VALDIDATING SELF-REPORTED VOTE: 1964-1988

Santa Traugott Institute for Social Research The University of Michigan

Prepared for the Annual Meeting of the American Statistical Association, Washington D.C., August 7-10th, 1989

The data utilized in this analysis were collected under grants from the National Science Foundation. The responsibility for any errors in analysis or interpretation is the author's.

VOTE VALIDATION IN THE NATIONAL ELECTION STUDIES

I. INTRODUCTION: WHY DO VOTE VALIDATION

For as long as we have been doing political surveys, it has been noted that the proportion of survey respondents who report to interviewers that they voted is substantially different than the estimates of turnout compiled from the aggregation of election returns. Table 1A (below) which lays out aggregate turnout estimates against turnout estimates derived from the self-reported vote in two national surveys of unquestioned quality and repute, shows the dimensions of the discrepancy. SRC/CPS survey estimates have never been less than 16% different, and since 1978 have been about 20% higher than aggregate estimates. The Bureau of the Census Voter Supplement Surveys, for reasons which we will note briefly below, do substantially better, especially in Presidential years, but usually are 6-7% higher.

And, for as long as we have noticed this discrepancy, analysts have tried to understand why it exists. Very early, it was concluded that a major source of the problem must be that respondents told interviewers that they voted when they had not. Why not look up the actual voting records of respondents? So, people have been trekking to election offices for at least the

past 47 years, trying to track down the presumed source of the error: who among the respondents claiming to have voted actually did not vote?

Since 1964, the election studies project at the University of Michigan (a project which is known now as the National Election Studies or NES since 1978, and in years prior to 1978 as the SRC/CPS election studies) has engaged in this effort for its biennial national surveys. A sizable body of data has accumulated, not always with careful thought for comparability of measures over time, and with not nearly enough attention paid to systematic evaluation of these efforts. This paper represents the beginnings of an effort to remedy this situation.

Why is it that data analysts should care about the discrepancy between survey and aggregate estimates of turnout? Why do we validate the self-reported vote of our respondents? There are two basic reasons.

1). Survey practitioners are or should be vitally interested in survey error. How much confidence can we have in the responses to questions we ask? What is the correspondence between what respondents say to a friendly, carefully neutral, middle-aged female white interviewer (the overwhelmingly typical situation) and what they actually think, feel, or, in this case, do? To the extent that we cannot believe in or even assess the correspondence between survey responses and individual attitudes or behavior, our enterprise is jeopardized. Misreporting is a "response error" of a kind. Who does it, and even, why and under

what circumstances?

Sometimes the determination of response error requires experimenting within surveys, and sometimes complicatedd statistical machinery must be brought to bear to disentangle response effects. One gets the impresion that in the early days of vote validation, the presumption was that vote validation would be a relatively trivial and straightforward operation. What could be easier than to look up the actual facts on administrative records? Later experience has shown, I believe, that vote validation is anything but an easy cut into the general problem of response error in surveys.

Thinking more broadly about surveys, fully to understand the discrepancy between aggregate estimates and survey estimates of turnout is to delve into the many different aspects of error in surveys, not necessarily related to response error. Clausen's (1967) dissection of the nature of the discrepancy between the 1964 election survey data-based turnout estimate and the national estimate is classic in this regard. He discusses a number of factors <u>aside from</u> misreport, which might lead to the discrepancy. He demonstrates first that denominators in the aggregate estimates which purport to measure the number of people who might have voted are questionable. More significant for this discussion at least, he shows that sample surveys have measurable error with effects on turnout estimates, in a) population coverage b) survey non-response, both initial and recontact (and both non-coverage and non-response are almost certainly highly

related to turnout) and in the stimulating effect on turnout of the interview itself: that is, the process of being measured has changed the object that is measured. Traugott and Katosh (1979) have elaborated on the latter finding. 1

The Current Population Surveys, which also evidences turnout estimates which are higher than the aggregated vote estimates, are consistently less far off the mark than the NES studies. The explanation for this appears to lie in slightly different popualtion coverage, a much higher response rate (high 90's for the single wave post-election interview as opposed to NES' response rate of .71 or .72 for the first wave of the panel, and .85 to .87 for the re-interview at which vote information is elicited), and probably the stimulating effect on NES respondents of the pre-election interview. ²

2) Political scientists, who are the chief consumers of election survey data, are interested in the effects on their data of misreporting. Indeed, reading the Clausen report of the first

¹As study manager for the election studies, I regularly review and read interview protocols and have in so doing have come across a disconcerting number of instances in which respondents spontaneously mention to interviewers that the whole process of being interviewed has certainly made them much more aware of and interested in politics, and they have made a certain effort to "study up."

²It would be handy if other surveys could somehow calibrate, or set outside bounds to, their over-report based on the Bureau's Voter Supplement Survey, since the Bureau's estimates are purged of most sources of discrepancy <u>except</u> for over-report. I think that this will not work. Response error such as misreporting of vote behavior must be based on survey conditions which vary from survey to survey: length of interview, question order, wording and context, personal or face to face and so on.

SRC/CPS vote validation, one is struck that the discrepancy between their survey estimates and aggregate estimates had been noted in the years 1948 to 1960, but it was the fact that 1964 survey respondents were for the first time discrepant in the two party division of the vote (Johnson getting 68% of the vote in the SRC survey compared to 62% nationally) that led directly to a concern with who the misreporters were, so that the error in vote division could be dealt with.

The chief effect of misreporting must be on models predicting voting turnout. If the sources of misreporting are correlated with the independent variables in these models, then these models will be wrongly specified and their results in error. What we need to do in vote validation is to give those who use the vote variables information to minimize the error with which a respondent is assigned to a vote/nonvote category. What is the best placement for an individual respondent in the voter/non-voter classification? Unless the vote validation exercise can shed light on this question, it has failed in its most important purpose, no matter how fascinating and interesting it is as a way to explore survey error.

II. VOTE VALIDATION EFFORTS AND RESULTING TURNOUT ESTIMATES Whenever vote records of individual respondents have been looked up, the results have been quite consistent.

"Here we may remark parenthetically that about 9 per cent of the individuals who told our interviewers that they had voted, had in fact not done so. This was established by referring to the Election Board records. Some had not

registered and therefore could not vote; others, though registered, did not cast their ballots on Election day. It turns out that the people who had previously expressed a <u>high</u> level of interest had a greater number of .. "liars" than people who had a medium interest level or no interest at all. "

Thus, Kitt and Gleicher describe vote validation efforts for the Elmira study, carried out in 1948. A number of other efforts, beginning with Parry and Crossley in 1942, have been made to validate self-reported vote in surveys. Most of these efforts across the last 45 years, have had results similar to those cited above.

