

BPC/MRS Enquiry into Election Polling 2015: Ipsos MORI Response and Perspective

Roger Mortimore, Paul Baines, Robert Worcester and Mark Gill

What went wrong?

The accuracy of the forecasts drawn based upon pre-election opinion polls generally, including those by Ipsos MORI, was disappointing at the 2015 general election. As all the polls showed very similar errors, it seemed likely that all were affected by the same factors, and an industry-wide enquiry (chaired by Patrick Sturgis, sponsored by the Market Research Society and the British Polling Council) was set up to investigate. Nevertheless, we at Ipsos MORI were naturally concerned to learn what we could from the evidence already available to us, and to apply the lessons to improving our regular polling as quickly as possible. In our recent book, *Explaining Cameron's Comeback* (Worcester, Mortimore, Baines, & Gill, 2016), we explained the results of our initial investigation into what went wrong; at the time we went to press, we were still awaiting the findings of the Sturgis enquiry. Their report is now published (Sturgis et al., 2016), and we have considered their conclusions and the extent to which they may help to improve polling accuracy in future elections. We find much of their analysis convincing and useful; but some important questions remain unanswered.

The respected pollster Humphrey Taylor, in his review of *Explaining Cameron's Comeback* for *IJMR* (Taylor, 2016), contrasts our explanation for the performance of the polls with that put forward by Sturgis et al. However, we are not so far apart as he suggests: it is more a matter of different perspectives from which the polls' results can be viewed. The starting point for the Sturgis investigation is, in effect, the theoretical 'perfect poll' (which can never be achieved), and the report explains the failure in terms of how the polls fell short of that perfection in practice. Our starting point, by contrast, is the knowledge that our polls - admittedly imperfect in many of their details - nevertheless produced satisfactory forecasts at the 2005 and 2010 elections, and so we focus on what was different in 2015. These two approaches seem on the face of it to throw up different answers, but they are not contradictory but complementary. The main thrust of the report's findings is probably right, and Humphrey Taylor (2016, 487) is also right that the key problem underlying everything else

is a response rate bias. Yet we still need to go further, and understand *what was wrong in 2015 that was not wrong in 2010*, if we are to translate our knowledge into a practical solution to improve the accuracy of our polls. That is where our analysis comes in.

There are many ways in which polls might go wrong. After the 1992 election, at which it was widely perceived that the polls had failed, an MRS enquiry (Market Research Society, 1994) diagnosed a combination of: ‘late swing’ (voters who changed their minds after the last poll was conducted) estimated to be circa 20-33% of the ‘error’; use of inaccurate demographic weights and quotas in sample control, largely due to the delay in providing up to date census data; and some under-representation of Conservative support, either through their voters’ reluctance to participate in the surveys at all or reluctance to admit their voting intentions – the so-called ‘shy Tories’.

The 2015 enquiry report considers that late swing was ‘not a significant factor’ in the discrepancy between the poll forecasts and the result (Sturgis et al., 2016, p. 37), and finds no significant problem with the demographic controls¹. Since our polls marginally over-counted rather than under-counted the number of Conservative voters, clearly there’s no ‘shy Tory’ problem either. There must be a new factor. The Sturgis Report shows clearly what that new factor is, but not why it hasn’t troubled polls in previous elections nor, more importantly, what pollsters can do about it.

To understand why knowing what went wrong is not the same as knowing how to put it right, it is first necessary to understand the nature of opinion polling intended to forecast election results. It is a very unusual form of market research. Normally, market and opinion research has four elements: (i) definition of the objectives and development of the methods, including the questions to be asked and the sample to be drawn; (ii) data collection; (iii) data analysis; and (iv) data interpretation and reporting.

¹ The Sturgis report did suggest that the polls might have under-represented the oldest voters because they relied on weighting or quota designs with very broad age bands, which proved too crude for the purpose, and some commentators have picked up on this (e.g. Mouncey, 2016). But their analysis only looked at online samples (Ipsos MORI has separated the 65-74 and 75+ age groups in the weighting of its telephone polls for many years), and in any case, the report concludes that “the under-representation of voters aged 75 and over in the poll samples is unlikely to have made a notable contribution to the 2015 polling miss” (Sturgis et al., 2016, pp. 54–55).

That interpretation must take into account the inevitable imperfections of the survey and the impossibility of measuring some phenomena directly.

All surveys have imperfections and all surveys are necessarily imprecise. No poll or survey in English can escape the use of the English language in forming the questions, an imprecise language at best. It is impossible to be sure of excluding interviewer or interviewee error. Sampling imposes practical limitations: telephone polls miss those not on the phone, or where the phone is out of order, or where respondents are on holiday or at work; Internet polls miss people who are not on the internet or who will not sign up to do surveys. Further, every poll or survey has an inescapable margin of error because of the random element in selecting a sample. And on top of this, every poll is at the mercy of its response rate, and response rates are generally low²: in fact, response rate bias is an ever-present threat.

