NES WORKING PAPER

IDENTIFYING BIAS IN VOTING MODELS

NOVEMBER 4, 1991

RUSSELL MAYER

Misreporting by survey respondents is a problem for gathering accurate and reliable data in many research contexts.[1] The National Election Study (NES), as well as other surveys designed to measure the political attitudes and behaviors of the American electorate, have documented problems with misreporting, especially with regard to voting (see Anderson and Silver, 1986; Jennings, 1990; Silver, Anderson, and Abramson, 1986). NES estimates of voting turnout rates have been consistently higher than rates reported in aggregate voting returns (see Table 1). Part of this discrepancy can be accounted for by components of the survey design and execution: known sampling biases, survey non-response, and stimulus effects (Clausen, 1969; Traugott and Katosh, 1979). However, NES estimates of turnout remain substantially higher than official reports; Clausen (1969) reports that approximately 8% of this overestimate is due to misreporting. In part, it seems that misreporting by respondents, i.e. respondents claiming they voted when they did not, inflates statistics on percentage voting as measured by self reported vote.[2]

These inflated statistics for voting participation, troubling in and of themselves, indicate an even more disturbing problem for researchers interested in studying the causes of voter participation. The National Election Study's self reported vote variable, a measure of voting commonly used in analyses of the cause of voting, incorporates not only actual vote, but also misreport. Under certain conditions, this contamination of the self reported voting variable leads to biased coefficients in models of voting participation. If people engage in misreporting randomly, then using self reported vote to examine the causes of voting presents no problems in determining point estimates; the relationships between voting and its causes are not biased in any by the phenomenon of misreporting, and these random occurrences cancel out.[3]

However if there are systematic causes of misreporting, then the coefficient estimates from self reported voting models for factors related to both voting and misreporting may be biased. When voting is the dependent variable, misreporting is relegated to the disturbance term. Standard statistical treatments require the assumption that those factors which make up this disturbance term are random occurrences, not systematically related to any variables important to the voting model. However, if this assumption of independence of the disturbance term and the exogenous variables is violated, the coefficient estimates on these variables will be biased. In the bi-variate case, the independent variable which is positively related to both misreporting and voting will be inflated. There will be a systematic relationship between the independent variable and the disturbance which will add to the size of the coefficient relating this variable to voting. In the multi-variate case, the direction of bias is not as easily determined, but bias due to correlation between measurement error in voting (e.g. misreporting) and factors related to voting still exists.

One response to this problem would be to use validated vote data to measure voting in place of self report. Since 1964, NES has conducted vote validation studies in which the self reported voting and registration statuses of respondents are compared with actual voting and registration records. Yet, using validated vote in models explaining the causes of voting may result in other problems. Sanchez (n.d.) contends that validated vote introduces biases of its own, due to NES's failure to locate the records of a substantial number of respondents. She suggests that the likelihood of locating a respondent's records is correlated with region, size of place, and gender (Sanchez, n.d.). If certain sub-populations are systematically excluded from the validated vote variable, and these sub-populations turnout at rates different from the rest of the population, then using validated vote may introduce as many biases as it corrects.

Clearly, both self reported and validated measures of vote are flawed in certain respects. Self reported vote inflates turnout rate and counts misreporters as voters, while validated vote may systematically exclude certain populations from the vote variable. Past attempts to identify misreporting biases have generally ignored the potential biases introduced by the vote validation process. In an attempt to compensate for this potential problem, my approach to identifying misreporting bias includes a correction for censored sample suggested by Achen (1986). I transform the validated vote model by including a variable measuring the probability of selection into the validated vote sample in order to achieve unbiased results. Only then do I compare this correct validated vote model to the model using self reported vote as the dependent variable. The differences in the coefficients on these two models show the degree of bias introduced by misreporting. By way of comparison, I include in my analysis several factors which theoretically should be related only to voting and therefore should have similar coefficients across the two models. I concentrate my efforts here on identifying NES misreporting bias in the most recent presidential election year 1988. My primary goal is to develop a methodology which might be extended to other years in future analyses.

