center is collaborating with JSTOR to digitize, preserve and extend access to *Legislative Studies Quarterly*.

<http://www.jstor.org>

PAUL GRONKE  
*Duke University*

## ***Overreporting the Vote in the 1988 Senate Election Study: A Response to Wright***

This paper discusses the misreport of votes in Senate elections, a potentially serious flaw discovered by Gerald Wright in the 1988 National Election Study's Senate Election Study (NES/SES). In a recent issue of this journal (November 1990), Wright reports that more respondents said they voted for the Senate winner than actually could have. The probability of erroneous reports increases as the gap between the election date and the interview date widens. This effect interacts with the margin of victory (i.e., the larger the margin, the larger the overreport). The consequence, Wright maintains, is biased coefficients in voting models—the influence of presidential vote on Senate vote is underestimated while candidate-based effects are overestimated (Wright 1990).

I question Wright's conclusions about both the cause and effect of misreporting vote. Wright's explanation of the cause of overreports suffers from two shortcomings. First, the purported cause does not explain why respondents overreport only their Senate vote and not presidential or House choices. Second, the model Wright uses to identify the cause is misspecified; the inclusion of a dummy variable that measures voting for the Republican candidate is inappropriate. When the model is specified as theory directs, with the vote for the winner as the variable of interest, the overreporting problem disappears.

Finally, I take issue with the way Wright demonstrates the effects of bias: he compares equations estimated using the NES/SES data with those estimated using data from network exit polls. Wright's argument is that exit polls are conducted on election day and are therefore free of time-based errors. Thus, structural equations estimated

using exit poll data are the standard against which all other polls should be evaluated. Wright lets exit polls off too lightly; he does not turn his critical eye on exit poll data. Many differences between academic surveys and exit polls besides interview date could cause differences in regression coefficients. If measurement error has been adequately modelled, the source of error can be included as an additional regressor. When this is done with the NES/SES data the resulting coefficient estimates do not look at all as Wright suggests they should (in fact, they are unchanged). I close by presenting my own model of overreporting, which may help set an agenda for future research in this area.

### Wright's Problem: Overreport of Senate Votes<sup>1</sup>

A sample survey attempts to measure as accurately as possible the attitudes, opinions, and demographic characteristics of the target population. The Senate study measures, among other things, Senate vote. Ideally, the correlation between vote choices reported by survey respondents and actual election outcomes should be 1.0, with error due to sampling variance disregarded. At the state level, the expected value of a survey estimate of the Republican, the Democratic, or the winning vote percentage should be the actual vote percentage. Variation across states in the proportion of respondents who said they voted a certain way ought to bear a 1:1 relationship with actual state-to-state variation. Following Wright (1990), let  $RV_n$  = observed percentage vote going to the Republican and  $AV_n$  = the actual percentage Republican Senate vote in state  $n$ . If we regress reported vote on actual vote, both measured at the state level, the slope of the regression line ought to be  $45^\circ$  (hence  $b_1 = 1$ ) and there should be no constant over- or underreporting (hence the intercept,  $b_0$ , should be 0). All this is presented in equation 1.

$$RV_n = b_0 + b_1 AV_n + e \quad (1)$$

$$\text{If } E(RV_n) = AV_n, \text{ then } E(b_0) = 0 \text{ and } E(b_1) = 1.0$$

As Wright shows, the observed relationship deviates substantially from expectations—some degree of overreporting for the winner is going on. The distribution is tilted, with reported vote increasing faster than actual vote (Wright, Figure 1). The estimated slope and intercept, obtained through OLS regression, are reported in my Table 1 (compare to Wright, Table 3). Is there overreporting? When the actual vote is greater than 80%, the predicted reported vote is greater than 100%. More NES/SES respondents are saying they voted for the

TABLE 1  
State-Level Evidence of Overreport of Senate Votes  
(aggregate data; standard error in parentheses)

Coefficient	Senate Vote	Presidential Vote
Constant $b_0$	-.238 (.052)	.091 (.119)
Actual Vote $b_1$	1.516 (.108)	.892 (.217)
N	32	50
R <sup>2</sup>	.869	.245
Standard Error	.081	.085

Source: 1988 NES/SES, 1988 official election returns.

