

PRESIDENTIAL ADDRESS METHODS AND STANDARDS: A CHALLENGE FOR CHANGE

WARREN J. MITOFSKY

Ever since I joined AAPOR, some 25 years ago, I have been struck by the tension that seems to exist between a relatively small group of survey practitioners on one side and more formally trained methodologists on the other. But it is easier to describe the conflict than the participants.

It is a conflict between the old and the new. It concerns methods and practices that were popular during another era, but they are not for today. It permeates conference sessions at AAPOR. It affects the things we say to each other about our colleagues. It even has found its way into other AAPOR presidential addresses.

For a while I thought the resistance to progress came mostly from older practitioners and those in the commercial sector of our organization. But it does not. Some are in institutions of higher learning. Many have been in the field since the early days of AAPOR. But there are younger people in their ranks also.

All of the people on both sides of the conflict have a deep commitment to survey research. And all share a love for AAPOR. Yet they take their positions at the barricades, like combatants at the Alamo, ready to capture or defend it, every time standards and practices are discussed.

I don't want to characterize this tension as a dominant theme of our meetings. That would be far from an accurate picture. The tension is more of a persistent undercurrent that keeps surfacing. I don't know if this same tension exists in other social science associations. But I do know we experience it at AAPOR.

For my own part, few who know me can have any doubt on which side of this controversy I stand. To me, the issue is whether we benefit from methodological advances, or whether we persist in using tools that were state-of-the-art too many years ago.

Would our public poll mentors—men such as George Gallup, Elmo Roper, Archibald Crossley—if they were alive and working today, still be doing what they did in days gone by, or would they adapt to the current methodological age? Of course these men of vision would be

leading the way. I cannot imagine them stuck in the past. It is these men who formed the National Council on Public Polls. They did it partly because of the conflict within AAPOR over formulating standards and partly to head off government regulation of polling. This new organization was supposed to be an instrument for self-policing of the polls, and to educate the news media and the public about polls.

And yet AAPOR today still has no *unambiguous* mechanism in its Code of Professional Ethics and Practices for condemning polls that are outright garbage, let alone so-called legitimate work that cuts corners and lacks scientific rigor.

In the sample design area I am referring specifically to surveys with known biases in sample selection and estimation, surveys that still rely on quota methods at some stage of the design rather than probabilities at all stages, or surveys that ignore unequal selection probabilities when it comes to estimating. We should also include surveys based on biased samples selected only from listed phone numbers.

Usually there is an attempt at legitimizing quota sampling. It is described in reports as a “modified probability” design. (I must tell you, every time I hear that phrase “modified probability,” I am reminded of the expression “a little pregnant.”)

And taking account of unequal sample selection probabilities, such as those we get when we select one person per household in our sample, requires weighting if we are to avoid a bias. But in some eyes weighting, I am told, is “cooking the data.” Any discussion of reducing the sampling error by use of ratio or regression estimates I suspect would move us, in the eyes of those who object, from a venial sin and punishment in purgatory to a mortal sin and condemnation to a survey hell.

There are techniques used for asking questions that introduce response bias. This seldom gets the same attention as sample design issues. I am thinking of survey questions that purport to set the agenda for the respondents’ frame of reference by providing background information on a complex or unfamiliar subject. Or questions that are deliberately given a context by placing them after other subject matter in the survey. An example is the intentional placement of presidential preference questions near the end of a questionnaire. Presumably, the issue questions that precede it are a mini-agenda for the campaign, which cannot help but bias the presidential choice.

Fifteen to 20 years ago, the more common method of sampling and interviewing switched from in-home and face-to-face to telephone. When this happened there were speeches by elder statesmen to the press and to marketing directors saying that they could not rely on interviews conducted by telephone. They gave no credence to telephone interviewing’s advantages of direct supervision of interviewers

or the ability to select unbiased probability samples of the telephone household population.

The methodological issues we dispute today are not new. They were with us in 1946 at Central City and at the first AAPOR meeting at Williamstown a year later.

At Central City¹ there was a session devoted to “probability” versus “quota” sampling. Morris Hansen of the Census Bureau was a lonely advocate of probability sampling. Representing nonprobability sampling, or quota sampling, were Norman Meier of the University of Iowa and a consultant to the Iowa Poll, Lucien Warner of *Life* magazine research, and Elmo Wilson, director of research for the Columbia Broadcasting System. The argument was mainly over the greater cost of the new method. At the conclusion of the discussion there was no consensus. There was no meeting of the minds among the participants.

Another panel discussion at Central City included some of the most distinguished researchers of the day: George Gallup, Julian Woodward, Clyde Hart, and Harry Field, discussing “Technical and Ethical Standards in Public Opinion Research.” Woodward argued that the people who do polls “must conduct themselves in such a way as to justify the responsibilities which will increasingly be theirs and to deserve the respect with which the public will regard them.” He was talking mostly about public polls, but others at the meeting quickly added market research to the discussion. Woodward saw polls as a “public utility.” He was concerned that cost alone should not deter improvement in methods. He wanted “a committee on standards that would ensure that no poll that did not live up to the standards agreed upon should have membership in the association. Such standards would include adequate sampling methods, competent interviewing staff, well-designed questionnaires, and the observance of certain ethical as well as technical standards.”