<u>The Censored Sample</u>

The sample that remains once vote validation has been applied to NES data is systematically different from the original sample.[4] If there are factors which affect the process of selection into the sample of validated voters which also affect the likelihood of voting, then coefficients on these variables for the vote equation will be biased. For example, Southern African-Americans are less likely to be validated, due to generally poorer record keeping in the South. Thus, the population of African-Americans in the validated or censored sample is disproportionately composed of non-Southern African-Americans. If African-Americans included in the sample vote at rates higher than their Southern counterparts, the relationship between this racial characteristic and voting will be overestimated due to sample censoring.

Adjusting for this sample censoring involves modeling a selection equation which predicts the likelihood of each case being selected into the sample. Such an adjustment is possible because observations on which cases have been included and excluded by the validation process are available. In addition, there exist data on independent variables related to probability of selection, at least one of which is unrelated to the probability of voting, so that the system of equations is identified. Thus, the correction for sample censoring involves accounting for voting by using two separate equations which model the process of selection into the sample and the process of voting.

The selection model involves a dichotomous dependent variable (selected or not selected), and is estimated using logit. The predicted values from this equation are incorporated as a independent variable in the voting equation, which is also dichotomous in nature. This variable controls for the probability of being selected into the censored sample when estimating the probability of voting.

Sanchez (n.d.) provides several insights into potential factors in the selection process that bias the validated vote measure. As mentioned, because of differences in quality of record keeping, respondents in the South are more likely to be excluded from the validated vote sample. Size of place also affects probability of inclusion in validated vote. People from large cities, where the number of records the county election office handles is large, may be more likely to have their voting records misplaced or mishandled than people from smaller localities. Thus, size of place affects whether or not NES was able to validate a person's vote. In addition, there exists an independent evaluation by the NES record checker of the probability that they found all the records in the election office in which they performed vote validation. This is a subjective measure which may have been used by record checkers in order to disguise their own incompetence instead of to assess the quality of access to records accurately. However, this measure still should be highly related to a respondent's chance of exclusion from the sample, whether this is due to poor record access or a poor search.[5]

Hence, the selection model includes a dummy variable for Southern respondents, and ordinal variables on size of place and access to records as the independent variables. The dependent variable is a dichotomy for inclusion in the validated vote sample. There are 1,775 cases for which pre- and post-election interviews were conducted; for all of these interviews a measure of self reported vote was obtained. Of these 1,775 cases, 1,652 were included in the validated vote measure, while 123 were excluded due to failure to validate these respondents (see Table 2 for a breakdown of the self reported and validated vote variables). The logit estimates for the selection equation are given in Table 3.

All of the coefficients on the causes of inclusion in the sample of validated vote are in the appropriate direction and all are statistically distinguishable from zero effects with at least 95% confidence. Being from the South makes a respondent less likely to be included in the validated vote sample, while being from a smaller geographic unit, for example, suburbs versus an inner city, makes a respondent more likely to be included in the sample. In addition, the validation checker's rating of his or her access to office records is positively related to a respondent's chance of selection into the sample.

Table 4 shows that the magnitudes of these effects are fairly small. Holding other variables constant at their means the entire effect of a respondent being from the South translates into only a 2.2% decrease in the probability of inclusion in the sample. Similarly, substantial changes in size of place and access to office records, from one standard deviation below the mean to one standard deviation above the mean, result in small changes in the probability of inclusion, 1.9% and 2.4% respectively. However, these data generally confirm the notion that selection into the validated sample has systematic components which can be identified.

Most importantly, this selection equation provides a predicted probability of inclusion in the sample for every case in the validated vote sample. The inclusion of these predicted values in a model of voting that uses validated vote as the dependent variable allows one to control for the potential biases introduced by the way in which this censored sample was constructed. In essence, including probability of selection in a voting model utilizing validated voting controls those biases introduced by the validation process itself.