Note: Both equations were estimated via ordinary least squares regression. The dependent variable in both cases is the reported vote for the Republican Senate candidate, aggregated to the state level. The independent variable is the actual Republican vote percentage, taken from the *Congressional Quarterly*. The substantive results are the same when the presidential equation is estimated only for states holding Senate elections. The coefficients and valid n in column 1 differ slightly from Wright (1990) because I eliminated Nevada from the analysis (due to a data error reported in the second release of the NES/SES).

winner than actually did. Curiously, the effect is not evident for presidential vote (see Table 1, column 2). The estimated values of  $b_0$  and  $b_1$  are statistically indistinguishable from 0 to 1. State-to-state variation in NES/SES presidential election vote percentages closely matches actual percentages.

The pattern of results at the individual level looks the same. Now I am regressing the vote report of the  $i$ th respondent ( $1 = \text{Republican}$ ) in the  $n$ th state against the actual vote in the  $n$ th state.<sup>2</sup>

$$RV_{in} = b_0 + b_1 AV_n + e \quad (2)$$

Individual respondents report voting for the winner 1.38 times as often as they ought to, if vote reports were unbiased (compare Table 2, column 1, to Wright, Table 4, column 1). The effect is not limited to the NES/SES. The National Elections Study's postelection survey (NES) data have the same bias (Table 2, column 2). Clearly, something is amiss in how Senate votes are being reported.

After examining a number of possible causes of overreports, Wright settles on a combination, the spiral of silence and time: "the results are consistent with the time effects hypothesis in which there is a spiral, if not of silence, then away from admitting losing votes for Senate candidates" (1990, 552). He reaches this conclusion after

TABLE 2  
Individual-Level Evidence of Overreport of Senate Votes  
(standard error in parentheses)

Coefficient	Senate Vote		Presidential Vote		House Vote	
	NES/SES	NES	NES/SES	NES	NES/SES	NES
Constant $b_0$	-.156 (.035)	-.264 (.067)	.063 (.114)	.062 (.163)	-.011 (.001)	-.056 (.020)
Actual Vote $b_1$	1.377 (.075)	1.506 (.145)	.937 (.204)	.863 (.304)	1.007 (.003)	1.126 (.041)
N	1224	842	1984	1209	1702	1054
Standard Error	1.03	.465	.491	.498	.045	.114

Source: 1988 NES, NES/SES.

Note: All equations were estimated via weighted least squares. The dependent variable in both cases is the reported Senate, presidential, or House vote (where 0 = Democrat and 1 = Republican). The independent variable is the actual Republican vote percentage, taken from the *Congressional Quarterly*. This variable is identical to the one used in the aggregate analysis (election results are attached to the respondent record). The coefficients and valid  $n$  in column 1 differ slightly from those in Wright (1990) because I eliminated Nevada from the analysis (due to a data error reported in the second release of the NES/SES).

searching for the combination of variables that reduces  $b_0$  and  $b_1$  to their desired values. He finds that when two variables are added they account for the gap between the date of the election and the date of the interview ( $Days_{in}$  and  $Days_{in} * AV_n$ , shown in equation 3).<sup>3</sup>

$$RV_{in} = b_0 + b_1 AV_n + b_2 Days_{in} + b_3 Days_{in} * AV_n + e \quad (3)$$

Voters err in their reports of their Senate vote as the gap between the election and the interview date increases; the size of the error increases in proportion to the size of the winning margin (hence the interaction term in equation 3).<sup>4</sup>

As Wright recognizes, it is one thing to point to a methodological flaw and cry "Error, error!" It is another to show that the error is something to be concerned about. Is this a worrisome situation? Wright thinks it is—he claims that overreporting biases coefficients in other equations. After comparing regression equations that predict partisan Senate choice using NES/SES data to those using ABC and CBS exit polls, he writes, "The conclusion from our estimation of the partisan choice model is that coattail effects are much larger in the exit polls; local factors, incumbency, and candidate spending are substantially larger in the NES/SES data. This pattern shows that the systematic bias in reported vote produces underestimates of the effects of

national forces and overestimates of the impact of incumbency/candidate variables in Senate elections” (1990, 557).

The relative importance of national and incumbency variables is at the center of academic debate about congressional elections. If Wright is correct, his is surely a damning critique of the survey method used by the National Election Study. In fact, Wright closes with a call for reorientation of voting surveys, asking them to emulate exit polls: “the ideal would be huge election-day polls that could tap voter attitudes and reports of behavior before they are contaminated” (p. 560). Is this harsh conclusion warranted?