This set the stage for what Paul Sheatsley² describes as the “ensuing tensions between [the] academic and commercial constituents.” He describes the position of the “protagonists” in the only debate that took place during the adoption of AAPOR’s constitution. It was over standards: “the commercial researchers . . . were on guard against any attempts by academics or others to restrict their research freedom.”

From the history, it is clear that the debate over the scientific integrity of methods used in the conduct of surveys is at least as old as this organization. Whatever mechanism for improving standards that was

1. As described by Don Cahalan, “Origins: The Central City Conference,” in the forthcoming *AAPOR History*.

2. The following discussion of Central City and Williamstown is adapted from Paul Sheatsley, “The Founding of AAPOR,” in the forthcoming *AAPOR History*.

envisioned but not enacted at Williamstown still remains nonexistent. The conflict is unresolved today. Thus, the tension remains.

For many years I was an advocate of more severe performance standards. I wanted a code that would define the parameters of a scientific poll and standards that would prohibit biased techniques. These standards would allow us to answer a journalist's inquiry about the quality of someone's poll with a forthright declaration. At last we could say: "That poll was not performed in a manner that meets AAPOR's standards." If we have criteria for evaluating the quality of a poll, then the press and other consumers of public opinion research would be able to distinguish research of high quality from survey salesmanship.

With great reluctance, I have changed my mind. As much as I believe that in most cases it is possible to write performance standards for specific survey research projects, I no longer believe that it is possible to codify performance criteria that are general enough for most needs.

For example, in my own work at CBS News, under most circumstances I could not imagine sampling precincts for the purpose of making an election projection by any means other than random selection of precincts where the probabilities are used in the estimator. This was the method used when we selected a sample in the Philippines for the Marcos election in 1986, just as it has been for every state sample in the United States that we have selected during the last 22 years.

However, had I been in Panama two weeks ago, I could have envisioned a circumstance where probability sampling would not have been feasible, but a quota sample of voting precincts would have made sense. I could have had very useful information about fraud in the election even with a purposive sample. I could have obtained the vote counts tabulated by the local election officials and compared them to the counts released by General Noriega's government. I would not have been able to quantify the election result or make an estimate of the magnitude of the fraud, nor could I have specified a sampling error for the data I did have. But I could have shown a systematic distortion of the counting in the precincts I selected. If most differences had favored the Noriega-backed candidate, it would seem reasonable to conclude that there had been election fraud.

A different example of an unscientific but informative exit poll occurred in Moscow a few weeks ago. The magnitude of the answers to some of the opinion questions showed a level of feeling that was a clear indication of strong voter sentiment. This conclusion was responsible even though there was no measure of precision possible for the results. Moreover, the fact that Muscovites answered the questions was as important as any substantive finding.

What I have tried to do with these slightly atypical examples is make

it clear that different surveys have different purposes. Defining standard methodological practices when the purpose of the survey is unknown does not seem practical. Some surveys are conducted under circumstances that make probability methods impossible. These special circumstances require caution against unjustified or unwarranted conclusions, but frequently legitimate conclusions are possible and sometimes those conclusions are important.

Several years ago, an American Statistical Association review of survey practices by Bailar and Lanphier (1978) contained many useful conclusions. It was based on a survey of surveys and was about the way they were conducted. It identified both good and bad survey practices. One of the conclusions in the report was that all surveys should be based on probability selection methods. It said this even though the surveys examined for the purpose of making this recommendation did not constitute a probability sample of all surveys.

Now I am not suggesting some rationalization for conducting non-probability surveys for most of the situations public pollsters deal with. If CBS News or some other organization were conducting a national survey of public opinion in the United States, and there was a reasonable amount of time to conduct the survey, and the results were to be reported as representing the views of all adult Americans, I still think a probability selection at all stages of the design is the *only* reasonable practice.

But I am now willing to accept alternative methods, methods that I personally would not use under the circumstances I just described. Even methods that many people may consider biased. I think it is reasonable for a researcher to conduct his or her research using any design that fits the problem, provided there is proper disclosure of their methods and its limitations.

But now we have come to the catch. To do this, I believe we need to change the disclosure requirements contained in the AAPOR Code. The code is wholly inadequate as it is now written. Let me explain.

The preamble to the code says in part: "Our goals are to support sound and ethical practice in the conduct of public opinion research." I emphasize the word "sound." It then says in the code under "Principles and Professional Practice in the Conduct of Our Work," "we shall exercise due care" in our designs so we can "assure the reliability and validity of results." The word I want to highlight here is "validity." The code goes on to say, we will not produce "misleading conclusions," or "knowingly make interpretations of research results . . . that are inconsistent with the data available."

In section three, the code calls only for minimal disclosure. It asks for identification of the sponsor, question wording, dates and method of interviewing, and details about the sample design and its precision.

