

INTERNET VERSUS CONVENTIONAL POLLS IN THE 2005 GENERAL ELECTION

Peter Kellner

Britain's pollsters went some way to rehabilitating themselves in the 2005 general election. Every poll published on the morning of 5 May was correct to within 2 percentage points for Labour, Conservatives and the Liberal Democrats. After the horrors of 1992 (when not a single poll in election week came anywhere near the final result), and the less-than-perfect polling performances in 1997 and 2001 (when most surveys still overstated Labour support and understated the Conservatives), the outcome of the 2005 election was met by Britain's pollsters with understandable relief.

However, it would be a mistake to believe that all the problems of the previous three general elections have been solved. Table 1 illustrates the point. It shows how many face-to-face and telephone polls in each campaign came within three points of the final result for Labour and Conservative, and how many deviated by three points or more. For 2005, it also shows the record of internet polls conducted by YouGov, BPIX, the British Election Study and Harris Interactive.

[KELLNER TABLE 1 NEAR HERE]

Before analysing these figures, two points should be made clear. The first is that nobody can be sure whether a 'wrong' figure in the early or middle stages of a campaign was actually wrong, or a correct depiction of a state of the parties that subsequently shifted. In 2005, all five overestimates of the Conservative share by YouGov (two), BPIX (two) and the BES (one) occurred within the first ten days of the campaign; our belief is that Tory support did in fact slip slightly after that. However, if one takes the weekly averages of each pollster in each campaign, there is no sign of any *large* movements during the course of any of the past four campaigns; so the table does give a broad picture of the scale and direction of systematic errors.

The second point is that even perfectly-conducted polls are subject to margins of error. Over the course an election campaign where nothing changed, one would expect the table above to record an 'error' rate of around 20 per cent - but with the 'wrong' figures evenly

distributed between too high and too low. (One reason why the Internet polls had a lower error rate was that they were all, in effect, panel surveys, in which people whose party allegiances had been measured before the campaign started were re-interviewed during the campaign. This reduced the likelihood of purely random sampling errors, beyond those embedded in the – very large – baseline samples.)

All that said, the table shows that the problems that first appeared in 1992 have reduced, but not disappeared. The overall ‘error’ rate fell from 87 per cent in 1992 to 53 per cent in 1997, but in the latest election it was only slightly better than that. And the direction of error has remained the same: ~~every~~ ‘error’ by conventional polls in both 2001 and 2005 took the form of an overstatement of Labour’s share and/or an understatement of the Conservatives’ share. It should be stressed that not every pollster was affected equally. ICM, which had made the greatest change to its methods after 1992, produced a high proportion of the more accurate polls in 1997 and 2001.

By 2005, all the pollsters had devised new techniques that, they hoped, would remove the Labour bias from their figures. The outcome was still patchy. The overall error rate was down, but only slightly. As in the previous three elections, the most common error was an overstatement of Labour’s support. The effect was to produce a number of stories in a wide range of papers saying that Labour was heading for a third landslide victory. This is not surprising when four of the conventional pollsters each produced at least one campaign survey showing Labour leading the Conservatives by ten points or more – that is, by a wider margin than in 2001, when Labour’s lead in the popular vote was nine points.

Had these aberrant polls all been conducted at around the same time, there would be a case for saying that they were right at the time, and that all detected the same campaign blip. But this is not the case. The fieldwork for ICM’s 10-point lead was 13-15 April, for NOP’s 10-point lead 22-24 April, for MORI’s 10-point lead 29 April-1 May, and for Populus’s three tracking polls showing leads of 12-14 points, dates between 27 April and 2 May.

The conclusion is clear. Conventional polls still seem prone to overstate Labour support and (though less often) to understate Conservative support.

It should also be noted that, although the polls finally converged on election day, they did so using different methods for excluding expected non-voters. MORI went furthest: throughout the campaign its main voting intention figures took account of only those people who said they were absolutely certain to vote. At the other end of the scale, YouGov did not

exclude anyone until its very final survey, when it looked at three different possible turnout adjustments, and in the end chose the smallest, raising the Conservatives' share and reducing the Liberal Democrats' share by just one point each.

In short, the conventional polls did better in 2005 than in the previous three elections, but they have not entirely eliminated their pro-Labour bias; and they do not agree completely on the best way to convert raw data into accurate published results. In contrast, the three Internet polling companies had a good election.

It is clear, then, that there are specific issues that still need to be resolved in relation to election polls. But is that all? Give or take the odd tweak to the methods for calculating voting intention, is everything well in the world of polling?