### **An Underspecified Theory: What About Presidential and House Elections?**

I think it is not. My initial question is whether overreporting occurs in other races (Wright looks at the presidency; I add the House). The answer is that it does not. Estimates of presidential votes do not exhibit an overreporting problem (Table 1, column 2; Table 2, columns 3 and 4). The values for  $b_0$  and  $b_1$  are within one-half of a standard deviation of their predicted values. Strangely, individual-level House votes appear to suffer from minor overreports in the postelection NES ( $b_1 = 1.126$ ) but not in the NES/SES (compare column 5 to 6). The degree of overreporting, though, is nowhere as severe as for the Senate votes.<sup>5</sup>

This pattern of evidence does not fit at all with Wright’s suggested cause. Wright explains the time effect this way: “in the days that pass between the election and the interview, the respondent is exposed to postelection coverage and to social interactions; these provide information and pressures that can affect recall and the respondent’s overall image of the election” (550). If information about the election outcome is influencing respondents to misreport the vote, surely the presidential results should be more prone to error than the Senate, not less so. The amount of news coverage about the presidential outcome is intense, far outweighing Senate coverage. If “abandonment of the loser” is occurring, I certainly would expect Dukakis voters to join in the fun. Any suggested cause would have to explain why only Senate results are affected. I do not think Wright’s account does this (though he can explain lack of House misreporting—postelection coverage is slight).

We are left with a puzzle. Clearly, there is serious overreporting of Senate votes in the 1988 NES and NES/SES. The inclusion of a date of interview variable brings the coefficients in line with

expectations. Time effects should operate in other elections but do not. Respondents are somehow immune to time effects when reporting their presidential and House votes. Wright's theory needs to explain why presidential and House voters do not overreport, as well as why Senate voters do.

### **A Misspecified Model: Overreporting for "Winners" or Republicans?**

This discussion is rendered irrelevant by a fundamental problem—the model Wright uses to isolate the overreporting problem is misspecified. In equation 3 (compare to Wright 1990, Table 4, column 5),  $RV_{in}$  is a dummy variable that indicates whether the respondent voted Republican (1) or Democrat (0). Yet Wright's whole theory revolves around overreporting voting for the winner. Is Wright's specification simply an easier way to find the source of the bias?

Unfortunately, it is not. Using Republican vote instead of winner's vote undermines the whole analysis. According to equation 3, all else being equal, respondents are more likely to vote Republican over time.<sup>6</sup> This can be seen by taking the first derivative of equation 3 with respect to date of interview ( $Days_{in}$ ):

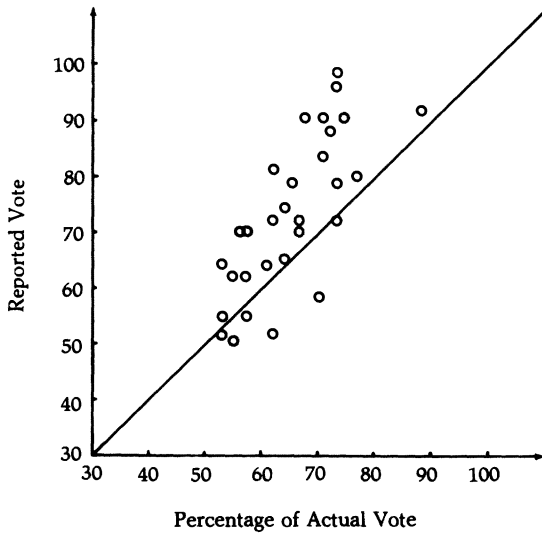
$$\frac{\delta RV_{in}}{\delta Days_{in}} = b_2 + b_3 AV_n \quad (4)$$

The second term in equation 4 is exactly right: we expect the probability of overreporting in favor of a Republican to increase as the vote total of a Republican increases. However,  $b_2$  is exactly wrong: it indicates that, over time, there is a constant increase (of  $b_2$ ) in the probability of reporting a Republican vote, independent of who actually won the election. Even if a Democrat wins, the equation predicts an increased probability of voting Republican over time. This can't be right. Wright's use of a dummy variable representing the party vote is inappropriate.

