DISCUSSION

Morris H. Hansen, Westat, Inc.

1. Introduction

I wish to commend the authors of the two papers for an excellent and straightforward job of presentation and analysis of their information. They have both come to the conclusion, with persuasive but not conclusive supporting evidence, that the problem was changes in preferences in the last few days before the election. They had hoped to protect themselves against changes by polling up to within a few days of the election date, and on the assumption that decisions would have been made by then. They report considerable evidence from their data favoring their conclusions. It seems to me they provide relevant supporting evidence but not proof. Some questions remain -- there are other plausible hypotheses.

I commend both papers for their frankness and and the quality of their reporting. I especially commend Mitofsky for going to considerable limits in reporting some details of what are generally methodologically superior results, such as showing the effect of weighting by probabilities of selection within households, and of post-stratification by demographic characteristics. His results are interesting because each of these steps is methodologically sound, even though it actually made the results less accurate in this particular case. It is important not to misinterpret the implications and infer that such methods should not be applied. The trouble is that the election is not a valid test of survey methodology, since prediction is involved.

2. Too Much Accuracy Has Been Claimed for the Polls

Too much has been claimed for the accuracy of the polls on the basis of recent performance -- in my judgment more than is supported by the evidence. It seems to me healthy to receive a setback that may force fuller recognition of error levels that should be expected, based on the available evidence, and that at the same time may stimulate additional research and possibly improvements in both methods and claims.

(a) Sampling and Response Variability

With respect to levels of errors that may be expected from the polls, just from sampling and response variability, I feel that there may be some confusion concerning the ranges of errors expected, and this should be carefully clarified by the polling organizations themselves, and especially in the press.

For example, Kohut cites (in his Table 1) the results of the Gallup Poll compared to actual election results for the past six presidential elections. He then states that "the average observed difference has been 1.3 percent, and in the light of this performance it is no small wonder that 1980 came as a shock to poll watchers." I don't understand the surprise, based on the same evidence. The estimated 95 percent confidence limits for 1980, based on the estimated variance from the prior six elections and the Gallup estimate for 1980, span the actual election results. Thus, based on recent observed performance, there should be no reason for surprise.

Mitofsky's results do imply a statistically significant difference of about four percentage points (after a rough adjustment for the four percent undecided). A 95 percent confidence limit of about 2.2 percent can be computed from his data. However, potential errors in estimated turnout and in voting performance as well as nonresponse problems and other sources of error seem to make it dubious to anticipate accuracy as measured only by the computed sampling error (which measures only sampling variability and some of the response variability, and makes no allowance for reporting or prediction biases). On such evidence it is my judgment that the pollsters should not be surprised, and therefore should have warned the public to expect such differences.

Another related point deserves some attention. The polling organizations seem to claim a range of error for election predictions based on variance computations from the survey or prior experience, or both, on the order of one to three percentage points. These, it seems clear, are appropriate for the error in the percentage for one of the candidates, say Reagan's percentage, and not for the difference between the percentages for Reagan and Carter.

If Reagan's percentage is subject to a standard error of about two percentage points, and Carter's is about the same, then the standard error of Reagan's percentage minus Carter's is about four percentage points. The standard error of the difference is very close to the sum of the two because of the near-perfect negative correlation between these two percentages in a sample estimate: if one is too high then the other is too low. (The correlation would be -1 if there were no third-party vote, and in this case the standard error of the difference is exactly the sum of the two standard errors.) I get the impression that this has not been made clear to the press or the public.

A misunderstanding of this negative correlation is shown by Stacks in his *Time* article,* and a resulting misinterpretation, where he says "The public opinion industry has christened Caddell's thesis the 'big bang' theory of the campaign: eight million voters moving to Reagan in 48 hours." Stacks failed to understand that a shift of four million out of 80 million (or of five percentage points) from Carter to Reagan changes the difference between their percentages from 0 to 10. If the effect of the high negative correlation is recognized, it makes the assumption of last-minute shifts more plausible. This simple relationship is know by the polling organizations, but needs emphasis in interpreting the polls and the potential of their sampling and other errors.

(b) Response Errors: Telephone Versus Personal Interviewing

Polls (and other surveys) have substantially changed their procedures in recent years due to the advent of random digit dialing together with telephone interviewing.