The honest answer is 'no'. There are forces at work that will, I believe, prompt a fundamental change in the way polling companies operate. In this instance I am not referring to the rise of internet polling (though I would expect it to overtake both face-to-face and telephone polling within the next decade or so). I am referring to something more basic: faith in the sanctity of the sample.

Consider conventional sampling theory. In an ideal world, one would contact pure random samples, with names and addresses drawn from the electoral register. There would be a 100 per cent response rate. We could then calculate margin of error, and be confident that 19 times out of 20, the true figure would lie within a defined range.

Of course, the ideal world does not exist. Pure random samples are impossible to obtain. Even to attempt to achieve such samples is extremely expensive and time-consuming. They take weeks to complete, and do well to achieve a response rate much above 60 per cent. Such surveys are generally regarded as the 'gold standard' of survey research, yet nobody can be quite sure how accurate their results are, for nobody can be certain whether the 40 per cent the researchers cannot reach are like or unlike the 60 per cent they do reach.

In any event, random surveys are of little use when – as in a general election campaign or in the aftermath of a specific event such as the Budget or the recent London bombings – results are needed within two or three days, rather than two or three months. This is why most polls conducted for the media apply different techniques. Most conventional polls are conducted by telephone. Each fieldworker is given a list of randomly-generated landline telephone numbers, and dials them in turn. In addition to the attitude questions, such as which party people would vote for, or who would make the best Prime Minister,

respondents are asked a number of questions about themselves – their age, gender, occupation of head of household, housing tenure and so on. (Data about region is generated by the telephone number of each respondent.)

Once all the fieldwork is complete, the polling company then weights the raw data by this demographic data. The variables used for this weighting process vary slightly from polling company to polling company; but they all operate according to the same basic principle: that if the weighted data accurately reflects the demographic profile of the population as a whole, then the results for the attitudinal questions are likely to be accurate (or as accurate as the size of the overall sample will allow: nobody pretends that the laws of probability can be repealed).

As with random polls, the underlying assumption is that the respondents that can be reached have the same range of views on every subject as those that are not reached, once the data have been adjusted to correct for any demographic imbalances in the raw sample. A moment's thought should demonstrate that this is an heroic assumption. There are a number of reasons why it may be wrong.

1. Response rates to telephone surveys are low. According to ICM's Nick Sparrow, 'Only about one in six people approached by a pollster agrees to be interviewed and the indications are that those who are interested in politics are most likely to vote in elections and are also more likely to agree to participate in polls on the subject.'¹. That comment, from one of Britain's best traditional pollsters, acknowledges that response rates are not only low, but that respondents are unlike non-respondents.
2. Not only are response rates low; they are falling further. Increasing numbers of people are using voicemail and/or caller ID technology to screen incoming calls. It is easier for potential respondents, even if they are at home, to screen out pollsters and other cold-call companies. Once again, this not only reduces response rates; it also introduces another potential source of error, as it may be that on some issues, people who do have and use voicemail and/or caller ID technology hold different views from those in the same demographic groups who do not.

¹Nick Sparrow, 'More anger than apathy', *Guardian*, 6 November 2001

3. Another of Britain's more thoughtful pollsters, NOP's Nick Moon, points out a further source of possible bias. In a 1999 book², he wrote that conventional polls 'will inevitably be biased towards those who are more easily available: someone who hardly ever goes out is far more likely to be in if a quota interviewer calls, than someone who is almost always out'. This would not matter if the attitudes and experiences of people who mostly stay in are the same as those who go out a lot. Not only can nobody guarantee this; it seems likely that on a number of issues, the attitudes of home-stayers are likely to be different from those of go-outers.
4. Not everyone has a landline telephone. This used to be a big issue, when telephone penetration was well under 90 per cent. The people without telephones tended to be older and poorer than those with phones. One of the active debates within the polling community concerned the best way of correcting for this bias. By the 1990s, this problem had largely gone, as telephone penetration passed 90 per cent. But it has now returned in a different guise: a number of people, especially students and other young adults, use mobile phone either wholly or mainly to make and receive telephone calls. As telephone polls contact only landline numbers, the mobile generation (in both the telephonic and geographical meanings of the word 'mobile') is proving increasingly hard to reach.
5. In any event, demography is often only loosely correlated with attitudes, especially on politics. Forty years ago, if a poll's weighted data contained the correct distribution by social class, then it had a good chance of producing accurate voting intention figures, as middle class voters overwhelmingly supported the Conservatives, while working class voters overwhelmingly voted Labour. This is no longer true. A poll can easily produce inaccurate political data even if it incorporates impeccable data by age, gender, social class, region, housing tenure, car ownership and the number of foreign holidays taken.
6. Quite separately from the challenge of achieving a representative sample, survey researchers occasionally produce results that are distorted by interviewer effects. There is now copious evidence from Europe and the United States that some respondents are prone to give 'social satisficing' answers to some questions –