Instead, as Wright's theory indicates, the winner's vote percentage is the variable of interest. At the aggregate level, there is still evidence of bias: in 1988, the actual mean vote for the winner was 60.9%, but the mean vote reported in the NES/SES was 66% and the mean in the NES data was 65.6%.

Figure 1 is a plot of the actual percentage of votes received by winning Senate candidates against the mean percentage of votes

FIGURE 1  
 Comparison of Actual Senate Winner's Vote  
 with Reported Winner's Vote  
 (state level, Senate Election Study data)



for the winning Senate candidate reported in the NES/SES (aggregated by state); the expected 45° line is added for reference. The distribution is tilted, with reported vote increasing faster than actual vote.<sup>7</sup>

However, the amount of overreporting is not large enough to be statistically discernable. The estimated values for  $b_0$  and  $b_1$  are shown in Table 3.<sup>8</sup> The values are signed correctly, and neither is discernably different from its expected value. When the analysis is performed at the individual level, with either the NES/SES or the NES data, no bias is evident (see columns 2 and 3 of Table 3). When the model is properly specified, with vote for the winner used instead of vote for Republicans, there is no evidence of bias in vote reports.<sup>9</sup>



TABLE 3  
Overreport of Senate Winner's Votes?  
(standard error in parentheses)

Coefficient	Winner, Aggregated Data	Winner, Individual Level	
	(NES/SES)	NES/SES	NES
Constant $b_0$	-.105 (.111)	-.005 (.084)	-.112 (.153)
Actual Vote $b_1$	1.28 (.179)	1.09 (.130)	1.29 (.253)
N	32	1224	842
R <sup>2</sup>	.617	.201	.067
Standard Error	.08	1.01	1.00

Source: 1988 NES, NES/SES, Official election returns.

Note: The aggregate-level equation was estimated with ordinary least squares, the individual level equations with weighted least squares. The dependent variable is the reported vote for the Senate winner (at the individual level, 1 = voted for the winner, 0 = voted for the loser; in the first equation, these reports are aggregated to the state level). The independent variable is the actual winner's vote percentage, taken from the *Congressional Quarterly*. These variables are simple transformations of the variables from Table 2.

### Demonstrating Bias: When Can We Compare Exit Polls and Academic Surveys?

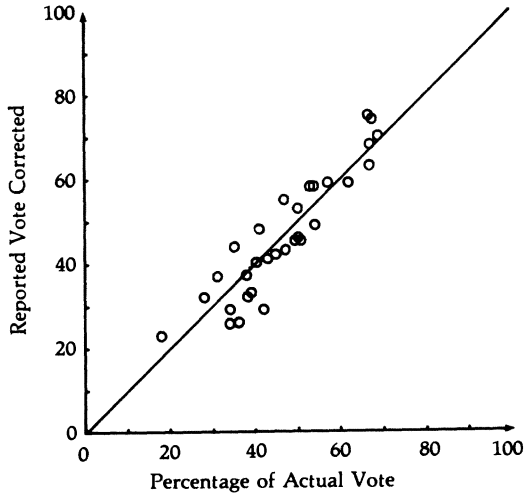
I have shown that Wright's model is misspecified. I also take issue with another part of the article: how he demonstrates the effects of overreporting. Comparisons of structural equations estimated with two very different datasets—an academic survey and a network exit poll—are not as unproblematic as Wright indicates. Furthermore, if the NES/SES is flawed, a simple correction is available. I will show the correction first.

Wright's proposed representation of the bias problem is shown in equation 3. If we are able to predict model bias accurately, a simple correction can be applied when descriptive statistics are desired. Moving AV to the left-hand side results in equation 5. Since RV ought to be identical to AV if the correction works, I substitute  $\hat{R}\hat{V}$  for AV.

$$\hat{R}\hat{V} = \frac{RV}{b_1 + b_3 * \text{Days}} - \frac{b_0}{b_1 + b_3 * \text{Days}} - \frac{b_2 * \text{Days}}{b_1 + b_3 * \text{Days}} \quad (5)$$

The coefficients reported by Wright can be substituted in this equation. Figure 2 is a reproduction of Wright's Figure 1, this time with

FIGURE 2  
Fixing the Bias in Vote Reports



corrected Senate vote (aggregated to the state level) plotted against actual Senate vote. The cloud of points reveals no evident bias: they are randomly distributed about the expected 45° line. The new variable, when reexpressed as the percentage voting for the winner, has a mean closer to the true value.<sup>10</sup> The correction works.