This paper reports on work done at the Institute for Social Research by the SRC/CPS National Election studies since 1964, to determine the validity of respondents' self-reported vote. The project itself has lagged in its own evaluation of the thoroughness and completeness of the exercise now carried out after each biennial elections study, and in clearly stating for consumers of the elections study data, with what degree of confidence they might use the vote validation data, and in what ways the various data collections are comparable over time.

A project staff has the most information to evaluate its own data collections. In recent years, the NES project staff, the principal investigators and the Board of Overseers of the National Election Studies have all agreed on the importance of technical evaluations in many areas related to its survey operations. The relentless schedule of field operations, from questionnaire design to data release, leaves few windows of time

for such tasks. This paper represents for us the beginning of the overdue evaluation and description of vote validation by the National Election Studies.

The election study at Michigan has conducted vote validation exercises for the elections of 1964, 1972, 1974, 1976, 1978, 1980, 1984, 1986 and 1988. Most of the discussion which follows will include only vote validation for 1964, 1976, 1980, 1984, and 1988. Moreover, the emphasis will be on the elections of 1984 and 1988. I have chosen to de-emphasize the off-year election validations because a) the actual turnout calculations (as Jerry Jennings has pointed out) are more problematic than those in Presidential years and b) much more important, the stimulus given to our respondents in the post-election interview in off-years is quite different than that in presidential years when they have been interviewed before the election. Presidential and off-year misreporting are probably not comparable, and in ways which we don't yet understand.

The results for 1972 and 1974 are even more problematic. This vote validation was carried out as part of the 1976 vote validation, in 1977, and 5 years is a long time to look into the past for a valid voting record.

Table 1B displays the results over time of vote validation efforts. Looking specifically at Presidential elections, the SRC/CPS/NES turnout is typically in the low 70's, while the aggregate turnout estimate is in the low 50's. When we revise our turnout estimates to reflect only those self-reported voters

whose report can be confirmed (or can not be checked because they voted out of the area or didn't give us their name) the "Valid vote estimate" is 9-12% below self-reported estimates. When we look at the distribution of self-reported voters by validation results, and summing all R's whose vote could not be confirmed (rows labled "R did not vote, "can't find record", and "registration record found but R was not eligible to vote," we find that 12-14% of self-reported voters do not actually have a record of voting. This result is consistent over time and consistent with other findings.

Some aspects of the vote validation results have changed over time. The "Record check can't be done" category has changed in magnitude. This category consists of respondents who did not give us their name (a small but steady component) ; those who lived outside of the PSU and for budgetary reasons, were not validated, and a scattering of administrative reasons. In 1984 and in 1986 we made some effort to track down the out-of PSU respondents over the phone, with mixed success. We did not do this in 1980 or 1988. 1980 also saw some administrative problems simply in getting the forms for individual respondents out to the field, probably a consequence of carrying out the vote validation exercise before the election study interviewing and coding was completed.

We have reduced very modestly since 1978 the proportion for whom records cannot be found, in tandem with the proportion classified as misreporting. It is probably relevant to note in

this connection that in 1984, in order to reduce misreport, we changed the position of the vote report in the post-election survey. Previously, since we thought of asking respondents if they voted and for whom as somewhat delicate, we waited until very close to the end of the interview to ask about voting behavior. In 1984, we began to consider that having just gone through 55 or 60 minutes of questions about political attitudes and behaviors might induce respondents to paint a consistent picture of themselves in their report of vote and it would possibly be better to elicit such information before interviewrelated factors took effect.

More significantly, or so it seems to me, the share of the unconfirmed voters which comes from the "Record not found" has been stable over time. Since 1972, the contribution to unconfirmed voters of the "not found" and "R did not vote" categories has been roughly equal. (The "R did not vote" designation means that a record was actually located for the R but was not marked to indicate that he/she voted, <u>or</u>, that a registration record was located for the respondent but it had been canceled or purged before the election in question.) And, who these "not-founds" might be has troubled users of the validation data. How do we know that these are not simply difficult to find voters?

Table 1B gives two other valid vote estimates (for Presidential elections only). One line gives the estimate of validated vote if we treat "not-founds" as missing data. The

following line shows what happens if we accept the self-report of voting for the "not-founds." It would be good to know, in evaluating vote validation efforts, whether we have reduced over-report by 5%, 7% or 11% in 1988.

We have over the years tried very hard to reduce the proportion of 'Not-founds" and have succeeded to some extent. But the further attack on the "not found" problem needs to be several-fold: we need to analyse who the "not-founds" are, compared to misreporters and validated voters, and see how much difference they make in analysis. The last section of this paper will address that issue specifically. We need also to improve our search procedures, <u>and</u> our description and analysis of office procedures which might explain findability of respondents. What are the procedures that have been used over time to carry out vote validation?

Vote validation almost always involves a visit by an SRC field interviewer to the actual office in which election records for the jurisdiction in which our sample areas fall are stored. There, the interviewer (not infrequently the interviewer who spoke to the respondent in the first place) tries to find a record for a respondent in the office and to determine from that record, (and/or from vote records if these are kept separately) whether the respondent voted. The indication that a respondent voted is different from office to office: sometimes it is a signature on a poll list, sometimes it is an "X" marked on the registration record, or a

date-stamp, or more latterly, a computer entry. The person looking up the records has no prior knowledge of the respondent's self-report on vote, although she does know that all those she is looking up have claimed to be registered. Regrettably, we have very little information about the 1964 vote validation exercise. It was probably done in late 1965 or early 1966. But starting in 1976, we have the following rough parameters of the validation operation:

Elections	Field dates	Number of OFfice
1972-74-76	Aug-Oct. 1977	147
1978	Feb-March, 1979	178
1980	Jan-Feb, 1981	178
1984	Feb-March,1985	120
1986	April-May, 1987	175
1988	May-June, 1989	120

The staff has struggled with the question of appropriate field dates for the study, although in practice our freedom of action is quite circumscribed by the volume of other work going on at the same time. Anecdotal lore suggests that there is a time beyond which records vanish or are sealed or access becomes more difficult. We have not, in recent years, found any particular relationship between accessibility of records and date of the field period. It's probably better to get to a jurisdiction in February-March, before the spring elections which

are held in many if not most jurisdictions because in some cases, the most easily checked form of vote recording is the computer file, and that file may only list the last time the individual voted (since one reason the files are kept with vote information is to determine who to purge for not voting). This is not an insuperable problem, since alternative sources, typically the poll lists which are the input to the computer, are usually available, but it means an extra step for the look-up person.

Registration and voting records are most often kept at the county level. Therefore, our "offices" are basically in the county seats in the counties where we have sample segments. In some states, for example, New England township states, registration records are kept in local or township offices, which are typically less professionally run than county offices. Large cities like Chicago and Detroit, and independent cities, also keep their own registration records. We have gone to the same offices in 1984, 1986 and 1988 (in 1986 we used a larger version of the SRC sample frame than 1984 or 1988; hence more offices).