In most research we allow for this in our interpretation: contrary to the old adage, the numbers do *not* speak for themselves. Interpretation calls upon our experienced judgment as researchers to produce insights which go beyond the numbers generated by the survey and to explain their meaning, their implications and their limitations. Opinion polls – surveys of (usually political) public opinion conducted for media publication – are different. They live by the numbers, and make and depend upon the assumption that the findings are a direct and literal reflection of objective fact: the primary output of an opinion poll is not the researcher's report interpreting the percentages, but the percentages themselves, published as they stand.

Nevertheless, with polls that are measuring opinions rather than attempting to predict behaviour, this need not be a serious problem. Suppose that 40% of the public would say, if asked, that they are 'satisfied with the way the Prime Minister is doing her job'. What are the direct consequences of that number, and how do they differ from 45% being satisfied? Only by context and comparison with

² It has been suggested that this is increasingly a problem for telephone surveys, with wide use of call screening and caller ID and a growing unwillingness to answer calls from unrecognised numbers, probably fuelled by the explosion in the volume of marketing/sales calls over recent years. While this is, indeed, a cause for concern it seems unlikely that difficulties specific to telephone polling played a key role in the performance of the polls at the 2015 election when the voting intention forecasts of the online polls were almost identical to those of the telephone polls.

similar other findings does a poll of this sort acquire a useful meaning – we can see, for example, how the number changes over time or how the Prime Minister scores by comparison with the Leader of the Opposition.

But when we poll voting intentions, the situation is different. Even though the same limitations which apply to any other survey necessarily still apply, the numbers are expected literally to reflect likely future voting behaviour. Using the poll as the basis for a prediction is not considered enough. The poll itself is expected by the media and therefore by the public to be the prediction: the numbers in the poll (somehow adjusted, perhaps, but using published, transparent and systematic methods) are expected to match the actual voting outcome in every way, except perhaps with the exception of late swing. (Pollsters argue with good reason that polls are only accurate at the time the fieldwork is conducted and cannot take into account extraneous events taking place after the poll is undertaken but before the electors cast their ballots - typically only election day itself in the UK – yet, typically, late swing is dismissed as the pollsters' excuse rather than being accepted as a factor entirely beyond their control.) Moreover, the media demand - and lead the public to demand - a degree of precision that no polling agency can consistently deliver (Worcester, 1996).

Even discounting the risk of late swing, and unrealistic expectations of the margins of error, this creates a problem. In theory, with a perfectly representative sample and questions that correctly identify the voting behaviour of each respondent, a poll could predict in this manner. But as practical pollsters, we know that our samples will always be imperfect, and that the answers some respondents give us may well be misleading if taken literally. This may arise from factors such as social desirability bias – respondents tending to give the answer they believe will be favourably viewed by the interviewer, rather than one that truly reflects their own opinion (e.g. Noelle-Neumann, 1993) - or simply because respondents might be unable to predict their own behaviour accurately - and nearly three decades of research indicates that behavioural intention does not correspond 100% with actual behaviour (Ajzen, 1991).

In other types of research, we could simply take account of this in the interpretation of the data and reporting of our conclusions. But for election polls, our methods have to be designed to overcome all these distortions mechanically and so lead to a ‘headline’ figure that is sufficiently close to the electorate’s actual voting to be a satisfactory forecast. We know that just polling ‘straight’ does not achieve this:

- We have to weight, to correct for important response biases and perhaps other biases in the sampling method;
- We have to incorporate some degree of modelling, certainly to deal with turnout and perhaps other complications such as the behaviour of ‘don’t knows’;
- We have to design our questionnaires in a way that persuades respondents to give us the least-misleading answers (for example, deciding which, if any, party names and individual candidate names³ to include as prompts in the voting intention question⁴, and using turnout questions that encourage the respondent to admit past or future abstention).

All of these adjustments are necessary to producing an accurate forecast. But their operation is mostly also dependent on factors that might change from election to election⁵. When the measurement goes wrong, as in 2015, we may need to understand why our adjustments stopped working before we can refine them or replace them completely to achieve a better measurement in future elections.

What follows focuses specifically on Ipsos MORI’s polls. If the presumption is correct that the error in all the polls had the same cause then, in so far as other companies’ polls have methods in common with ours, the same lessons apply to all of them. But for companies using very different methods, for example relying on internet panels rather than telephone sampling, it may be that the practical solutions needed to overcome the obstacles to accurate polling are different despite the underlying situation creating those obstacles being the same. Nevertheless, we believe that an exploration of the

³ In 2001, we were criticised by some commentators for ‘changing our methodology’ in mid-election, because we started informing respondents of the names of the candidates once the identities of those standing were confirmed – as we had at many previous elections. As far as we know, none of the national polls in 2015 included candidate names, but their inclusion or omission was an issue of vocal controversy in constituency polling.