I see nothing wrong with this development so long as it is executed and reported carefully. Much greater control with central telephoning is feasible (and I hope is being executed) as compared with field interviewing. In my judgment such increased control can reasonably offset the loss of no-telephone households, although it is important to report this undercoverage. A post-stratification adjustment (by educational level, for example) presumably can partially correct for but cannot fully offset this loss.

Gallup's case for the secret ballot is interesting and adds an important argument for personal interviewing. However, I question if this advantage is sufficient to offset the necessarily lower level of control over personal interviewing in the polls. Moreover, I am not aware of a serious effort at quality control of field interviews. If the secret ballot is worthwhile it should have equivalent advantages in the continuing attitude surveys. It is not clear to me if it is used there. Presumably an alternative procedure could be the use of randomized response to achieve anonymity. Much has been said about the potential advantages of randomized response to achieve similar kinds of gains. However, in my uninformed judgment, a secret ballot will be less expensive (especially because of reduced variance and will be trusted just about as much as a randomized response approach by a respondent who may or may not have confidence in either, or who may be embarrassed in reporting opinions or voting behavior to a field interviewer.

There are those that suggest that the less personal contact by telephone may have some of the same effect, that is, that telephone interviewing should reduce the tendency to report what is an unpopular opinion or position. Much more needs to be learned, and again, maybe the 1980 election discrepancies will help the recognition that we know much too little. Increased research is needed. It should be supported and executed in many different ways.

The problem of more or less systematic response errors in surveys is serious for factual items, and presumably more so in attitudinal items such as predicted voting behavior.

An illustration, to emphasize the potential problem, is the clearly factual question in a post-election survey: Did you vote in the recent election? Every two years the Census Bureau, in a supplement to the Current Population Survey, shortly after the election asks a question equivalent to "Did you vote in the recent election?" They get about a 10 percent overestimate of the actual voting. There are many potential reasons, including the fact that they do not necessarily interview the respondent himself. Nevertheless, the discrepancy is relatively large on a clearly factual question. The problems and potential biases are far greater on less clearly defined questions such as expectation to vote in a forthcoming election a few weeks or possibly just a few days away.

A result reported by Kohut is relevant and interesting in this respect. If I interpret him correctly, he reports, from earlier studies involving a follow-up to voting registration records for a subsample, that about 78 percent of those that claim they will vote shortly before an election actually do vote. From this it is obvious that there is much prediction instead of simple measurement in the pre-election polls.

(c) Nonresponse

I am concerned also about nonresponse. The nonresponse rates have not been clearly reported along with their explicit treatment, or their implicit treatment in estimation. They should be.

3. Reporting and the Media

The media is a major problem. The pollsters seem to be willing to accept and respond to what the media want, oversimplification. Of course the reports from the polls should be as simple and straightforward as feasible, but with clear reporting of such things as nonresponse rates and imputation for nonresponse, and the quality of measurements as evaluated from earlier experience (an excellent illustration is the 78 percent figure as cited from Kohut earlier and the results of the post-election survey report by Mitofsky). If the media want to oversimplify and risk misrepresentation they should not be aided and abetted by the polling organizations.

In this respect I am especially concerned by what seems to be the increasing emphasis on being first, and lowering quality in order to do so, at increased risk of error. It is well known that it is easy to be right most of the time, with quite loose methods. Thus a record of being reasonably close to right over a half-dozen experiences is hardly relevant. Of course it is better than being wrong in those experiences, but far from proof of future correctness. The best way to attempt to assure future correctness is high standards of performance. With the problems of measurement in the social sciences being what they are today there will still be substantial risks. These should be identified and reported to the extent feasible.

4. Final Remarks

Finally, the pre-election polls should not be represented as providing an evaluation of factual measurement in surveys through comparison with actual performance. They provide a test of predictions based on survey results. Valid tests of factual measurement in surveys can be done in post-election polls.

I hope the discrepancies in the pre-election predictions from actual 1980 election results will result in additional attention to some of the above problems and issues. If some additional progress is made the 1980 experience will have served an exceedingly useful purpose.

---

*John F. Stacks. "Where the Polls Went Wrong," Time, December 1, 1980, pp. 21-22.