Access to registration records has virtually never been refused, at least since 1980. There have been a few situations in which the vote records themselves are not accessible at all. In 1984, for example, we discovered that a particular jurisdiction was burying its vote records in the garbage dump, although the state had recently passed legislation requiring records to be purged for non-voting. (The registrar was beginning to feel uneasy about the necessity of purging on the

basis of buried information about voting; in subsequent years the situation has been corrected.) It is somewhat more common that a portion of the vote information is not accessible, at least temporarily. Office records or master files are frequently updated with vote information that is generated at the polls, signature lists or ballot applications, or other materials used by election workers to check off when a person has voted. While we are free to look at the office records or master files, these original materials may be not available for one reason or another. Where possible, our practice has always been to look at these original materials as well as the office records, usually computerized by now, to determine if a record of notvoting is due to an error in transcribing the vote information generated at the polling place onto the office records. In 1988, for example, we determined that 7 respondents who were not recorded as voting on official records in the central office were shown to be voting on the poll records. This is in fact an astonishing low level of recording or transcription error, reflecting on the high regard for accuracy and meticulous record keeping that characterizes the great majority of the offices we have visited.

In 1984, we made a number of innovations in our vote validation procedures. I think we have made our operations more efficient, better documented and the life of the project staff somewhat simpler, but I am not convinced that we have dramatically affected our results. We made in 1984 special

effort to involve field office <u>supervisors</u> rather than interviewers in the vote validation effort. We believed that some offices might require especial diligence in working through disorganized records, or that some election officials would require special persuasive efforts to cooperate and tell us what we needed to know about how records were kept. In addition to the supervisors we asked for and felt that we got assignments of interviewers known to be perseverant, tenacious, accurate and undeterred by cranky bureaucrats.

In 1986 and 1988, the SRC's schedule of field operations did not allow us the luxury of specifying the interviewers we wanted, much less field supervisors, so we retreated from the 1984 policy, without, I think, deleterious effects on the data. In 1984, we were unable to find any record of 6% of our selfprofessed voters; in 1988 this figure was 7%. In 1984, we found registration records for 17 people, but were not able to gain access, for one reason or another, to appropriate voting records. This number had reduced even further to 12 and 11 in 186 and 1988 respectively.

A number of other innovations in data collecting for the vote validation data were introduced in 1984. In 1984, we stopped validating respondents who told us that they did not vote and were not registered. (In 1980 only 2 individuals who had told us they were not registered and did not vote were recorded

as having voted.³) It seemed fruitless for interviewers to spend time and morale combing through records for information which did not exist. Indeed, this may have been the source of the many complaints about how hard it was to find records in 1980 and the lamentable disorganization of office record keeping. However rational this project decision was, it does introduce noncomparability into the series, because we did not ask selfreported non-voters whether they were registered until 1978. Therefore, we cannot exactly replicate our search procedure before that.

Information on respondents to be checked are sent out on "forms." In 1980, and earlier, there were as many forms per respondent as there were conceivable addresses for that respondent: his pre-election address, his post-election address, if different; the address where he lived and the address where he told us he was registered, if different; and so on. The proliferation of forms was intended to cover all possible bases; it seemed to me to have caused a certain amount of wheelspinning, and we decided in 1984 to simply send out the address at which the respondent <u>said</u> he was registered. This has not noticeably affected our "found" rate, probably because of the

³It is theoretically possible to report that you did not vote when you did. There have always been a scattering of people whom we find listed as voting who report not-voting. While in the 1970 studies that number might reach 30, our latter experience has been only 5 or 6 cases (alhtough we reached 12 in 1988). I think this is probably transcription error, and as such, is still another indication of the care with which records are kept.

fact that voting at addresses different than where one lives is not at all uncommon and not particularly hard to deal with for the lookup person. I would think that up to 15% of respondents are registered at an address different than the one at which they are now living. In 1988, 17.4% of the respondents whose record we found had registration addresses different than their sample address. (This is based on the interviewer's judgement as to whether this was a different address; closer examination which I intend to do may show that many of these seeming differences are between box numbers and rural highways, or otherwise non significant.) If the respondent's real registration address is within the same jurisdiction (usually county) that his/her sample address falls in, they are findable. Very few offices are set up so that you must have the exact address of the respondent to locate him or her. We are, though, surprised by the magnitude of the outdated addresses, especially in view of the number of offices that assure us of the regular address updating of their files by one means or another (usually by mailing a nonforwardable card or letter to the registrant).

Also in 1984, we began administering what we call an Election Official interview with the person in charge of the records, before we start the lookup procedure itself. In 1976, there was also an interview with an election official, but this had as its chief focus the understanding of how the election laws were administered in that jurisdiction. While we were concerned in 1984 and remain so in 1988 with election law administration,

particularly as it affects registration rates, when we began in 1984 to interview election officials, our chief purpose was to <u>educate our interviewer</u> about how the records were kept in that office. Secondarily, we wanted to build rapport and a common vocabulary of terms between the interviewer and the election official. Mostly, we think that effort has succeeded. Interviewers at least walk away from the interview thinking that they understand how the records are kept; and frequently the election official becomes sufficiently interested and involved to take part in the lookup process itself.

The form itself (see attached) on which information is recorded has been standardized. It takes interviewers on a complicated path through election records, trying to make sure that for each respondent for whom we are looking, they have touched all possible bases. I think that the procedures we have developed now give us the capability of evaluating the officerelated reasons why we find or don't find information for specific respondents, but it is not as clear that we have had any measurable effect on the <u>results</u> of the lookup procedure. In particular, the proportion of self-reported voters that we can't find any record of has not changed between 1980 and 1988.

Virtually all of the data that we have collected in the course of vote validation is now or will shortly be available for the use of others. All of the individual voter record data has been released, usually merged with the election study data itself. The election office interviews for 1986 have been

released, and the 1988 interviews are coded and will be released soon. (Because they are not yet released, I have not been able to draw on them in any significant way for these analyses.) The 1984 election office data are not coded; but we plan to do this. We have also the beginnings of a file pooling vote validation data from the last several elections; we would like, as time and funds permit, to expand and elaborate this file, adding contextual data, including that from the election office interview to it.

As we continue to try to improve the current vote validation procedures, basically focusing on how to reduce the proportion of records we are unsuccessful in locating, we will try in 1990 to implement a two-stage search procedure, in which interviewers can call the Ann Arbor office for instruction about individual cases when they are unable to locate a record. It is barely conceivable that with detailed reading of information available to us in coversheets and protocols, we could find some clue for a further search, or for a search in another area. If this strategy involves sending interviewers <u>back</u> to an office they have already visited, or to another office entirely, the likely cost will be a deterrent, since costs vary quite directly with number of office visits.