⁴ In recent years we have found that using questions which do not include party names seems to deflate support for the Liberal Democrats, but that prompting for too many parties inflates support for the smaller parties.

⁵ And which, for that matter, may also operate differently at devolved elections, local elections or referendums from their operation in general elections.

problems facing our methodology, and of the possible solutions we are considering to overcome them, is of wider interest in throwing light on the difficulties that pollsters face, even if it is more of a case history than a general theory.

Ipsos MORI’s final poll in the 2015 election was conducted by telephone on 5-6 May, with its prediction published on 7 May (election day) in the *London Evening Standard*; it consisted of interviews with 1,186 adults resident across Great Britain (of whom 1,096 said they were registered to vote). Table 1 compares the numbers in the final poll with the actual election result.

Table 1: The final poll and the election result (including abstentions)

	Election result			Ipsos MORI final poll projection			Difference		
	%	millions	Voters only	%	millions (projected)	Voters only ('headline forecast')	%	millions	Voters only
Total		50.5			50.5				
Conservative	22.4%	11.3	37.7%	24.7%	12.5	36%	+2.3%	+1.2	-1.7%
Labour	18.5%	9.3	31.2%	24.1%	12.2	35%	+5.6%	+2.9	+3.8%
Other parties	18.5%	9.3	31.2%	20.7%	10.4	30%	+2.2%	+1.1	-1.2%
Will not/did not vote	40.6%	20.5		30.5%	15.4		-10.1%	-5.1	

Source: Worcester et al. (2016: 214).

Although the Sturgis report states that “[t]he methods the pollsters used to collect samples of voters systematically ... under-represented Conservative supporters” (Sturgis et al., 2016, p. 4), this does not accurately characterise our poll, as can be seen from the table: our poll had *too many* Conservatives (24.7% in the poll against 22.4% actual). The problem was that our excess of Labour supporters was even larger (24.1% in the poll, 18.5% in reality). The headline prediction, of course, was re-percentage to exclude the non-voters and so under-estimated the Conservative share of the vote, but

this was not because there were too few Conservatives in our sample. We predicted too many Labour votes and not enough non-votes.⁶ This is the error that we need to explain and correct. The question is, why did this discrepancy arise? Was it because we really had too many Labour voters and not enough non-voters in our samples, or because some of those in our sample that we identified as Labour voters were in fact non-voters, or a bit of both?

The enquiry report points clearly to part of the answer. From other survey evidence (with large probability samples), it concluded that the polls' samples under-represented the politically disengaged. Because the disengaged are concentrated in Labour-supporting demographic groups, this means that the quotas used in recruiting the raw sample and the weighting used to construct a demographically-representative sample from this during the data analysis phase ended up replacing non-voters who were not willing to take part with otherwise similar Labour voters who were willing, and so Labour voters became over-represented in the sample. We find this explanation convincing. (And, interestingly, it would explain an excess of Labour voters and a shortage of non-voters but would not imply any shortage of Conservative voters. This was exactly the error we found in our own polls.)

Why did the same methods work in 2010?

But it still leaves a puzzle. If Labour voter over-representation, arising from the failure of our quota and weighting approach to fully correct a response bias, was the main cause of the error, what was different in 2010 (when the polls achieved satisfactory results using similar methods)? This conundrum is not solved in the Sturgis report. We can't simply assume that the sampling bias was new in 2015. It may equally have existed already, but been less closely correlated with voting behaviour. Since the methodologies were no different, something must have changed about the (potential) respondents. Perhaps the disengaged became harder to find or less willing to participate in

⁶ It should be noted that the table records the level of non-voting as 40.6%, implying a turnout of 59.4%, whereas the commonly-quoted "official" turnout was 66.1%. This is because the official turnout is calculated on the basis of the registered electorate rather than on the size of the whole adult population, a significant part of whom are not registered to vote (and some of whom are not eligible to vote at all, therefore not eligible to register). Since the universe from which opinion polls are sampled is the not the registered population but the adult population (from which the unregistered must then be excluded), and since the available sources of statistical information about population characteristics, necessary for setting quotas and weighting, also relate to the entire adult population, this is the correct basis for comparison here.

polls than they had been, so that fewer of them were interviewed – that would create a new sampling bias. But perhaps they were always under-represented, and it was a change in the voting behaviour of those who were over-represented, or a change in the relationship between their voting behaviour and the answers they gave to the pollsters’ questions, that was what was different.

