Horses for Courses: How Different Countries Measure Public Opinion in Very Different Ways

By Humphrey Taylor

I have just completed a small survey, or attempted census, of the leading survey research firms in 13 countries including North America, Europe, Australia, Japan and South Africa. It compares and contrasts the different methods researchers use to measure public opinion and forecast elections, and finds huge differences. Survey methods which are standard practice in some countries would be regarded as gross malpractice in others.

Having completed this survey, I argue that national differences are inevitable and probably desirable. What works well in one country does not necessarily work well in others. Having said this, I also insist that many of the differences between countries are undesirable. Research firms everywhere could and should learn from survey firms abroad much more than they do, to improve the quality of their work.

Background

When I joined the research industry in Britain in 1963 there were two regularly published national public opinion polls, Gallup and NOP. Both used the same quota sampling methodology. When I assumed responsibility for NOP's polls for the Daily Mail and other media, I don't think I seriously questioned our methods—what we did or why we did it. I just continued doing what we had done before.

Similarly, most of the marketing research work I have done over 30 years in the industry has used tried and trusted methodologies with little questioning of their validity. Much effort and attention was directed to making the research more useful to the client but relatively little to improving the quality and reliability of the data.

In so far as we did worry about the quality and accuracy of our samples, questionnaire design, fieldwork and weighting, we did so in a uniquely British context. We sometimes looked to other countries, especially to the United States, for new research applications—for new services we could develop and sell and for new ways to measure key marketing and advertising variables—but only rarely, if at all, for changes in our basic sampling methods or weighting procedures. In this respect, I believe not much has changed. *And fortunately for us, but unfortunately for the quality of our work, there is usually no way for our clients to test the accuracy of our findings.*

My interest in this problem is prompted by my experiences as a British-trained marketing researcher who has worked for the last 18 years in the US and who has managed, with varying degrees of involvement, surveys in more than 70 countries. My interest was given a powerful nudge by the failure of the British polls to predict a Conservative victory in 1992, and the excellent Market Research Society report on the polls by David Butler, John Barter and others.

Over the years I have been startled, and sometimes shocked, by the huge differences in basic survey methodologies from country to country. Yet very little attention has been focused on these inter-country differences and few people have questioned whether the methods used elsewhere might be better than those in general use at home.

Let me give you a few simple examples. For most purposes, quota sampling has been regarded as unacceptable in the US since the 1950s. Probability sampling is the norm here—in market research and opinion polling alike. In most of Europe, however, most in-person surveys for both market research and opinion polling have used quota samples. When it comes to telephone surveys, it is unacceptable in the US not to use random-digit-dialing (RDD), to include both unlisted numbers and recent movers. Yet, in many countries telephone surveys often do not use RDD.

Looking specifically to political opinion polls, one also sees large cross-national differences. In some countries the data are routinely weighted by "past voting," bringing claimed voting in the last election into line with the actual results. In the US we can demonstrate that this would have made some of our very accurate election forecasts extremely inaccurate.

In other words, what French and German pollsters have regarded as normal good practice—quota samples, weighted by past voting—Americans would regard as gross malpractice.

A Survey of Public Opinion Polling Firms

Awareness of these national differences prompted me to conduct a modest survey of the leading polling firms in 13 countries. It was not a sample survey. I attempted a census of the best known polling firms—with a little snowball sampling thrown in. I wrote to what I believed to be the best known, established organizations which conduct pub-

Survey Methods:. . .

I. Normal Methodology for Public Opinion Surveys

1. Normal Sample Sizes

•Vary from 500 (one firm in the U.K. and one in Italy) to 3,600 (one organization in Japan) with substantial variations within countries.

2. Basic Methods

Uniform

•All or mostly in-person probability: Australia, South Africa.

•All or mostly in-person quota: U.K., Mexico.

•All or mostly telephone: US, Canada, Germany, Denmark.

Mixed

•Japan (in-person probability, telephone, mail and self-completion).

•Brazil (in-person probability, in-person quota).

•Portugal (telephone, in-person quota).

•France (telephone and in-person quota).

•Italy (telephone and in-person probability).

3. Timing

Number of days in field: varies from 1 to 14 (several firms said "usually only one day").

Whether usually includes weekends: Yes (almost everywhere, except for one German firm).

II. Weighting

•Germany: Past voting, geography, sex, age.

•UK: Geography, sex, age, socio-economic class, housing, working/not working*, past voting*, number of eligibles in household.*

•Australia: Geography, sex, age.