III. THE VALIDATED VOTE AND POLITICAL PARTICIPATION.

In the remainder of this paper, I begin on two tasks: how do vote validation results affect analyses of voter participation in both registration and voting, with particular empahsis on

whether the distinction between "not-founds" and "misreporters" makes any difference, and, the analysis of source of the response error: who are the over-reporters, again distinguishing between the not-founds and the misreporters.

Table 2 pools data from the 1984 and 1988 election surveys and vote validations. These two studies were chosen because the data and sample frames were most comparable. 1986 was not used because of concern about differential panel mortality and stimulation in an off-year election study, which has only a postelection interview, as opposed to the Pre-Post interviews in Presidential years. We validate only those cases which were interviewed in both the Pre and Post election waves; since 1980 we have validated the "no-posts" but I have not included these cases in these analyses. Those who did not give their name or who reported being registered outside of election study sample areas are treated as missing data -- not included in this table.

Table 2 displays the most elemental bivariate relationships between factors related to registration and turnout, and the results of vote validation. For vote validation results to have a measurable impact on models of voter participation, misreporting should correlate with the independent variables. If its effect is similar at all levels of independent or conditioning variables, then the results of analyses of turnout using validated and unvalidated data won't differ greatly. Some who write on validated data results suggest that this is the case (Sigelman, 1982).

The first two columns of Table 2 relate to registration, showing the percent in each category of various independent variables who are validly registered, and the difference between valid registration and reported registration for each category. The second set of columns repeats the exercise for validated voting. Looking first at validated voter differences, we see that the differences between categories of independent variables are not generally large. With the exception of race, a difference of 6% between largest and smallest proportion of overreporting is the maximum difference among categories of independent variables. It's hard to see, for example, that the relationship between age and either registration or voting would be affected very much by use of validated data, although overreporting diminishes slightly with age, a factor which seems to show up in other parts of the table; where over-reporting is less for retired and disabled, for widowed, for less than eighth grade education -- all factors relating to age. Certain other categories seem most prone to misreporting: the unemployed, renters, divorced. Perhaps these are somewhat more marginal people, and perhaps the over-report of vote problem we have with these people is related to their mobility: they have recently changed residences in some way, and are less findable than others as a result of their status change.

Race is a case of special interest. It has the strongest differences in the table, although a 10% difference is not large. (The racial differences result is probably echoed in the urbanism

variable.) Anderson <u>et al</u>. (1986) have demonstrated that this difference is related to whether or not the respondent was interviewed by a black interviewer, with the notion being that the social desirability of reporting voting is higher in that situation. The worry here is that blacks might be located in electoral offices where record-keeping is less well-done than other areas. Is black over-report due to not-being found, or to confirmed registration without a record of voting (misreporting)?

Table 3, which lays out the results of validation in detail for each category of independent variables, allows us to compare the misreporters and the not-founds, and only modest differences emerge. Compare column 3, which includes the misreporters, and column 4, including the "not-founds" who reported voting. They track very similarly, across almost all categories.

There are exceptions, and they are instructive. For the oldest group in the sample, aged 75 and up, misreporting contributes noticably more to over-report than not being found. It's possible that short term memory fails some older respondents. In our measure of "transience" or life-cycle stability, combining unemployed, divorced, renters, young, and recent movers, we find that at the highest category, Not-founds contribute more to over-report, as we would expect. Very importantly, for black respondents the contribution to overreport of <u>mis</u>-reporting is twice that of not-being-found, which makes me think that an analysis of office record-keeping is not going to do much to reduce the phenomonon of black over-report.

In the non-demographic measures, where respondents report on political attitudes, media exposure and other kinds of political participation, things start to look a little different. If there is any movement across categories of an independent variable, it seems to go in opposite directions depending on whether the source of the over-report is misreport or not being found. People who are very attentive to the media, for example, are more likely to misreport than to be elusive in the records, while the least attentive are more difficult to locate than likely to misreport.

At this point, I view the "not-found" category as somewhere between an irritating source of noise in data analysis and a moderate problem for the study. The preliminary analysis does not yet show me that the the respondents whose records we cannot locate are markedly different than misreporters.

I bring to this discussion my complete faith in the thoroughness and skill with which the search for the missing respondents is waged. Almost always, every conceivable office record has been checked. Names have been searched under any number of possible mis-spellings, and many times, we have found other members of the same household registered at the same address, but not the respondent. It is not at all uncommon for the ranking election official to become challenged by the search and devote herself to ransacking records in hopes of finding an elusive respondent. Mostly, when we look at the records of searches for the unfound self-reported voter, it is hard to

imagine how more could have been done to find the respondent. So when the interviewer writes, "if this person were registered in this office, I would have found him/her," I tend to believe that the person is indeed not registered in this office, under this name.

It is certainly the case that in spite of thorough instructions (and impassioned pleas) to the survey interviewers to spell and record the names of respondents carefully, we regularly find respondents under entirely different names than expected. If the names of respondents are radically misspelled, especially in the first letters, we can only find them when there is a cross-reference address directory, which shows a person of a highly similar, but radically mispelled name, and of about the same birthdate living at the same address as the respondent. And given the number of people not listed on their registration address at the same place where they now live (where they were interviewed by us) it is certainly possible to believe that some people are actually registered in an adjacent county or township, near where they work, or used to live.

The alternative to believing that our "not-founds" are really registrants about whom we lack sufficient information to locate, is that there are adults willing to be interviewed about politics, not once but twice, in the Pre-election and Postelection wave, who have not registered to vote in living memory in the community in which they live. (Remember, these respondents are not even found in purged, canceled, or inactive

files.) After some consideration, I believe that for the most part, that's who these people are.

In the election study, we only ask registration status of persons who told us they did not vote. People who tell us that they voted, but who are really not registered, are reporting registration only by implication. (That is why I have not stressed the use of the vote validation data for the study of registration.) In fact, the variables which make it more difficult to find people -- being divorced or separated, unemployed, renting rather than owning your home, are also related to the probability that people are registered in the first place. (See table 2.) My tentative conclusion about the difference between those self-reported voters and whom we cannot find and those for whom we find a record, which does not indicate that they voted, is that the not-founds are not registered, and Both are misreporters of vote, but the the misreporters are. not-founds are in effect also misreporting their registration status.

The literature on the validated vote is accumulating. It is hard to compare, because the measures of voting or not voting do not always use the same denominator. In particular, a number of important recent articles report results based on the notion that in order to understand the true effects of misreporting on models of voter participation, we need to calculate misreport as the proportion of respondents who did not actually vote who told you that they did. (See particularly Anderson and Silver, 1986.)