Comparing Models of Voting

Traditional voting models use demographic and political variables (e.g., age, income, education, race, political efficacy, and political interest) to explain voting (Wolfinger and Rosenstone, 1980; Katosh and Traugott, 1981; Abramson, Aldrich, and Rohde, 1987). However, several of these factors may also be related to misreporting. For example, one reason why education and income make a person more likely to vote is because voting is more a social norm for highly educated, upper income people. However, this social pressure to vote may also cause well education, wealthier individuals to lie about having voted. Thus, in models using self reported vote as a dependent variable, the coefficients for these variables may be biased estimates of the actual relationship with voting. They will overestimate these relationships, since they systematically account for misreporting as well as voting.

However, there are other variables which while related to voting, should not be related to misreporting at all. Specifically, structural variables measuring the relative ease of registration, a prerequisite for voting in nearly every state, should affect the likelihood of voting without affecting the likelihood of misreporting. Wolfinger and Rosenstone (1980) report that four voter registration variables significantly affect voting. These variables are closing date of registration, existence of alternate registration locations, registration by mail, and extended election office hours. These variables should all contribute to the ease (or difficulty) of registration and therefore the likelihood of voting, without providing any apparent incentive for respondents to misreport. Detailed descriptions of the political, demographic, and structural variables used in the models are located in Appendix 2.

Thus, those political and demographic factors correlated with social or moral pressure to vote (education, income, political interest, and political efficacy) are theoretically linked to both voting and misreporting. In a model using self report as the measure of voting, the coefficients on these variables will be biased. Comparing the coefficients for these variables in the model using self reported vote to those in the model using validated vote corrected for the censoring bias should reveal the true degree of bias in these coefficients. In contrast, the structural variables should display nearly identical coefficients in both models, because they should not be subject to any of the bias due to misreporting associated with the self reported vote measure.

Analysis

The dichotomous nature of the dependent variables requires the constraint of predicted Y to the zero-one interval and suggests a non-linear functional relationship. Logit was used to meet these specifications and the resulting analysis is shown in Table 5. The coefficients produced by logit cannot be directly interpreted in terms of the dependent variable. These coefficients represent the expected change in the log of the odds ratio for a unit change in X, not a change in the probability of the dependent variable [P(Y)].

In order to understand the effects of the independent variables on the dependent variables more clearly, I transformed these logit coefficients into equivalent expressions given in terms of P(Y) (Table 6). The relationships between the Xs and Ys are non-linear in the parameters, meaning that the effect of a unit change in X on Y varies depending on the levels of all the Xs. Thus, in order to calculate a change in the probability of Y for a change in any one X, one must specify the level of that particular X and the levels of all other Xs. In models with only a few independent variables this can be expressed by showing several logit curves that represent changes in Y given the X of interest for several values of other Xs. However, since voting models contain so many X variables, I present the expected value of P(Y) for each X one standard deviation above and below its mean, given all other Xs take their mean values.

In general, the estimates for the political and demographic variables seem to support my hypotheses. Comparing the two models the coefficients on the political and demographic variables generally behave as expected. Age seems to be related to misreporting, as the coefficient for the model using self report is a larger than the coefficient for the model using corrected validated vote (0.040 > 0.035). This inflation is probably due to misreport attributable to memory decay by older respondent in the sample. However, Table 6 shows that when this coefficient is used to calculate the change in the probability of Y caused by the increase in age from one standard deviation below the mean to one standard deviation above the mean, the self reported vote model underestimates the increase (0.232 < 0.260). In all likelihood this inconsistency is due to the range for which this effect is calculated. If misreporting occurs only at high values for age, then the effects of misreporting bias, or overestimates of the effects of X on P(Y), will not be apparent at middle range values of age.