•Denmark: Past voting, geography, sex, age, socio-economic class, education, working/not working, housing.

•USA: Sex, age, race, education, number of telephone lines*, number of eligibles in household*, how often at home*.

•South Africa: Geography, sex, age, race.

•France: Past voting, geography*, sex*, age*, socio-economic class*, working/not working*.

•Brazil: No weighting.

•Mexico: Sex, age, socio-economic class, working/not working.

•Canada: Sex, age, race (language)* (one Canadian poll: no weighting).

•Portugal: Past voting, geography, sex, age, working/not working.

•Japan: No weighting.

•Italy: No weighting of telephone surveys; weighting of in-person survey by geography, sex and age.

*Some firms only.

III. Telephone Surveys

1. Sampling Methodology

•Random digit dialing (RDD): USA, Canada.

•Directories: Denmark, Portugal, Germany, Italy.

•Other (including "database quota"): France.

•Mixed (RDD and directories and "database"): UK, Australia.

•Mixed (voter registers and directories): Japan.

2. Clustering

•Unclustered: Australia, UK, Portugal, Denmark, Italy.

•Mixed (some clustered, some unclustered): US, Canada, Japan, France, Germany.

3. Telephone Surveys—Selection of Individual

•Random selection grid: Germany*, USA*, Portugal*, Denmark, Canada*.

•Youngest male" method: Australia*, USA*, Canada*.

•Quota: UK*, Portugal*, France, Australia*, Italy.

•Birthday method: UK*, Canada*, Germany*.

•List of named individuals: Japan.

*Different firms use different methods or some firms use different methods.

4. Telephone Surveys—Number of Call-backs (re-calls)

•Varies from zero (0) to 6 or more.

lic opinion surveys in these countries. I surely missed some excellent companies that I should have included. With one reminder, 35 of the 37 firms I wrote to have sent me replies.

The main purpose of this small survey is not to make judgements about what is good or bad, or to criticize the methods used in some countries—although I will do that. Rather it is intended to make people aware of these differences, to stimulate debate and to encourage survey researchers—whether they do opinion polls, academic research, public policy research or marketing research—to question their methods and to test possible improvements.

Henry Durant, for many years head of the Gallup Poll in Britain, once described polling as "the stupidest of professions." Who else is stupid enough to make a forecast on Thursday morning which may be proved wrong on Thursday evening? No doubt many market researchers share this view, but put it differently: "Why do pollsters have to embarrass the whole marketing research industry by getting elections wrong?" One answer might be that we usually get elections right and, thereby, validate the use of the sample surveys used for marketing research.

I have chosen for my little survey to focus on public opinion polls, for several reasons:

. . . Cross National Comparisons

IV. In-Person Surveys

1. Clustering In-person Surveys

•Number of clusters varies from 50 to 400.

2. In-person Quota Samples: Quota Controls

•Mexico: Sex, age, socio-economic class, working/not working.

•UK: Sex, age, socio-economic class, working/not working.

•Brazil: Sex, age, education, working/not working, industry.

•Portugal: Sex, age, working/not working.

•France: Sex, age, profession.

3. In-Person Probability Samples—Basic Methodology

•Germany: Area probability, random walks.

•Australia: Area probability.

•South Africa: Area probability and household listings.

•Japan: voting register, residents/household (inhabitants) listings.

V. Election Forecasts (Last National Elections)

1. Methodology

•Telephone: Denmark, Germany*, Australia, USA, Canada, Portugal*, Japan*.

•In-person quota: UK, Portugal*, France, Mexico.

•In-person probability: Japan*, Brazil, South Africa.

*More than one method used in country.

2. Use of Turnout Questions to Determine Likelihood of Voting

•Not normally used in countries with mandatory voting. Otherwise generally used, except for two firms in Canada.

3. Timing

•Number of days in field: from 1 to 8.

•How long before election fieldwork ended: from 1 to 8 days (except where laws forbid publication of late surveys, where it ends sooner).

4. How Undecided (Including Not Sures/Don't Knows) Voters Allocated

•In proportion to decided voters: Denmark, UK*, USA*, Australia*, Canada, Portugal*, Japan, Germany*

•Based on replies to other questions: Australia*, Mexico, US*.

•Based on experience of previous elections: Australia*, Portugal, Germany*.

•Other methods: UK*, Brazil.

*Some firms only.