The argument is that we should be studying misreporting as a function of the opportunity to misreport and on the basis of populations "at risk" for misreporting, which is those who actually did not vote (since the number of those who told you that they did not vote but for whom the records show that they did is usually qutie trivial). When looked at from this perspective, the amount of over-reporting gets estimated at between 20 and 30%. Most of this analysis, as I understand it, puts in the "did not vote" category, those respondents for whom we can not find a record.

For comparison purposes, I am including Table 4, which looks at the non-voters in the sample -- all those for whom a record of voting could not be confirmed -- by three categories: admitted non-voters, vote misreporters, and self-reported voters for whom a record could not be found. The results are basically similar to the earlier analysis. By and large, the two misreporters and not-founds track along together, with some interesting exceptions, including age, race and "transience".

We see clearly in Table 4 however, that the propensity for nonvoters to tell interviewers that they voted increases strongly with social class, education, political interest and degree of partisanship. Looking at education specifically, it is striking that 87% of the respondents with low education admitted not voting, while less than 50% of those with a higher than college level eduation did so.

I am inclined to think that the social desirability phrase

most often used in these and similar analyses to describe motiviation for misreporting oversimplifies the situation. If it were, misreporting should be different over the phone than in person, and it appears not to be. (In 1984 we administered half of the sample in the Post-election survey by phone, half face to Distributions on the validated vote variable are virtually face. identical.) Certainly, the black respondent over-reporting when talking to a black interviewer is operating from a different mindset than the 83 year old who can't really remember one presidential election from another at certain points. Some have argued that the over-reporter is really a person who usually votes but who, more or less accidentally, was unable to vote in the particular election about which we asked. Presser and Traugott (1983) have shown, however, in their analysis of validated vote in the 1972-74-76 Panel, that the tendency to misreport in one election is highly correlated with the tendency to misreport about another election. These are habitual misreporters, rather than habitual voters.

The variables which we typically include in an election study do not give us much purchase on the psychological motivations for misreport. But I believe it is worthwhile to follow up on trying to understand its source, because if a respondent has misrepresented his political behavior in a socially desirable direction on voting, why not on other measures such as political participation, media exposure, interest in politics, following the campaign, and so on? Only if we

understand the sources of the misreport can we hope to think about altering the survey situation itself, to mitigate the factors which seem to trigger misrepresentation.

References

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

Anderson, Barbara A., Silver, Brian D., and Abramson, Paul R. 1986. The effects of race of interviewer on measures of electoral participation by blacks (in) SRC National Election Studies. Populaion Studies Center, Univ. of Michigan, Research Report.

Clausen, Aage R. (1968) "Response Validity: Vote Report." <u>Public Opinion Quarterly</u> 41: 56-64.

Kitt, Alice S. and David B. Gleicher. "Determinants of Voting Behavior: A Progress Report on the Elmira Study." in <u>Public</u> <u>Opinion and Propaganda</u>, ed. by Daniel Katz, Dorwin Cartwright, Samuel Eldersveld, Alfred McClung Lee, NY: The Dryden Press. 1954, pp. 413-425.

Miller, Mungo. "The Waukegan Study of Voter Turnout Prediction." in <u>Public Opinion and Propaganda</u>, ed. by Daniel Katz, Dorwin Cartwright, Samuel Eldersveld, Alfred McClung Lee, NY: The Dryden Press. 1954, pp. 751-760

Parry, Hugh J. and Helen M. Crossley. "Validity of Responses to Surey Questions." in <u>Public Opinion and Propaganda</u>, ed. by Daniel Katz, Dorwin Cartwright, Samuel Eldersveld, Alfred McClung Lee, NY: The Dryden Press. 1954, pp. 738-750.

Presser, Stanley, and Traugott, Michael. 1983. "Correlated response errors in a panel study of voting." paper prepared for presentation at the Annual Meeting of the American Associatonof Public Opinion Research, Buck Hill Falls, 1983.

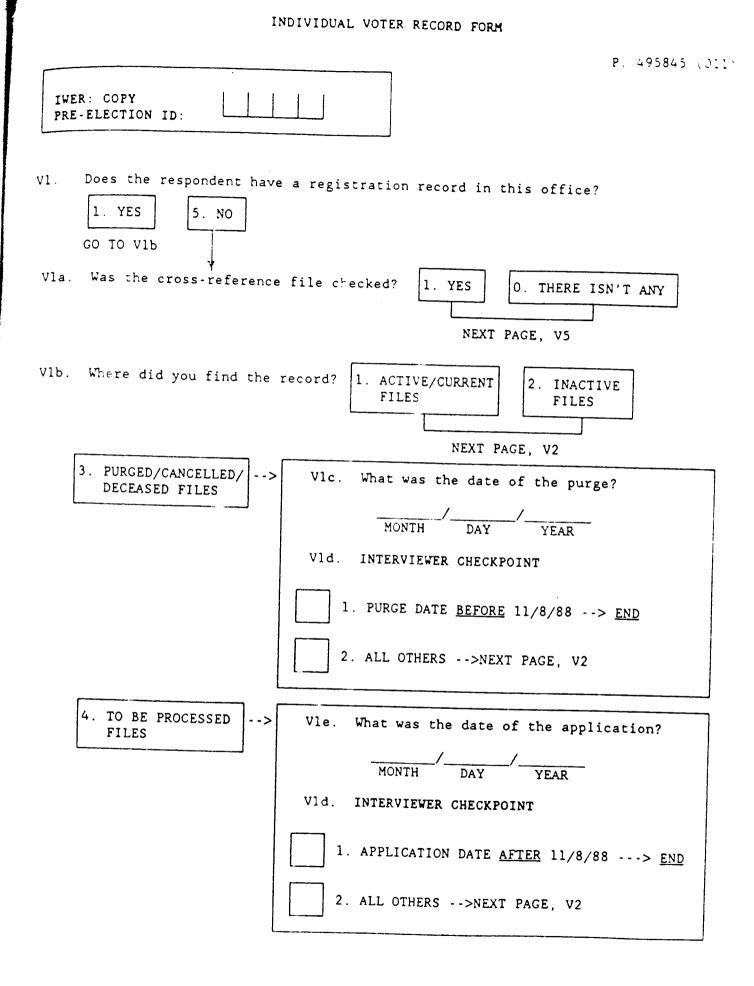
Sigelman, Lee. 1982. The nonvoting voter in voting research. American Journal of Political Science. 26:47-56.