My interpretation of these words, taken as a whole, is that the code, including the preamble, is ambiguous. It is quite rigid in its insistence on the use of sound scientific methods and that we not mislead the public. I assume that means we should not mislead each other also. But then it is quite weak when it asks for only minimal disclosure. It implies that a researcher can be relieved of the burden of sound scientific methods as long as there is disclosure conforming to the minimal requirements of the code.

It is not satisfactory, in my view, for the researcher just to disclose the items that are called for in AAPOR's "Standard for Minimal Disclosure." That information is not informative enough for most consumers of survey research, certainly not for most members of the news media and other members of the public. It is barely useful to trained survey researchers. What we have done with our minimal disclosure is to place the burden on our users for making sense of the limitations of our surveys. We have provided users with a few technical details called for by the code and then we tell them either to hire a professor to interpret these details or get trained themselves in survey research.

That is wrong. It is the responsibility of the researcher to spell out the limitations of his or her own survey when there are limitations. That is the clear implication of the section on principles of professional practice. If the disclosure section were consistent with the requirements contained in the section on the principles of professional practice, we would disclose much more than is presently called for in the disclosure section.

Having said this, I want to propose a more meaningful set of disclosure principles. My interpretation of proper disclosure would still permit a researcher to tailor a survey to the purpose at hand. It also would obviate the need for performance standards. The disclosure principles I am about to propose would shed light on the interpretation of the results. It would not permit surveys to be cloaked in the pseudo-scientific aura bestowed by following the present disclosure standards. And finally, I believe this proposal is more consistent with the purposes spelled out in the preamble and the professional practices section of the AAPOR Code.

Along with reporting the items contained in the minimal disclosure section, the researcher should go further. If there are biases in the research design—for good reason or not—the researcher conducting the study will be the one to make these design flaws and their consequences known to the consumer of the research. This to me is proper disclosure. I would like to repeat what I just said: If there are biases in the research design—for good reason or not—the researcher conducting the study will be the one to make these design flaws and their consequences known to the consumer of the research.

If there is a sample selection bias it should be disclosed. If there is a weighting bias it should be disclosed. If there is a question order effect known to the researcher it should be disclosed. And furthermore, the limitation of any conclusions should be spelled out if there are such biases.

It is absolutely the obligation of the researcher to spell out the limitations of his or her survey. Let me be even clearer about what I want to see disclosed. When I say limitations, I want to know how the design affects conclusions based on that survey. Specific statements about what can and cannot be concluded should be part of the disclosure.

For example, under my proposal, nonprobability samples are all right, provided the bias of the method is properly highlighted and there are proper cautions against generalizations from the sample to some larger population. If there are no call-backs or if only respondents at home are interviewed, the limitations they impose on the interpretation of these results should be made clear.

If weighting for the probabilities of selecting one adult in a household is not part of the estimation procedure then I want to be told in the disclosure statement about any relationship between household size and characteristics correlated with this variable. I should also be told that people in households with more than one adult are underrepresented due to the lack of weighting for this selection probability.

If the sampling error is not reported, then the disclosure statement should have different information. It should say that the sampling error is not reported either because the design is not random at all stages and therefore a sampling error computation is inappropriate, or the researcher thinks reporting sampling error would be misleading, given all the other possible sources of survey error. Either would be all right, provided the disclosure statement made it clear that there can be no justifiable conclusions drawn from the survey about any change over time in any characteristic obtained in the survey, or that there could be no meaningful interpretation about the differences between two subgroups in the same survey on some opinion item.

If there are question biases I want to know what they are. In surveys about racial attitudes, I want to know whether there is an interaction between the race of interviewers and respondents and whether the conclusions are affected by this interaction.

To me, this type of disclosure would come closer to "truth in packaging" of survey research. What we have now does not do it. This type of disclosure would reduce the tension about methods. I want to conclude by offering another proposal that could reduce tensions and shed light on the various reasons for using alternative survey methods. Rather than ignoring our differences, or having one-sided discussions about them, we should bring them into an open forum for public discus-

sion. My proposal is for a series of debates about methods to take place at our annual conferences. Last year at AAPOR, we had a public discussion about the reporting of sampling error when a survey is made public. It was a useful beginning. However, I would like to take it a step or two further.

These debates should have formal rules. Both sides should be prepared to produce a written report of the arguments and counter arguments. These reports could be published in *Public Opinion Quarterly* or issued as separate papers by AAPOR. I can imagine lively discussions, useful to the participants, to AAPOR members, and perhaps even beneficial to the consumers of our work. I would even hope for the amelioration of the issues. Perhaps, the reduction of tensions.

I offer these thoughts today in the hope of improving the quality of work produced by members of our association. In recent years the credibility of survey research has been attacked. Most of the work by members of this association has been of high quality, professionally conducted and accurately reported. But just often enough, the caliber of the surveys reported to the public is suspect. I know raising standards involves increased cost, but I believe we must pay the price. If polls are a public utility, as Julian Woodward said, then we owe it to the public as well as ourselves to raise the level of our performance. And we must do it through our own initiative, and not because of pressure from outside.

Reference

- Bailar, Barbara A., and C. Michael Lanphier (1978)
Development of Survey Methods to Assess Survey Practices. Washington:
American Statistical Association.