Income, education, the measures of efficacy, and political interest also all demonstrate inflated coefficients in the self reported vote model. These biases are probably due to the fact that voters who are high on these variables are more likely to think that voting is important and may try to disguise the fact that they failed to vote by misreporting. Of course NES's direct measure of citizen duty would be a an ideal way to identify this type of misreporting bias, but ambiguity in the question wording in 1988 made analysis of this variable problematic.[6] Table 5 shows that the magnitudes of these biases are fairly substantial for education, income, and political interest, around two standard errors. For the political efficacy scales the bias is much smaller, approximately one quarter of a standard error.

Table 6 shows that these biases translate into small changes in predicted probability of Y, at least for the middle range changes (one standard deviation below the mean to one standard deviation above the mean, holding all other independent variables at their mean) which I examine. A change in annual income from 10,500 to 47,500 dollars leads to an increase in probability of voting of 14.6%, with all other variables constant at their means, in the model using self reported vote as the dependent variable. However, this 14.6 increase is 2.4% higher than the predicted increase in P(Y) for a similar change in income using the corrected validated model (14.6 - 12.2 = 2.4). Similarly, the self report model overestimates the effects of middle range changes in education and political interest on P(Y) by 0.4% and 1.2% respectively.[7]

The effect of misreporting bias on the coefficients for the race variables is much more complicated. Apparently, Tables 5 and 6 show that the model using self reported vote underestimates the effects of being black on voting, while overestimating the effects of being a member of a racial group other than white or African-American. While race itself may be unrelated to misreporting bias, it is undoubtedly correlated with factors such as education and income which are related to misreporting. The point estimates for the race variables depend in part on the relationship between the dependent variable and other independent variables correlated with race. If the relationships between these other variables and voting are biased, it follows that the point estimates for race will be biased as well. Correlation patterns between the Xs in the sample effect the direction of the bias in complicated ways. Race is negatively related to voting, and negatively related to education, but the relationship between voting and education is positive and contains positive bias. This complex pattern of correlations makes the expected direction of the bias much more difficult to predict than in the simple bi-variate case.

There are three other factors to consider in terms of making sense of these unusual results for the race variables. First, Anderson, Silver, and Abramson (1986) establish that there is a tendency for African-American respondents interviewed by American-American interviewers to misreport voting. This misreporting bias, which applies to only small part of the sample, may need to be controlled for in order to examine the extent of the misreporting bias more clearly. Second, race of respondent is one of the primary potential sources of bias in the uncorrected sample (recall the example of Southern African-Americans being excluded by the validation process). Given that this is one of the biases which I suspected might exist using uncorrected data, it is not surprising that the magnitude of the bias between models is fairly large. In addition, if there are several correlations between race and biased independent variables all affecting the size of the biases in the race variables, it is not surprising that the size of these biases is substantial. Finally, the corrected validated model seems intuitively to match expectations of the effects of race on voting. Being African-American should lower the likelihood of voting substantially, as it does in the corrected model. Being a member of another racial group should lower the likelihood of voting as well, and it does. The direction of bias in the race variables for the self reported vote model is complicated due to positive and negative correlation among biased independent variables. However, the size of the revealed bias and the results of the corrected model fit with expectations.

The structural variables, which are measures of the election environment, should exhibit no bias due to misreporting. In election environments in which people are more likely to vote because of easier registration requirements, there is little reason to suspect that people are more likely to misreport voting. In addition, unless there is some relationship between the validation process and the registration arrangements, the correction for sample censoring should not affect these coefficients either. Consequently, no change in the coefficient estimates on these variables between the self reported and corrected validated vote models is expected.

The registration by mail dummy variable corresponds to this predicted relationship. The presence of registration by mail produces a nearly identical effects across equations (.39 for self reported, .40 for corrected validated). In terms of change in P(Y) resulting from the presence of registration by mail, the self reported vote slightly underestimates the increase due to this type of change, by 2.2%. While this is approximately the same size of the bias in income induced by using self report, one must consider that the change in the registration by mail dummy represents the entire range of the variable, while the change in the income variable only represents a sizeable change around the mean. Thus, a bias of 2.2% for the entire range of the variable seems relatively small.