NOTE: In USA, France, South Africa, Canada and Japan some firms do not allocate undecided voters at all.

5. Use of Follow-Up Question to "Squeeze" Undecided Voters

Generally used except for most firms in Japan, one firm in France and one firm in Brazil.

1. Opinion researchers share a common objective, that of generating accurate data about the general public. We are comparing apples and apples.

2. They worry much more than most marketing researchers about the precise accuracy of (at least some of) their measurements, because their results are tested in the highly visible crucible of elections.

3. For most of the firms surveyed, published opinion polls are 5% or less of their total business, so there is a big spin-off into the mainstream of marketing and other survey research. Historically they have often led the way in the adoption of new and better methodologies to improve accuracy—followed by much of the marketing research industry.

Arguably, many of the most important improvements in the research industry followed polling debacles. The 1936 failure of the Literary Digest's straw poll, and George Gallup's success, greatly spurred scientific polling. The 1948 forecast of a Dewey victory over Truman led to much more timely polls and big changes in sampling—including the demise of quota sampling in the United States. The 1970 British election led British pollsters to worry much more about differential turnout and late swings—but not enough to avoid their 1992 debacle.

In the most recent US presidential election Harris tested, at our own expense, three different ways of weighting, two different ways of asking the crucial voting question, different question orders, and different ways of selecting the individual to be interviewed within a household. Over the years we have evaluated our public opinion surveys based on how many days and which days of the week we were in the field, different methods of drawing our RDD samples, and the effects of multiple callbacks and weighting for non-response.

We did this because we were running scared. Samuel Johnson said that nothing concentrates the mind like the prospect of being hanged. The prospect of picking the wrong winner on election night comes close, believe me.

Current Polling Practices

With that as a long preamble, let me present the main findings of my little survey which relates, of course, to nationwide opinion surveys.

As the summary table (pp. 4-5) indicates, there are huge, fundamental differences in what the main polling firms do. The biggest differences, by far, are inter-country, but some intra-country differences are large too. Had I included more firms I would probably have found even greater differences.

Some National Variations Are Inevitable And Desirable—But Many Aren't

While I hope researchers will ask what they can learn from other countries and should import, I am not arguing that every country should use the same methods. It's not true that one size fits all. The reverse is true. Quota samples, historically, have worked well in many countries. Some countries publish (and researchers there can use) complete voter lists, electoral registers or household lists; in other countries the lists aren't available. Only some countries have high telephone penetration. Weighting procedures must depend upon the demographics of the population and the availability of reliable data for weighting.

This past summer the Harris firm in Mexico successfully used techniques that I have not used before, including "secret ballot" voting and the judgmental assignment of marginal voters based on their replies to questions other than their voting preferences. I'm inclined to advocate the same techniques in Northern Ireland, but not in the United States, most of Europe, or Australia.

Some country-to-country differences in methods are inevitable, then. But many others make little or no sense to me and seem to reflect national chauvinism, inertia, ignorance, and even incompetence. I'll mention five research practices that I find entirely unacceptable.

1. *Not using RDD in telephone surveys.* Most polling organizations outside the US and Canada don't use RDD in most of their public opinion surveys, although we know that failure to use RDD excludes recent movers and many others who are not listed. Several of the firms which I surveyed tell me that RDD is more expensive because it involves calling non-working numbers and businesses. Some report receiving complaints when calling unlisted residential numbers. While we face both these problems here in the US, the biases of not using RDD are unacceptable.

2. *Using clustered samples in telephone surveys.* Several leading companies in the United States and Canada, and all those in Japan who completed the questionnaire, say they use clustered telephone samples, a vestigial relic of in-person interviewing. It substantially increases sampling error.

3. *Using no weights or limited weights.* The Japanese firms, one Italian firm and one Brazilian firm surveyed report that they normally use no weighting. Several other firms use relatively few weights, particularly for telephone surveys. (There are also many within-country variations in weighting.)

While there have to be differences in weighting between countries (for instance, weighting by race and education works much better in the US than in Britain), some of the differences suggest that little thought is given to finding good weighting factors. This is one area where I believe large improvements could be made by many of the world's leading polling firms, whether they use telephone or in-person survey, quota or probability sampling. We have to apply several substantial weights to our telephone survey data in the US to correct biases which I suspect are not adequately corrected in some other countries.

