

## *The Case For Publishing (Some) Online Polls*

by Humphrey Taylor

Although many of the world's media regularly publish online polls which are not based on probability sampling, there are still some media, including *The Polling Report*, which are reluctant to do so.

The three main arguments heard most often against the use of online polls are:

- Traditional telephone polls are “scientific” and have a theoretical basis behind their methodology; online polls based on non-probability sampling are “not scientific” and do not have theoretical underpinnings.
- Online polls are biased because the online population is not a representative cross section of all adults because not everyone can be represented in an online poll.
- Online surveys based on non-probability samples have not yet established a track record that demonstrates their reliability.

In my opinion, none of them stand up to careful scrutiny, when considering some—but not all—online polls.

### **What Is “Scientific”?**

The belief that traditional telephone polls are “scientific” depends on your definition of science. Typical media telephone polls have a response rate of 20% or less. Many of the leading peer-reviewed academic journals will not accept papers based on surveys for publication unless they have a response rate of 50% or more. Forty years ago an “acceptable” response rate for many journals was over 70%, and even then there were examples of significant non-response bias related to the availability of the respondents. Using the 50% threshold would mean that the media should not publish, and we should not believe, any of the polls that are now conducted for, and published in, the media. Of course, that's nonsense. Not to publish these polls would deprive the public and our leaders of important, and we believe reliable, information. The main reason that we should trust traditional in-person and telephone polls (when conducted by serious polling organizations) is not that they are *theoretically* “scientific” but that their track record of predicting elections and other phenomena is good. If well-conducted telephone and in-person surveys are “scientific” it is because they make predictions which can be validated.

Newton had no theory that explained gravity or that justified his “laws” of gravity, dynamics or optics. But they came to be widely accepted because they worked in practice. They made predictions which could be validated (although we now know they were not perfectly accurate).

Science provides thousands of similar examples. Aspirin was widely used for many years to reduce pain even though there was no scientific theory to explain how it did this. When the FDA approves new drugs today, it relies on clinical trials that measure efficacy and safety not on any theory (if there is one) of why the drug works. A drug *must be shown to work in practice, not just in theory*. The same should apply to polls.

If polls had not demonstrated their reliability by predicting elections with a considerable degree of accuracy, we would not and should not trust them. Common sense suggests that we should continue to trust them as long as their track record continues to be good—but not a moment longer. In other words, the trust we have in opinion polls and the different methods they use (whether in person, telephone or online) should be based on empirical evidence of their track record. This, of course, is the scientific method.

In addition to low response rates there are several other reasons why it is misleading to describe most telephone polls as “scientific.” If you're picking otherwise identical black and white balls out of a bag,

sampling error is probably the only source of error. When you are measuring opinions, or the propensity to vote one way or another, there are several other sources of error which we cannot quantify, including interviewer effects, question wording, question order, and the ability and willingness of respondents to answer accurately and honestly. Furthermore, the exclusion from telephone polls of those with no telephones and those with only cell phones (a rapidly growing segment of the population in many countries including, to a lesser extent, the USA) means that the sampling frame is far from perfect.

This is why I avoid using the words “margin of error” which suggests that this is something we can compute. Statements such as “the margin of error in this survey is plus or minus 3 percent” are dangerously misleading as they suggest that we can calculate the maximum possible error from *all* sources of error, which of course is impossible.

If there was a truly “scientific” method of designing a telephone poll, one would expect all telephone polling organizations to use it. In fact, as others have noted, the design of polls is as much an art as a science. In 1997 I surveyed 83 of the world’s leading opinion research organizations, including ten in the United States, and found that no two firms used the same methods. Even in this country there were substantial differences between the ways polling firms drew their samples, selected respondents, called households back, substituted (or not) for individuals not reached, designated likely voters and weighted their data.

### **The Online Population Is Not Representative**

One of the more widely held criticisms of online polling is that the population online is not representative of the total adult population. But, neither is the population with landlines who are surveyed in telephone surveys. In a growing number of countries, the percentage of the population who are online exceeds the percentage of the population with landline telephones. Evidence suggests that this may now be true of U.S. college students.

Over the last seven years, Harris Interactive has run hundreds of parallel telephone surveys using RDD in tandem with online surveys of members of our panel of about six million respondents. The raw data from both types of polls differs significantly from the total population. Both our telephone and our online samples need to be weighted quite substantially to make them representative. The issue we address with both our online and our telephone polls is not whether the raw data are a reliable cross-section (we know they are not) but whether we understand the biases well enough to be able to correct them and make the weighted data representative. Both telephone polls and online polls should be judged by the reliability of their weighted data.