The presence of alternative registration locations and extended office hours for registration both exhibit small biases in the self reported vote model. The coefficient for office alternatives increases by 0.068 (0.074 - 0.006), a small change relative to the standard errors of the coefficients which are around 0.16. Such a fluctuation could easily be the result of chance. The corresponding difference in the change in P(Y) associated with the presence of office alternatives is less than 1%.

The coefficients for extra office hours are unexpectedly negative, and the coefficient on this variable for the model using self reported vote is of a magnitude such that it would be statistically significant were a two tailed test employed. However, a negative relationship between the presence of extra office hours and voting makes little sense. The bi-variate relationship between this variable and voting is extremely small and positive (.02). The corrected validated model brings this variable more in line with the expectation that the effect of this variable should be positive or at least zero, but not significant and negative.

As expected, the coefficient on closing date is negatively related to voting in both models. The further from election day the registration rolls close, the less likely people are to vote. However, rather than being unaffected by the correction for misreporting, the effect of this variable decreases substantially. In other words, the coefficient moves closer to zero using the corrected validated vote model. Increasing the closing date of registration from 15 to 40 days decreases the likelihood of voting by 7.1% in the self reported vote model and by only 1.8% in the corrected validated vote model. This difference should not be attributed to misreport, as I have no reason to suspect that closing date has any effect on respondent's likelihood of lying about voting. Another possibility is that the correction for sample censoring reduced the bias in this estimate. The sample censoring may have systematically excluded respondents who were less affected by the closing date's depressive effect on voting. While this remains a theoretical possibility I have not discovered any systematic relationships between closing date and sample inclusion that confirm this hypothesis. For now, the unexpected substantial decrease in the explanatory power of this variable resulting from correction for misreporting and sample censoring remains something of a mystery.

<u>Conclusion</u>

This paper has been an attempt to correct for two potential measurement biases affecting voting models. Validating vote corrects for bias due to misreporting. For voting models, factors related to voting may also be related to misreporting, and therefore coefficient estimates on these variables will be biased. However, using validated vote instead of self reported vote may introduce other measurement biases. Systematic exclusion of certain sub-populations which vote at different rates or for different reasons than those included by the validation process leads to biased coefficient in the validated vote model. A selection equation was developed to account for this censored sample problem.

Overall, the results from the censored sample correction were encouraging. The preliminary equation was able to identify variables significantly related to inclusion in the validated vote sample. The predicted values from this equation were used to control for sample censoring bias, and to produce a corrected validated model which appeared to capture the causes of voting accurately.

The comparison of the corrected validated vote model with the self reported vote model revealed generally small but identifiable misreporting biases. The coefficients expected to be positively related to both voting and misreporting did exhibit bias. For the range of values I examined, this did not always translate into overestimates of the changes in the probability of voting due to changes in the independent variables. In some instances these findings may be a result of the particular range of values examined. These cases suggest that one must be very careful in interpreting substantive changes across limited ranges of typical changes in logit models. Any real bias correction may occur at the extreme values and so analysis of a middle range can obscure the true worth of the methodological improvement.

Similarly, in cases where variables affected by the misreporting biases are not simply positively correlated with voting, such as the race variables, the expected direction of bias becomes difficult to establish. In this case, the data show that the corrected validation model, which adjusts for both misreport and sample censoring, produces coefficient estimates which are in keeping with expectations. The adjusted coefficients on race are both substantial and negatively related to voting, with African-American status showing a stronger relationship with voting than membership in non-white, non-African-American groups.