² Nick Moon, *Opinion Polls: History, Theory and Practice* (Manchester: Manchester University Press, 1999).

consciously or sub-consciously suppressing feelings that they fear might offend the interviewer. One of the likely reasons for the tendency of many polls to overstate Labour support and understate Conservative support in the past 15 years is that there was a 'spiral of silence', whereby some Conservative voters were unwilling to admit their allegiance to a stranger.

For Internet polls such as those conducted by YouGov, some of the problems are similar to those conducted by telephone, while others are different.

Advantages of Internet polling

- Very low marginal costs mean that larger samples can be polled for the same cost.
- As Internet questionnaires are completed completely anonymously, there is no interviewer effect, and hence no 'social satisficing'.
- Respondents can complete their surveys when it is convenient to them – they do not have to be at a landline when a fieldworker calls. The risk of availability bias, identified by Nick Moon, is likely to be less.
- YouGov has a great deal of information about its panel members. It can therefore use more sophisticated techniques than are open to telephone companies to select its samples and weight its raw data.

Disadvantages of Internet polling

- Only 60 per cent of the public has Internet access, and not all of that 60 per cent are willing to join Internet panels such as YouGov's. As with other kinds of polling, there is a danger that the people online polling does reach are unlike the people that they do reach.
- On some issues there is an obvious bias. For example, YouGov panel members are far more likely than the rest of the public to support the idea of allowing people in general elections to vote online; not only do YouGov panel members have computer access; by virtue of joining the company's panel, they demonstrate a willingness to express their opinions online.

It is precisely because we at YouGov are conscious of the potential weaknesses of the raw data obtained from online samples that we have developed special techniques to address them. These include a sophisticated, computerised system of sample selection, and an even more sophisticated method of weighting our raw data. Together these strive to ensure that our published data is representative not only of the demographic character of Great Britain (for normal, nationally representative polls), but *attitudinally* representative. For political polls, we make sure that our weighted data contains the right proportions of the readers of each national newspaper. We also weight by past political affiliation. For our 2005 election polls we weighted by recalled past vote (after aiming off for false memory syndrome). Since the general election, we have weighted data our data to achieve the correct proportions of Labour, Conservative and Labour identifiers. The percentages for each were derived from data collected from more than 40,000 respondents during the first half of May.

This is a special, albeit highly important, example of a wider point. Given the hazards of conventional as well as Internet polling, we believe it is no longer sufficient for any of us to use only demographic weights on all occasions. Wherever possible, YouGov applies weights that are relevant to the subject being surveyed, drawing on census or other official data, or large-scale research data such as the National Readership Survey.

Such methods cannot guarantee accuracy. Nothing can. In the end there are two options. One is to say the hazards of survey research are so great that none is done. The other is to acknowledge the problems and do as much as one can to solve them. We at YouGov believe that solutions are available, but that, because of the potential weaknesses in all raw samples (whether obtained face-to-face, by telephone or online), it is vital to use weights that are relevant to the survey being done, and not to assume that standard demographic weights are enough for all purposes.

Naturally we expect online survey research to become the dominant method for conducting political surveys over the next decade or so. But in some ways the more important intellectual revolution for the polling community will concern techniques for drawing samples and weighting the raw data.

KELLNER TABLE 1

(Figures show numbers of polls)

	1992	1997	2001	2005	
				Conven- tional	Internet
LABOUR					
Overstated by 3 percentage points of more	48	42	24	21	0
Within 3 percentage points of final share	2	7	5	10	25
Understated by 3 percentage points or more	0	0	0	0	0
CONSERVATIVE					
Overstated by 3 percentage points of more	0	2	0	0	5
Within 3 percentage points of final share	11	38	16	26	20
Understated by 3 percentage points or more	39	9	13	5	0
OVERALL					
Within 3 percentage points of final share	13	45	21	36	45
Variance of 3 percentage points or more	87	51	37	26	5
<i>% 'Error' rate</i>	87	53	64	42	10