### **The Track Record**

So, if polling methods should be judged on their track record, what is the track record of polls using non-probability sampling? Before telephone polling became widespread, most polls in most countries were conducted in person using quota sampling, not probability sampling. In countries with low telephone penetration, most polls still do so. North America was very unusual in relying on probability sampling for its in-person surveys. The reason why polls based on quota sampling were, and continue to be, believed and published is their successful track record.

When I worked in Britain in the 1960s and 1970s, there were seven regularly published national polls. Two of them used probability sampling and the rest used quota sampling. Their track records on predicting elections were pretty good, but on balance the polls based on quota sampling did somewhat better than those based on probability sampling. Furthermore, the average error in the British polls, based on quota sampling, between 1945 and 1975 was somewhat smaller than the average error in the American polls based on probability sampling. Interestingly, American media have not hesitated to report the results of foreign surveys based on quota sampling. And, given their track record, they are right to do so.

If the media wish to make decisions on which polls to publish based on some measure of quality and reliability, they should focus primarily on the track record of the polling organizations and the methods they use. There is obviously a strong case for not publishing polls by organizations using methods that have a bad track record. Where insufficient data on track records exists, it is reasonable to suspend judgment. But where the track record is good, decisions not to publish poll data are a form of censorship that prevents the public and decision makers from having access to interesting and sometimes important measures of public opinion.

One problem is that there are some online polls that use methods for which there is little or no track record or where the track record is not very good. The same can be said for some telephone polls.

## The Track Record of Harris Interactive's Online Polls

Harris Interactive has used its current methodologies to predict 78 elections starting with the American elections in 2000. We believe the results validate our methodology.

In 2000, our national online survey of the presidential vote was actually more accurate than that of any of the telephone surveys. However, the results of one election alone can be based in part on luck, and we were probably a little lucky. For this reason, in the 2000 elections we also conducted surveys in 36 states and forecast the results of all the 71 races for the presidential vote, gubernatorial vote and Senate vote in these 36 states. Overall, our results, while not perfect, were substantially better than the average for the final telephone polls in these elections. (For full details see *International Journal of Market Research*, Vol. 43, Quarter 2, 2001.)

In the 2004 elections, the final prediction in our online poll of the presidential vote was not as accurate. Our average error on the two main candidates was 2.5 percentage points, and subsequent internal research suggested that this was not directly related to the mode the surveys were conducted in. Whatever the reason, an error of 2.5 percentage points on any number is rarely enough to change the outcome of polls presenting public opinion data, although in this close race it was enough to show Senator Kerry slightly ahead of President Bush. In addition we used online polling to generate predictions for the presidential vote in three states. Our average error on the votes for the two main candidates in these three states was 2.7 percentage points.

Our most recent, and only other, use of our online methodology to predict elections was in the British general election of 2005. We made two forecasts: one for Great Britain as a whole and one for Scotland. The average error on the three main parties' share of the vote in both these predictions was 0.8 percentage points. These results compared very favorably with the final predictions of telephone polls, only one of which was more accurate.

In 58 of the 78 races we have covered there were also telephone polls conducted just before the elections. Our average error on the spread between the two main candidates or parties was 3.3 percentage points compared to 4 percentage points in the telephone polls.

There are other ways of validating the reliability of polling methods. Harris Interactive, like all major polling firms, measures ratings of the President and (since 2003) attitudes to events in Iraq. If our trend data are shown on a chart over several years compared to the data provided by telephone polls, an observer could not tell which polls used which methods.

## Some Other Points

Comparisons of the many hundreds of parallel online and telephone surveys we have conducted show some clear and systemic differences. However, these appear to be method effects or mode effects rather than sampling effects. One effect is related to the respondents' *reading* questions rather than *hearing* them. This influences responses to scales and how many people give "not sure" or "don't know" as an answer. However, there is no hard evidence that the online/written surveys are more or less accurate than the telephone/verbal surveys.

The second systemic difference relates to questions where respondents may be uncomfortable or embarrassed if they give an honest answer to a live interviewer. Where members of our panel answer questions about sexual orientation, church going, belief in God, drinking, giving to charity or other topics where there is a "socially desirable" answer, substantially more people in our online surveys give the "socially undesirable" response. We believe that this is because online respondents give more truthful responses than telephone respondents, and is not due to sampling differences. (However it has been suggested that gays, atheists and agnostics, drinkers, etc., rush to join our panel!) For example, we believe that the 6% of respondents who self-identify as gay, lesbian or bisexual in our online surveys is a more accurate number than the 2% who do so in our telephone surveys (even though we have no way of knowing what the correct percentage is).

Fortunately, the great majority of the media publish our online polls and we are proud to conduct online polls for some of the world's most prestigious media here and in Europe, who have been convinced of the validity of our surveys. It is unfortunate that *The Polling Report* does not and we hope it will not be long before it and all the American media will publish them whenever they find the results interesting or important. ■