Finally, for the structural variables (with the exception of closing date) the empirical results obtained by using the corrected validated model are consistent with theoretical expectations. Ideally, these variables were expected unaffected by misreporting and censored sample biases. Thus, the magnitude of the coefficients on these variables was expected not to change across models. This was the case for the presence of registration by mail. For the presence of office alternatives there seemed to be no relationship with voting across both equations, although the relationship in the corrected validated model was slightly stronger. The unexpectedly negative relationship between the existence of extra office hours for registration and voting in the self reported model was reduced in the corrected validated model. The resulting reinterpretation - that this factor was unrelated, rather than negatively related, to voting - was more in keeping with theoretical expectations. The relationship between closing date and voting in the corrected validated model is difficult to explain. While the self reported vote model may overestimate the strength of this relationship due to a correlation between selection into the sample and closing date, I have been unable to establish the nature of this bias statistically.

Correcting for misreporting bias by using validated vote is not enough. Censored sample biases introduced by the validation process can and should be corrected in models estimating voting. This technique is feasible in the case of NES data and results in improvements in accurately estimating coefficients. The results here generally confirm previous research which suggests that biases due to misreporting are small, but identifiable. Using the correct functional form, validation to address misreporting biases, and a selection equation to address censored sample biases yields results which are more believable, robust, and methodologically sound than previous efforts.

TABLE 1
Turnout Estimates Over Time

	1964	1972	1974	1976	1978	1980	1984	1986	1988
Aggregate	61.9	55.2	35.9	53.5	34.9	52.6	53.1	33.4	50.0
CPS	69.3	63.0	44.7	59.2	45.9	59.2	59.9	46.0	57.4
NES	77.7	72.8	52.5	71.6	54.5	71.4	73.6	52.5	70.0
NES - Agg	15.8	17.6	16.6	18.1	19.6	18.8	20.5	19.1	20.0
NES - CPS	8.4	9.8	7.8	12.4	8.6	11.2	13.7	6.5	12.6
CPS - Agg	7.4	7.8	8.8	5.7	11.0	6.6	6.8	12.6	7.4

Figures reported are turnout estimates (percentage of the voting age population that voted) and the difference between these estimates for the National Election Study, the Current Population Study, and aggregate voting returns.
Source: Traugott, 1989.

TABLE 2
Distribution of 1988 NES Vote Variables

Variable	Voters	Non-Voters	Missing Data
Self Reported	1235	540	---
(%valid)	(69.6)	(30.4)	
Validated	1033	619	123
(%valid)	(62.5)	(37.5)	

n = 1775 for both variables.
Source: 1988 NES.

TABLE 3

Logit for Selection into the Sample of Validated Voter

	b	s.e.
Constant	-0.296	.992
South	-0.403	.197
Size of Place	0.113	.068
Office Access Rating	0.277	.107

All coefficients except the constant are significant at $p < .05$

TABLE 4

Values for Selection Equation Variables in Terms of P(Y)

	LowX	HighX	diff
South	.945	.921	-.022
Size of Place	.924	.943	.019
Office Access Rating	.921	.945	.024

The table entries represent the corresponding values in terms of the probability of Y for the LowX, HighX, and the change in moving from LowX to HighX given in the other Xs are held at their mean values. For the interval variables the HighX and LowX values equal X +/- one standard deviation and for the categorical variables these values equal 1 and 0.

TABLE 5
Logit for 1988 Voter Turnout
(Self Reported and Corrected Validated Vote Models)

	Self Reported		Corrected Validated	
	b	s.e.	b	s.e.
Constant	-6.09*	0.49	-11.12*	2.56
Age	0.040*	0.004	0.035*	0.004
Income	0.069*	0.012	0.045*	0.012
Education	2.46*	0.39	1.90*	0.36
Black	-0.105	0.20	-0.52*	0.21
Other Race	-0.657*	0.33	-0.40	0.33
Internal Efficacy	0.0029*	0.0008	0.0027*	0.0007
External Efficacy	0.0033*	0.0008	0.0031*	0.0007
Political Interest	2.76*	0.32	2.03*	0.30
Office Alternatives	0.0066	0.171	0.074	0.160
By Mail	0.39*	0.15	0.40*	0.14
Extra Hours	-0.27	0.15	-0.18	0.14
Closing Date	-0.0160*	0.0054	-0.0033	0.0054
Selection	---	---	5.96*	2.69

* $p < .05$
Entries are logit coefficients and standard errors.