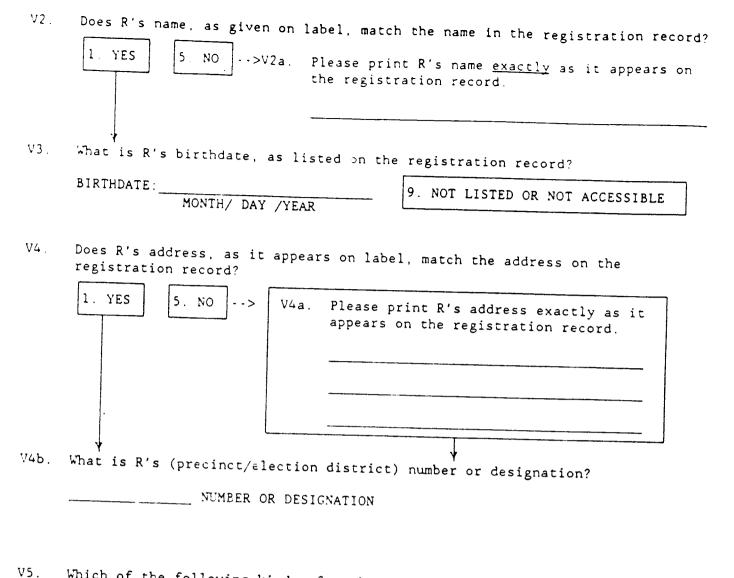
Silver, Brian D., Anderson, Barbara A., and Abramson, Paul R.. 1986. "Who Over-reports Voting?" <u>American Political Science</u> <u>Association.</u> 80: 613-624.

Traugott, Michael and Katosh, John P. 1979. "Response Validity in Surveys of Voting Behavior." <u>Public Opinion Quarterly</u>. 43: 359-377

III. INDIVIDUAL VOTER RECORD FORM

Đ

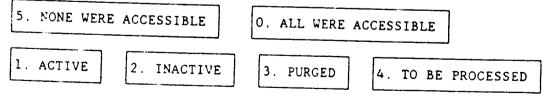


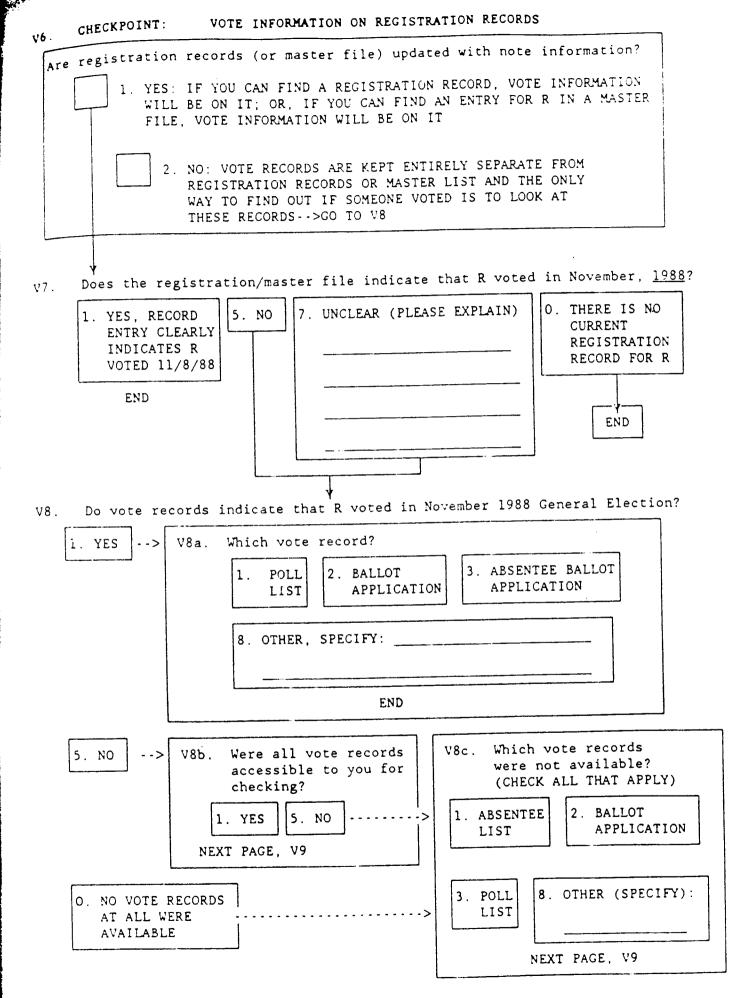


V5. Which of the following kinds of registration records, whether accessible to you or not, are kept by this office? (CHECK ALL THAT APPLY).

1. ACTIVE 2. INACTIVE 3. PURGED	4. TO BE PROCESSED 8. OTHER: (SPECIFY)

V5a. Of these sources, which ones were <u>not</u> accessible to you for checking? (CHECK ALL THAT APPLY)





MONTH / DAY / YEAR

HAS NEVER VOTED

THUMBNAIL SKETCH:

Please note any ambiguities in the registration/voting record or any other pertinent information that you think project staff should know about this case.

Table 1A- 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

TABLE 1B. DISTRIBUTION OF NES SELF-REPORTED VOTERS BY VOTE VALIDATION RESULTS

	1964	1972	1974	1976	1978	1980	1984	1986	1988
SAMPLE N	1450			2416		1407	1991		1775
SELF-RPT N Records show R	1126	1574	1320	1732	1250	1004	1466	1142	1237
voted	75.5	57.2	65.8	80.3	74.5	77.5	858	79.4	83.8
Record check can't									
be done	15.8%	15.3	12.9	3.2	2.7	9.0	2.3	4.1	1.7
Didn't vote	2.2%	12.0	10.2	5.2	9.8	6.2	6.1	8.7	7.4
NOT FOUND	6.1	15.6	10.5	11.4	11.9	7.4	5.8	7.8	7.1
PERCENT OF									
SAMPLE CONFIRMED	70.9	52.3	41.2	59.9	42.1	61.8	64.3	43.8	59.5
DIFF: SR-VALID	6.8	20.5	11.1	11.8	12.4	9.6	9.3	8.7	10.5
CONFIRMED(2 CONFIRMED(3	•			64.7 68.0		65.1 67.0	67.8 69.1		62.8 64.8