When multivariate analysis is used, the bias should be treated as an omitted variables problem; date of interview should be an additional regressor. If Wright is correct, the coefficients from a vote equation including the new variables ought to look like those obtained from exit poll data.<sup>11</sup> Instead, the coefficients are virtually identical to Wright's (see Table 4). Adding regressors to account for the bias makes no difference in the results, nor does it improve the fit of the model to the data. The relative impact of presidential preference and candidate factors remains the same. The overreporting bias that Wright discovered does not alter our substantive conclusions about Senate voting.

However, the discussion so far sidesteps the larger issue—how to compare exit polls with academic surveys. Wright claims that exit poll data are correct and that NES/SES data are wrong because exit polls are conducted on election day and are presumably free of time-induced bias. Yet, as my analysis in Table 4 shows, some other difference between the ABC and CBS exit polls and the NES/SES besides

TABLE 4  
Vote Equations: Without and With Bias Variables  
(standard error in parentheses)

Variables	Wright Model	Adding Regressors
Party Identification	.341 (.061)	.340 (.061)
Presidential Vote	.789 (.121)	.791 (.121)
Log of Democratic Spending	-.371 (.053)	-.364 (.053)
Log of Republican Spending	.326 (.045)	.316 (.047)
Democratic Incumbent	-.377 (.123)	-.381 (.124)
Republican Incumbent	-.079 (.133)	-.088 (.134)
Date	—	-.041 (.026)
Interaction of Date and Actual Vote	—	.091 (.055)
Constant	.108 (.319)	.121 (.312)
Maddala R <sup>2</sup>	.311	.313
Log-likelihood	-788.41	-788.41
Percentage correctly predicted	72.2	72.6
N	1138	1138

Source: 1988 NES/SES.

Note: Because the dependent variable—partisan vote choice (0=Democrat and 1=Republican)—is a dichotomy, these are all probit estimates. Party identification is a three-point scale, with strong Democrat through Independent-leaning-Democrat coded into 1, pure Independents coded into 2, and Independent-leaning-Republican to strong Republican coded 3. Presidential vote is a dichotomy (0=Dukakis, 1=Bush). Democratic and Republican spending are logged per capita measures:  $\ln$  ((Campaign spending)/state population). Democratic and Republican incumbency are dummies. Asymptotic standard errors are in parentheses. These estimates differ in minor ways from those reported by Wright, due to a different coding of partisanship and the deletion of Nevada (his  $n = 1258$ ).

date of interview must be causing the differences in coefficients reported by Wright. I cannot believe that overreporting is the culprit until a series of alternative hypotheses are eliminated. What is the demographic makeup of the CBS and ABC exit poll sample—are the exit poll and NES/SES samples comparable? If the effect of overreporting is to bias regression coefficients, then regression models for races in which overreporting is not a problem (e.g. the presidential race) should look the same across these studies. Is this true? (Wright does not report comparable vote analyses for the House and presidency.) Most important, why does time affect the Senate reports and not those for the House or the presidential vote.

What we are seeing here, I think, is the danger in comparing two very different surveys. Are there good reasons to suppose that an exit poll is a better measure of voter opinion than the NES/SES poll? Plissner and Mitofsky (1982) liken participation in an exit poll to vot-

ing a second time. In an important way, they are correct: the exit poll becomes part of the single act of voting. Every respondent is a voter. The exit poll minimizes the effect of history, be it contamination from postelection coverage, social interactions, rationalizations, bandwagons (since the winner is not yet known), or simple forgetting. All this improves the reliability of exit poll data.

There are reasons to suspect exit polls. Some relate to administration. Exit polls attempt to convert initial refusals, but do not attempt anything like the multiple callbacks that a well-run academic poll will employ (low refusal conversion rates can cause serious bias in political measures, particularly those relating to participation; see Brehm, 1989, 1990). News deadlines mean there is little time to adjust to problems that might arise during interviewing. Other concerns relate to the sample. Sampling within strata is done either in proportion to the total number of votes cast in some base year or in proportion to current voter registration (Levy 1983; Mitofsky and Waksberg 1989). Large variations in turnout relative to either figure result in unequal selection probabilities. News organizations differ in their treatment of nonresponse: some continue the skip pattern while others take the next available voter, thus converting an every  $k$ th voter skip pattern to an every  $k + 1$  pattern (Levy 1983). Finally, Mitofsky and Waksberg notice a slight but persistent bias in exit polls toward reporting a vote for the Democratic candidate (1989, 16–17).