TABLE 6
Values for Variables in Voting Models in Terms of P(Y)

	Self Reported			Corrected Validated		
	LowX	HighX	diff	LowX	HighX	diff
Age (28/63)	.643	.875	+.232	.530	.790	+.260
Income (10,500/47,500)	.699	.845	+.146	.609	.731	+.122
Education (HSgrad/BA)	.674	.859	+.185	.576	.757	+.181
Black (0/1)	.786	.775	-.011	.709	.635	-.074
Other Race (0/1)	.801	.759	-.042	.689	.656	-.033
Internal Efficacy (187/382)	.729	.824	+.095	.612	.728	+.116
External Efficacy (175/375)	.718	.832	+.114	.601	.738	+.137
Political Interest (.42/.96)	.639	.877	+.238	.550	.776	+.226
Office Altern. (0/1)	.780	.781	+.001	.666	.680	+.014
By Mail (0/1)	.746	.811	+.065	.628	.715	+.087
Extra Hours (0/1)	.802	.757	-.045	.691	.654	-.037
Closing Date (15/40)	.814	.743	-.071	.682	.664	-.018
Selection (.90/.96)	---	---	---	.640	.704	+.064

The table entries represent the corresponding values in terms of the probability of Y for the LowX, HighX, and the change in moving from LowX to HighX given in the other Xs are held at their mean values. For the interval variables the HighX and LowX values equal X +/- one standard deviation and for the categorical variables these values equal 1 and 0.

ENDNOTES

[1] This paper draws heavily on research and analysis performed by Lisa D'Ambrosio and myself for the National Election Study in 1990. While a report of our findings has not as been completed, this work borrows heavily from drafts of papers related to that project. However, responsibility for any errors contained herein rests solely with me.

[2] A very small number of respondents, always less than one percent of the sample, report not voting when in fact they did.

[3] Even if misreporting is random, it will increase the error variance and therefore will create standard errors for the coefficients which will be inflated.

[4] Potential sample biases in NES such as non-response bias are not the focus of this paper and therefore the sample is assumed to be representative. Of course adjustments for this and other sampling biases are always possible.

[5] In addition to these election office variables I had suspected that certain sub-populations might be particularly difficult to validate due to name changes (e.g., divorced women) or unusual names (e.g., non-white, non-African Americans). However, in my preliminary analysis these variables did not seem to affect the probability of selection into the sample significantly. In addition, their exclusion from the selection equation did not seem to induce much omitted variable bias. In light of these findings and since the theoretical justification for their inclusion was not overwhelming, I decided to omit these variables from my final analysis.

[6] In 1988 NES asked respondents how much they agreed or disagreed with the statement, "If people don't care how an election comes out they shouldn't vote". Agreeing with this statement might be a response high on citizen duty, as a respondent might believe that people should vote no matter what. However, disagreeing with this statement might represent a form of citizen duty since people might feel that the democratic value of an informed electorate would be threatened by the participation of people who don't care.

[7] The even smaller biases in the political efficacy coefficients have a counter intuitive effect of underestimating the effects of this type of change in X on P(Y) for the model using self reported vote. However, since the biases in these coefficients are so small, and the effects of these changes are interpreted over a very specific range of values, it is quite possible that these variables demonstrate range specific effects

similar to those observed for age. The small, but real effects of political efficacy on misreporting may occur only at extremely high or low levels of political efficacy.