TABLE 2: Validated Registration and Voting by Social and Political Groups

	-			_
	8	DIFF	8	DIFF
	VALID REG	REPORTED- VALID	VALID VOTER	SELF-REPORTED- VALIDATED
	KEG	VALID	VOILK	***************************************
OVERALL AGE				
17-24	51.4	8.8	37.6	10.0
25-34	67.5	8.9	56.2	9.8
35-44	76.2	9.1	64.5	12.5
45-54 55-64	80.4 81.3	6.6 7.9	71.0 71.5	9.6 9.0
65-74	84.4	4.7	75.6	7.0
75-79	77.6	4.3	63.4	6.7
<u>SEX</u>				
MALE	74.2	7.9	63.0	10.4
FEMALE	72.4	7.6	61.8	9.2
RACE				
WHITE	74.8	7.0	65.4	8.6
BLACK	65.3	11.7	44.2	18.3
OTHER	55.0	14.0	44.0	7.0
<u>URBANISM</u>				
	75 4	9.0	62.2	13.5
CENTRAL SUBURBAN	75.4 74.0	5.9	65.0	8.0
NON-URBAN	70.4	9.1	59.0	9.1
HOMEOWNERS				
OWN	80.5	6.3	70.5	8.7
RENT	58.6	10.5	46.1	11.7
WORK STATUS				
-		0.5	(2.0	10.0
WORKING NOW	74.0	8.5 9.5	62.9 40.1	10.6 12.2
UNEMPLOYED RET/DIS	54.1 78.8	9.5 5.5	69.6	6.9
STUDENT	71.5	6.9	60.6	8.9
HOUSEWIFE	64.1	6.9	-	-

Table 2 (Continued)

		१ VALID REG	DIFF REPORTED- VALID	ै VALID VOTER	DIFF SELF-REPORTED- VALIDATED
MARITA	<u>L STATUS</u>				
MARRIE NEVER DIV/SE WIDOWE	MARRIED P	77.3 64.8 64.5 75.1	7.0 8.4 11.1 6.0	68.3 51.6 49.1 64.0	8.1 12.1 14.7 8.5
TRANSI	ENT				
LO	0 1 2 3 4	84.0 70.7 63.2 50.8 53.4	5.0 9.5 8.8 11.0 16.4	75.9 59.2 44.4 37.6 34.9	7.2 11.4 12.7 11.0 12.3
HIGH EDUCAT		55.4	10.4	34.2	12.5
LO	1 2 3 4 5	59.7 57.7 66.4 75.2 81.8	9.6 8.2 8.2 8.1 7.2	47.6 44.9 54.8 63.4 70.7	6.7 8.0 9.4 12.6 11.0
HIGH	6	90.0	5.9	82.6	9.4
<u>SOCIAL</u>	CLASS				
TOM	1 2 3 4 5	64.2 62.8 76.9 79.4 82.5	9.0 9.3 7.8 6.5 10.0	51.6 48.8 64.4 69.7 77.5	10.4 11.7 10.6 9.2 9.6
HIGH	6	85.6	6.8	79.1	
INCOME					
Percen 0-16 17-33 34-67 68-95 96-10	ક ૪ ૪ ૪	54.5 69.6 74.0 82.5 91.1	9.7 7.6 8.3 6.2 4.8	41.7 56.5 62.8 73.5 84.9	8.0 9.4 10.6 10.3 8.2
PARTIS	ANSHIP				
IND/AP LEAN I WEAK STRONG	ND	52.4 68.6 72.9 85.1	8.7 7.9 8.1 6.6	41.7 57.7 61.0 75.4	8.0 9.7 10.0 10.0

Table 2 (Continued)

		% VALID REG	DIFF REPORTED- VALID	% VALID VOTER	DIFF SELF-REPORTED- VALIDATED
POL IN	TERES	ST			
NOT MU SAME VERY	СН	51.1 74.6 87.7	10.0 7.8 5.8	37.5 63.4 79.3	6.7 11.1 9.9
PARTIC	IPATI	<u>LON</u>			
LOW HIGH	1 2 3 4 5	65.2 83.0 87.2 92.2 100	7.9 6.8 7.8 5.2 0.	53.1 73.2 80.9 81.9 -	9.3 10. 10.6 12.9
MEDIA	EXPOS	SURE			
LOW HIGH	1 2 3 4 5	52.5 68.4 79.5 83.7 85.5	9.4 8.3 6.0 6.2 8.3	39.1 56.8 68.5 73.8 79.8	7.1 9.9 10.3 10.3 10.4

TABLE 3. DISTRIBUTION OF 1984 and 1988 ELECTION STUDY RESPONDENTS by Validation and Social Group Categories

	1 Valid Voters	REGI 2 Vali- dated Non- Voters	STERED 3 Self- Re- ported Mis- Report	4 Self- Re- ported Voter	NOT REGI 5 Self- Re- ported Reg/ No R Vote	STERED 6 Self- Re- ported Non egistered	И
OVERALL							
<u>AGE</u> 17-24	37.6	8.9	5.0	5.0	2 0	30 7	410
25-34	56.2	6.7	4.6	5.2	3.8 3.7	39.7 23.7	418 849
35-44	64.5	5.6	6.1	6.4	2.7	14.8	791
45-54	71.0	4.6	4.8	4.8	2.2	12.6	459
55-64	71.5	3.2	4.5	4.5	3.4	13.0	471
65-74	75.6	4.9	3.9	3.1	1.6	10.9	385
75-99	63.4	5.5	5.5	1.2	3.1	18.1	254
SEX							
Male	63.0	6.2	5.0	5.4	2.5	17.8	1571
Female	61.8	5.7	4.9	4.3	3.3	20.0	2070
RACE							
* • • • • • • •							
White Black	65.4 44.2	5.3 9.7	4.1 11.4	4.5 6.9	2.5	17.8	3107
Other	44.0	9.0	2.0	5.0	4.8 9.0	23.0 31.0	421 100
<u>URBANISM</u>							
Central							
City	62.2	5.7	7.5	6.0	3.0	15.5	895
Suburban	65.0	4.5	4.5	3.5	2.4	20.0	1530
Non-Urban	59.0	7.8	3.6	5.5	3.6	20.5	1216
HOMEOWNERS							
OWN	70.5	5.2	4.8	3.9	2.4	13.1	2408
RENT	46.1	7.3	5.2	6.5	4.0	31.0	1211

.