However, none of these clues are sufficient evidence to convict the exit polls. None of these explain the obvious differences in structural equation results reported by Wright. Exit polls provide highly reliable measures of election results. Response rates (around 75%) are similar to academic surveys and are higher than many telephone polls such as the NES/SES. Wright is joined by Levy (1983) when he claims that exit polls measure voters' attitudes, opinions, and actual vote choice better than academic polls such as the National Election Study do. The evidence presented by Wright is convincing on this score.

Nonetheless, academic surveys should not be replaced by huge election day polls. The central purposes of an exit poll and an academic poll like the NES/SES diverge. What the exit poll gains in speed and sample size it loses in breadth and sample reliability. An exit poll provides a snapshot of voter opinions as they leave the polling booth. It allows news organizations, and the public, to learn the outcome of the election quickly. But it provides limited information about the reasons behind the vote, and virtually no information about other politically important topics. The NES/SES survey instrument ran, on average, 35 minutes. It asked questions about senators not up for

reelection, contacts with senators and House members, political interest, media use, reactions to prominent political figures, and much more. The sample is carefully drawn, and great effort is expended to complete each interview.<sup>12</sup>

Where the exit poll questionnaire is driven by the need to construct a compelling account on television news and in next morning's byline, the academic poll is driven by questions and issues that concern the scholarly community. At a 1985 meeting of the Midwest Political Science Association, the rolling cross-section portion of the 1984 NES was criticized by media pollsters because it did not ask the "right" questions about the 1984 campaign. Media polls were criticized by academics because they asked only the "right" questions and little else. Neither situation is ideal.

### Future Directions

I can think of two possible ways to fit Wright's account with the empirical results. One cause of overreports could be the survey instrument. Over time, the name of the Senate winner receives prominent play in the media while the loser leaves the public stage. Over time, then, information about the winner, particularly the name, is readily accessible, while the loser's name becomes less and less available (and in particular its availability declines relative to the availability of the winner's name). Any information provided within a survey that stimulates recall of the winner's name might result in misreports. The NES provides just this kind of stimulus, and only for House and Senate contests: for these races, the candidates' names are provided. In our attempts to simulate the polling booth, we may be unintentionally biasing vote reports. This explanation accounts for the appearance of overreporting only for the Senate and House. The weaker effect for House voting is a result of the heavier coverage given to the Senate elections. Time comes into play because the relative availability of the winner's name increases as time passes.

A second explanation relies on a model of the process of misreporting.<sup>13</sup> Three variables determine the scope of misreporting: confusion, exaggeration, and the election result. Voters may be confused about whom they voted for. Their confusion would vary with their interest in the election and the cognitive effort they put into the vote choice. Voters also exaggerate in favor of the winner due to social pressures (desire to be with the winner, conform to the majority's choice, etc.). Finally, the possibility of misreports is affected by the probability

that a respondent could have voted for the loser. This is determined, of course, by the actual election result.

Assume that misreporting  $m_i$ , for the  $i$ th individual, is affected only by confusion and exaggeration. The probability of misreports is equal to the joint probability of confusion and exaggeration:

$$\Pr(m_i) = \Pr(\text{Conf}) + \Pr(\text{Exag}) - \Pr(\text{Conf}) * \Pr(\text{Exag})$$

Next, assume that overreporting operates in one direction, toward the winner (you can misreport only if you voted for the loser). Therefore, the probability statement above has to be multiplied by the chance you could misreport at all (i.e., by the loser's vote percentage):

$$\Pr(m_i) = \Pr(\text{Loser}) * (\Pr(\text{Conf}) + \Pr(\text{Exag}) - \Pr(\text{Conf}) * \Pr(\text{Exag}))$$

Wright's account includes only the first two factors: confusion (forgetting the vote) and exaggeration (as a result of time, media coverage, and associated social pressures). The probability of voting for the loser is left out yet plays a critical role. In the extreme case, voters cannot misreport voting for the winner when a candidate runs unopposed. More practically, the distribution of winning percentages by state assumes a relatively normal shape for the presidency and the Senate, since there are not many blowouts. The House distribution is trimodal—a fair number of close races and a large number of Democratic and Republican blowouts. For blowouts, the likelihood that you could have voted for the loser is low; therefore, the overall likelihood of misreporting is also low.