APPENDIX 1

<u>Self-Reported Vote</u>

Voters

- Said Voted, Vote Confirmed
- Said Voted, Registration Confirmed, Vote Not Confirmed
- Said Voted, Registration Not Confirmed
- Said Voted, No Name Given
- Said Voted, Unable to Check

Non-Voters

- Said Did Not Vote, Vote Confirmed
- Said Did Not Vote, Registration Confirmed, Vote Not Confirmed
- Said Did Not Vote, Registration Not Confirmed
- Said Did Not Vote and Said Not Registered, No Validation Performed
- Said Did Not Vote, No Name Given
- Said Did Not Vote, Unable to Check

Missing Data

- No Self Report (No Post Election Interview)

<u>Validated Vote</u>

Voters

- Said Voted, Vote Confirmed
- Said Did Not Vote, Vote Confirmed

Non-Voters

- Said Voted, Registration Confirmed, Vote Not Confirmed
- Said Did Not Vote, Registration Confirmed, Vote Not Confirmed
- Said Did Not Vote, Registration Not Confirmed
- Said Did Not Vote and Said Not Registered, No Validation Performed
- Said Did Not Vote, No Name Given
- Said Did Not Vote, Unable to Check

Missing Data

- Said Voted, Registration Not Confirmed
- Said Voted, No Name Given
- Said Voted, Unable to Check
- No Self Report (No Post Election Interview)

APPENDIX 2

<u>Political/Demographic Variables</u>
Age (from v414)
Income (from v520) 0-4,999; 5,000-9,999; 10,000-14,999; 15,000-19,999; 20,000-29,999; 30,00-34,999; 35,000-49,999; over 49,999
Education v422
Race Variables (from v414)
Internal Political Efficacy (from v940-943)
External Political Efficacy (from v937-939)
Political Interest (pre-election question v97)

<u>Structural Variables</u>
Other Registration Locations (from 1222)
Registration By Mail (from 1277)
Additional Office Hours (from v1286)
Registration Closing Date (from v1210)

REFERENCES

Abramson, Paul R., John H. Aldrich, David W. Rohde. 1987. Change and Continuity in the 1984 Elections. Congressional Quarterly Press.

Achen, Christopher H. 1986. Statistical Analysis of Quasi-Experiments. University of California Press.

Anderson, Barbara A., and Brian D. Silver. 1986. "Measurement and Mismeasurement of the Validity of the Self Reported Vote." American Journal of Political Science 30:771-785.

Anderson, Barbara A., Brian D. Silver, and Paul R. Abramson. 1986. "The Effect of Race of the Interviewer on Measures of Electoral Participation by Blacks In SRC National Election Studies." Public Opinion Quarterly 52:53-83.

Clausen, Aage R. 1967. "Response Validity: Vote Report." Public Opinion Quarterly 32(4):1-38.

Jennings, Jerry. 1990. "Misreporting Voting in the Current Population Survey." Unpublished paper.

Katosh, John P. and Michael W. Traugott. 1981. "The Consequences of Validated and Self-Reported Voting Measures." Public Opinion Quarterly 45:519-529.

Sanchez, Maria. n.d. Untitled Report for NES.

Silver, Brian D., Paul R. Abramson, and Barbara A. Anderson. 1986. "The Presence of Others and Overreporting of Voting in American National Elections." Public Opinion Quarterly 50:228-239.

Silver, Brian D., Barbara A. Anderson, and Paul R. Abramson. 1986. "Who Overreports Voting?" American Political Science Review 80:613-624.

Traugott, Michael W. and John P. Katosh. 1979. "Response Validity in Surveys of Voting Behavior." Public Opinion Quarterly 43:359-77.

Traugott, Santa. 1989. "Validating Self-Reported Vote: 1964-1988." Paper Presented at the Annual Meeting of the American Statistical Association. Washington D.C., August 7-10th, 1989.

Wolfinger, Raymond E. and Steven J. Rosenstone. 1980. Who Votes? Yale University Press.