Table 3- Continued

		l Valid Voters	2 Vali- dated Non- Voters	3 Self- Re- ported Mis- Report	4 Self- Re- ported Voter		6 Self- Re- ported Non Registered	N
WORK S	<u>STATUS</u>					Vote		
WORKIN UNEMPI RETIRE	-	62.9 40.1	5.7 9.0	5.4 5.0	5.2 7.2	3.3 2.3	17.5 36.5	2263
DISABI STUDEN HOUSEW	IT	69.6 60.6 58.1	5.3 6.7 3.2	3.9 4.2 4.8	3.0 4.7 3.2	2.5 2.2 3.2	15.8 21.6 27.4	644 449 62
MARITA	L STAT	US						02
MARRIE NEVER	D	68.3	5.2	3.8	4.3	2.7	15.7	2123
MARRIE DIV/SE WIDOWE	P	51.6 49.1 64.0	7.6 7.2 5.4	5.6 8.2 5.7	6.5 6.5 2.8	1.9 4.6 3.6	26.8 24.5 18.4	589 527 386
TRANSI	ENT					5.0	10.4	200
LO	0	75.9	4.0	4.1	3.1	1.9	11.0	1697
	1	52.9	6.1	5.4	6.0	3.5	19.8	833
	2	49.4	7.0	6.8	5.9	2.9	28.0	629
	3	37.6	8.5	4.7	6.3	4.7	38.2	319
HI	4	34.9	13.7	4.8	7.5	8.9	40.0	146
EDUCAT LOW	<u>ION</u> 1 2 3 4 5	47.6 44.9 54.8 63.4	9.6 8.2 6.9 4.7	2.5 4.6 4.7 7.1	4.2 3.4 4.7 5.5	5.4 4.8 3.5 2.6	30.6 34.1 25.5 16.7	353 414 943 621
HIGH	6	70.7 82.6	5.2 3.2	5.9 4.2	5.1 5.2	2.1 .7	11.0 4.2	574 696
<u>SOCIAL</u>	CLASS							
LO	3	51.6 48.8 64.4 69.7 77.5	7.3 9.3 7.2 4.8 5.0	5.3 4.7 5.3 4.9	5.1 7.0 5.3 4.3 2.5	3.9 2.3 2.5 2.2 7.5	26.7 27.9 15.3 14.1 7.5	1958 43 320 1271 40
HI		79.1	1.9	4.6	4.9	1.9	7.6	368

Table 3 (Continued)

		1 Valid Voters	2 Vali-	STERED 3 Self- Re- ported Mis- Report	4 Self- Re- ported Voter	5 Self- Re- ported Reg/ No	SISTERED 6 Self- Re- ported Non Registered	N
INCOM	<u>1E</u>					Vote		
PERCE 0-168 17-33 34-67 68-94 96-10	28 18 18	41.7 56.5 62.8 73.5 84.9	8.8 8.1 5.9 3.4 2.1	4.0 5.0 5.3 5.6 4.1	4.0 4.4 5.3 4.7 4.1	5.7 3.2 3.0 1.5 .7	35.8 22.8 17.7 11.3 4.1	547 496 1184 952 146
PARTI	SANSHI	2						
IND-A LEAN- WEAK- STRON	IND	41.6 57.7 61.0 75.4	7.2 5.7 6.7 4.7	3.5 5.1 5.2 5.0	4.5 4.6 4.8 5.0	4.2 3.3 3.3 1.6	38.8 23.6 19.0 8.2	425 877 1217 1109
POL I	NTERESI	2						
NOT M SOME VERY		37.5 63.4	10.5 5.6	3.1 5.6	3.5 5.5	6.5 2.3	38.9 17.5	819 1712
INTER	ESTED	79.3	3.0	5.4	4.5	1.3	6.5	1094
PARTI	CIPATIC	N						
LO HIGH	1 2 3 4 5	53.1 73.2 80.9 81.0 92.3	7 4.0 2.8 1.7 1.9	4.5 5.8 3.5 8.6 5.8	4.6 4.6 7.1 4.3	4.7 2.2 .7 .9	36.1 10.1 4.9 2.6	2168 948 283 116 <2 0
MEDIA	EXPOSU	RE						
LO	1 2 3 4	39.1 56.8 68.5 81.9	7.1 5.1	3.9 4.5 5.9 8.6	3.8 5.4 4.4 4.7	5.6 2.9 .6 1.5	38.2 23.4 14.4 10.3	665 851 994 751
HIGH	5	79.8	1.8	3.9	6.5	1.8	6.2	337

TABLE 4- DISTRIBUTION OF NON-VOTERS

	Self-Reported <u>Non-Voter</u>	Self-Reported <u>Not Found</u>	Registered Voto <u>Misreported</u>	ers _ <u>N</u>
OVERALL AGE				
17-24	84.4	7.6	8.0	262
25-34	78.7	11.0	10.2	263
35-44	65.4	17.7	17.0	381
45-54	68.6	15.3	16.1	283
55-64	68.9	15.6	15.6	137 135
65-74	72.0	11.8	16.1	93
75-99	82.5	3.1	14.4	93 97
<u>SEX</u>				
N/- 7				
Male	73.4	13.5	13.1	594
Female	76.1	11.1	12.8	800
RACE				
White	76.0	12.3	11.7	1095
Black	67.7	11.9	20.4	235
Other	87.7	8.8	3.5	57
URBANISM				
Central	65.5	15.2	19.3	348
Suburban	77.7	9.5	12.8	540
Non-Urban	78.6	12.8	8.6	509
HOMEOWNERS				
Own	71.4	12.6	16.0	722
Rent	78.9	11.6	9.5	732 662
WORK STATUS				
Working now	72.5	13.2	14.3	856
Unemployed/		11.9	8.2	134
Laid off	78.9	8.5	12.6	199
Retired/Dis Student/		11.9	10.7	177
Housewife	-	-	-	27
				_ ·

	Т	ABLE 4- (CONTIN	UED)	
	Self-Reported <u>Non-Voter</u>	Self-Reported <u>Not Found</u>	Registered Vote <u>Misreported</u>	rs <u>N</u>
MARITAL STATUS				
Married Never Married Div/Seperated Widowed	75.4 75.7 72.2 76.9	12.7 12.8 12.3 7.7	11.9 11.5 15.5 15.5	679 288 277 143
TRANSIENT				
0 1 2 3 4	70.5 73.0 76.2 83.2 81.6	12.8 13.9 10.5 9.4 11.2	16.7 13.0 13.3 7.4 7.1	
EDUCATION				
Low	87.7 85.7 79.8 66.4 64.7	7.5 6.1 10.1 14.7 15.3	4.8 8.2 10.1 19.0 20.0	187 231 435 232 170
High	48.4	27.9	23.8	122
SOCIAL CLASS Lo 1 2 3 4 5 Hi 6	79.1 71.6 71.0 55.1	10.1 13.8 13.4 23.1	10.8 14.7 15.7 21.8	712 22 116 396 9 78
INCOME				
Percentile 0-16% 17-33% 34-67% 68-94% 96-100%	86.3 78.7 72.7 62.4	6.9 10.0 13.3 16.9	6.9 11.3 14.0 20.8	321 221 451 255 22
PARTISANSHIP				
Indep or APOL Leaning Ind. Weak Strong	87.3 77.8 75.1 59.8	6.8 10.3 11.8 19.9	6.0 11.9 13.1 20.3	

TABLE 4- (CONTINUED)

	Self-Reported <u>Non-Voter</u>	l Self-Reported <u>Not Found</u>	Registered Voter: <u>Misreported N</u>	5
POL INTEREST				
Not Much Some Interest Very Much	89.8 70.5 53.3	5.4 14.4 20.7	4.8 15.0 26.0	521 638 227