I combine this observation with conjectural mean values on the other variables in Table 5. For illustrative purposes, I have used three values in the cell entries, low (.25), medium (.50) and high (.75). These values are arbitrary, but the ranking of the institutions is not. A good argument can be made for each cell value (the logic for cell entries is contained in the note to the Table).

Though purely speculative, the analysis targets Senate voting as most likely to suffer from overreports. On many dimensions, the Senate falls between the House and the presidency—in citizen interest, campaign intensity, media coverage, and institutional visibility (this is one reason why studying Senate elections holds so much promise for improving our understanding of electoral behavior). Ironically, middling values on the dimensions of interest here—moderately high citizen interest (resulting in less confusion), moderately high media coverage (increasing the potential for exaggeration), and relatively competitive races—results in the highest probability of overreporting bias. Less to its credit, the analysis indicates that the presidential race

TABLE 5  
A Speculative Model of the Bias  
in Vote Reports

Election	Probability of			
	Confusion	Exaggeration	Voting for Loser	Overreport
Presidential	.25	.75	.50	.59
Senate	.50	.50	.50	.625
House	.75	.25	.25	.30

*Note:* For illustration, I have used three values, low (.25), medium (.50) and high (.75) in the cell entries.

For the presidential race, I assume that voters were most interested in this race. Therefore I suppose that confusion is low. I assume that exaggeration is high because media coverage is heaviest. Finally, the mean percentage vote won by Dukakis was 45%, with few blowouts—loser vote is scored medium.

Voters are generally less interested in Senate races than in the presidential race but more than in House contests. I score likelihood of confusion medium. Media coverage of the Senate results also falls in between results for the presidency and for the House (Westlye 1987); I score exaggeration medium. Finally, the distribution of Senate loser vote percentage centers on 39%, with few blowouts—I score loser vote medium.

For the House, I score confusion high, since voter interest and involvement in these races has traditionally been low. I score exaggeration low, since there is little coverage of the outcome, certainly nothing rivaling coverage of the presidential or Senate results. Finally, I score loser vote low, since there are many lopsided House elections.

should suffer from overreports more than the House, a result not supported by the empirical findings (though individual-level results on this point are inconclusive). Also, the predicted probability of overreporting the presidential vote is not much smaller than that of overreporting the Senate vote, whereas the observed difference is large. Obviously I could jiggle the numbers to make the results come out cleaner, but that would obscure rather than clarify. Suffice it to say, the next step is to insert real values into this Table.

This model of overreporting implicates both systemic and individual-level influences. It suggests where researchers might profitably look for causes of misreporting. Individual-level variables such as education and political interest determine the likelihood of confusion; attentiveness to campaign coverage and group affiliations will effect exaggeration. At the same time, the content of postelection coverage and postelection discussion in a community will determine, in part, the degree of exaggeration. To measure voting for the loser, we could even disaggregate Senate and presidential outcomes to smaller areas—what was the margin in county A or congressional district B? Most of

the measures suggested here are already available: levels of political interest, education, media usage, and group affiliation can be obtained from survey data; election results are publicly available. The missing component is the content of postelection coverage. The model and Table 5 suggest a more complex model of misreporting could be worth pursuing.

### Conclusion

The 1988 Senate study is an invaluable resource for congressional and electoral scholars. At present the response rates are unacceptably low; the NES needs to expend considerably more effort in converting initial refusals. The potentially most damaging problem with the study is the bias in vote reports discovered by Gerald Wright. I disagree with Wright both on the cause and effect of overreports. When the question is framed correctly—overreporting votes for the winner—the bias is much less severe than he supposes (failing to meet conventional statistical significance levels). When the date of interview is included in a regression equation predicting vote, the other coefficients are unchanged.