4. *Stopping fieldwork well before elections.* More than a few of the firms surveyed say they make election forecasts based on polls which do not include interviewing in the last day or last two or three days before the election. In my opinion, this is suicidal. I have personally worked on several major elections (from the British General Election of 1970 to the New York mayor's race in 1993), where we only got the winner ahead because of the results of interviews conducted the day and evening before election day. Obviously, national laws banning polls or their publication, at the end of the campaign are a problem. But one firm told us that even when there were no such restrictions, they stopped interviewing 8 days before the election (possibly to meet the deadline of their client?).

5. *Weighting by claimed past voting.* Many polling firms in Germany, Denmark, France and Portugal routinely weight their polls by past voting. In the US this would have destroyed the accuracy of all of our recent forecasts, as it would have in Britain in the 60s and 70s. Over the last thirty years I have often observed that claimed past voting differed substantially from the actual result. After elections in the U.K. and America, more people routinely are found claiming to have voted for the winner than did so. When governments become unpopular, more voters tend to deny they voted for them. If race is a

factor, people who voted against a black candidate sometimes claim to have voted for him.

In the most recent election we covered in the United States, the New York mayoral election, claimed past voting bore little resemblance to the actual vote of four years earlier. The final result was a 2.5% margin of victory for the white challenger, Guiliani, over the black incumbent, Dinkins. Our poll showing Guiliani one point ahead was the most accurate of all the polls conducted in the last week of the campaign. Had we weighted by claimed past voting we would have shown Guiliani winning by a huge 22 point landslide. (He actually won by a 2.5 point margin.

Quota Versus Probability Sampling For In-Person Surveys

Among all the international differences, perhaps the most startling is the widespread use, mainly in European countries, of quota sampling for in-person surveys—something which would be unthinkable in the US and which scarcely any of the survey firms outside Europe say they normally use. However, Harris used quota sampling very successfully in the recent Mexican elections.

It is difficult to find a reputable statistician who will defend quota sampling: there is so little control, the bias toward available respondents is so strong, and the dependence on the judgment of interviewers—on socio-economic class for example—so great. And yet, quota sampling has a long history of working well in Britain, France, Germany—and now it has worked well in Mexico.

In Britain, NOP, under Peter Hyett, pioneered the use of probability sampling, based on the electoral registers, for opinion surveys in the mid-60s based on the electoral registers. Both NOP and Harris (from 1969) used probability samples for most of their opinion surveys, including their pre-election polls, until the mid-70s. There were only two major problems. On average, their results were slightly less accurate than the quota samples in predicting elections. And the fieldwork costs were much higher. Sadly, but unsurprisingly, NOP and Harris joined the pack and went back to quota sampling.

Of course, the British polls, following several elections where they did pretty well at forecasting the results, were spectacularly wrong in the 1992 general election. We will never know how much better polls based on probability sampling might have performed. But were I running a British poll, I would be sorely tempted to experiment with random sampling again.

Experimental Work in Progress

One of the questions in my little survey of polling firms asked what new methods were being used or considered. Nothing very radical was mentioned—although the British pollsters, scared by their recent disaster, mentioned several possible changes, including the use of "secret ballots," new quota controls, and new weights. Several polling firms in different countries are looking at new ways to allocate the "don't knows" (or "not sures") in their final pre-election polls, which point up a cautionary tale from the 1992 US election. Following analysis of hundreds of pre-election polls which showed a clear tendency to overestimate the vote for incumbents and underestimate that for challengers, Gallup allocated all the undecided votes to Clinton and thereby transformed a decent forecast into the worst forecast of all the published polls. The moral might be that what works for some elections doesn't necessarily work for others, that they should have "squeezed" the undecided harder, or that things change and that our polls must keep on changing also.

If so, it's bad news that my little survey found little evidence of change. Indeed, with the exception of the Brits, only the Japanese polling firms seem to be thinking of making changes. Several Japanese firms expect to use more telephone and fewer in-person surveys in the future. There is, although it was not widely mentioned in the survey, a general trend away from in-person to telephone surveys.

Unfortunately, there are many reasons why the research industry does not worry enough about the validity of its findings, invest enough in researching its methods or spend enough to use the best methods. Most of the results of survey research are never validated and cannot be. Few of our clients push us to use the best methods, or are willing to pay much more for us to do so. And, for the media, it is usually true that polls are judged on their newsworthiness, not on their quality. Alas, for most editors a poll is a poll is a poll, regardless of its accuracy or validity (except, of course, on election day).

Humphrey Taylor is president and CEO, Louis Harris and Associates