However, there is no doubting that Wright has identified an area for further research. Even if the degree of overreporting in favor of the winner is statistically discernable, a 5% overestimate is cause for concern. Much more work needs to be done on the reasons for the overreport of Senate outcomes. Scholars need to explore the costs and benefits of our current survey strategies, focussing especially on the reason why regression models look so different across exit polls and academic polls. Wright has done the discipline a service in this regard. I have suggested two possible reasons for the overreports that implicate the Senate alone. Both need further study to be proven or disproven. I do not agree with Wright, though, when he recommends large, election-day polls as a solution to the problem. For scholarly concerns, academic polls such as the NES/SES remain the best vehicle. Given current funding constraints, I see no reason to take on the responsibility that is being handled rather well by networks and newspapers.

*Paul Gronke is an Assistant Professor of Political Science, Duke University, Durham, North Carolina 27706.*

### NOTES

The data utilized in this paper were made available by the Inter-university Consortium for Political and Social Research. Neither the collector of the original data



nor the consortium bears any responsibility for the analyses or interpretations presented here. I would like to thank John Brehm and Charles Stewart particularly for help and guidance. I am indebted to Graham Kalton, Donald Kinder, Steven Rosenstone, Mike Traugott, Santa Traugott, and Gerald Wright for advice and assistance. They are of course absolved of responsibility as well.

1. This section is a summary of Wright (1990) with some additional analysis of my own.

2. Because of heteroskedasticity due to a dichotomous dependent variable, the individual-level regressions are estimated with weighted least squares (Hanushek and Jackson 1977, 180–84).

3.  $Days_{in}$  = how long after the election the interview took place (date of interview–November 8).

4. The coefficient estimates from this equation can be found in Wright, Table 3, column 5.

5. Sample sizes at the congressional district level are too small to reliably estimate House results.

6. I am indebted to Charles Stewart for pointing this out.

7. Graphically, this distribution is identical to Wright's Figure 1, with the points below and to the left of the (50%,50%) point rotated about that point.

8. RV is the estimated vote for the winner, derived from reported votes in the NES/SES. AV is the actual vote for Senate winners, as reported in *Congressional Quarterly*.

9. Notice that restating the model in terms of winners explains the lack of overreporting in presidential elections. In 1988, voting Republican and voting for the winner are identical.

10. The corrected estimate for the winner's vote percentage from NES/SES data is 61.5%.

11. Table 5 in Wright compares vote models estimated using NES/SES data and exit poll data. Part of that table is reproduced here (all variables are described in the notes to Table 4). The coefficient on presidential vote is much larger in the exit poll equation and the coefficients on candidate-based measures (incumbency and spending) are much smaller.

	NES/SES	ABC News
Party identification	.509	.528
Presidential Vote	.742	1.20
Log Democratic	-.371	-.271
Log Republican	.301	.256
Democratic Incumbent	-.368	-.117
Republican Incumbent	.015	-.011
Constant	-.226	-.590
R <sup>2</sup>	.475	.529
N	1258	63,855

12. Admittedly, the NES/SES fails woefully at getting complete interviews, however.

13. John Brehm gracefully clarified my thoughts in these paragraphs.

## REFERENCES

- Brehm, John. 1990. "Opinion Surveys and Political Representation." Unpublished dissertation, Department of Political Science, University of Michigan.
- Brehm, John. 1989. "How Survey Nonresponse Damages Political Analysis." Presented at the annual meeting of the American Political Science Association, Atlanta, GA.
- Hanushek, Eric A., and John E. Jackson. 1977. *Statistical Methods for Social Scientists*. New York: The Academic Press.
- Levy, M. R. 1983. "The Methodology and Performance of Election Day Polls." *Public Opinion Quarterly* 47:54-67.
- Mitofsky, Warren, and Joseph Waksberg. 1989. "CBS Models for Election Night Estimation." Presented at the American Sociological Association Conference, San Diego, CA.
- Plissner, M., and Warren Mitofsky. 1982. "Voting Twice on Election Day." *Public Opinion* August/September:14-16.
- Wright, Gerald C. 1989. "Misreports of Vote Choice in the 1988 NES Senate Election Study." Presented at the conference on "Electing the Senate," Houston, TX.
- Wright, Gerald C. 1990. "Misreports of Vote Choice in the 1988 NES Senate Election Study." *Legislative Studies Quarterly* 15:543-64.