The visible difference between 2010 and 2015 in our case was a change in answers to our turnout likelihood question. Every respondent was asked how certain they were to vote, and these responses were used in the turnout adjustment to our forecast. Our unadjusted voting intention data in 2010 and 2015 was very similar. The election results in 2010 and 2015 were very similar. If we had assumed that turnout patterns would be the same in 2015 as in 2010, and applied the 2010 turnout adjustment to the 2015 data, our poll results would have been well within the accepted margins of error (Table 2)⁷.

Table 2: Ipsos MORI’s 2015 poll using the 2010 turnout adjustment

Party	All declaring a voting intention		Projected turnout	Turnout corrected		Imputed votes of refusers	Final projection		Final Result
	Weighted number of people	Share		Weighted number of people	Share		Weighted number of people	Share	
Conservative	322	36%	85%	275	38%	23	298	38%	36.9%
Labour	309	34%	74%	230	32%	21	251	32%	30.4%
Other parties	267	30%	79%	212	30%	16	228	29%	32.7%
Total voting	898			717	100%	60	777	100%	100%

Base: For 2015 poll, 1,096 GB adults claiming to be registered to vote, of whom 923 declared their voting intention; turnout projection based on 2010 poll with 313 adults intending to vote Conservative, 289 intending to vote Labour and 323 intending to vote for other parties.

Source: Based on a table in Worcester et al. (2016: 223).

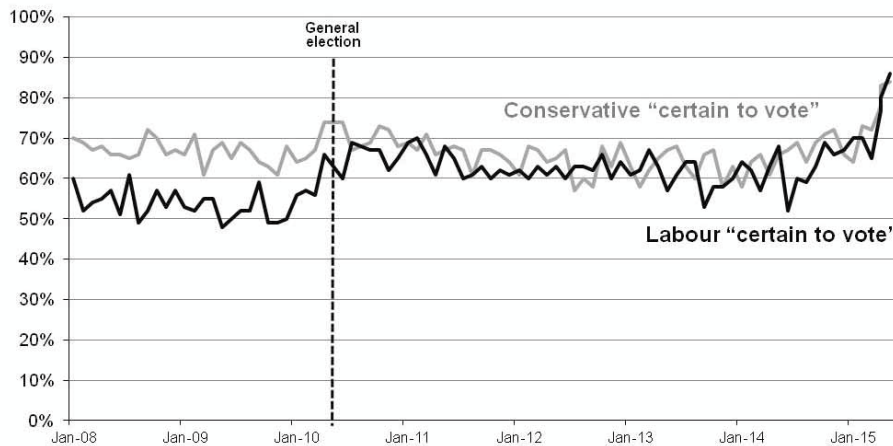
However, there was a change in the likelihood of voting responses and therefore a change in our turnout adjustment. The overall level of expected turnout (based on the proportions telling us they were ‘absolutely certain to vote’) rose. This rise was mainly among Labour supporters; the expected

⁷ Table 2 includes a column headed “imputed votes of refusers”. In every election poll a proportion of respondents tell us that they are certain to vote but refuse to reveal which party they are intending to support. In our final poll, rather than ignoring this group altogether we impute a voting intention for them on the basis of their newspaper readership, a question which is a relatively good predictor of voting intention for other respondents and which very few are unwilling to answer. This is a precaution against the possibility of refusers being overwhelmingly supporters of one party rather than another, which might skew the final forecast; but in 2015, as the table shows, the proportion of refusers was relatively small and the imputation had no effect on the headline figures.

turnout level for Conservatives was the same as in 2010. (In 2010, 85% of Conservatives and 74% of Labour supporters said they were certain to vote; in 2015, the figures were 84% for the Conservatives and 86% for Labour.) So, whereas in 2010 we had expected turnout to be significantly lower among Labour supporters than among Conservatives, in 2015 we expected the Labour turnout to be marginally the higher of the two. From our unadjusted figures of Conservative 36%, Labour 34%, we therefore projected an outcome of 36% to 35%, instead of the 38% to 32% which would have resulted if there had been no change in the pattern of the likelihood of voting answers. Hence the discrepancy between the forecast and the result. Since this posited a significant rise in overall turnout, and we now know that in the event there was no such rise, any explanation of what went wrong with our poll must account for this change.

Since it was the turnout answers of Labour supporters that changed, the likeliest explanation is that in 2015 Labour supporters were more prone to exaggerate their likelihood of voting than they had been in 2010. This might have been because they were less likely to turn out than Conservatives in both 2010 and 2015, but in 2015 they were less willing to admit it. On the other hand, suppose Conservatives were *more* likely than Labour supporters to exaggerate their turnout in 2010 (which is not unlikely in itself, as they are drawn more from the groups who might be most unwilling to admit non-voting), but this bias was cancelled out by some other bias elsewhere in the data (for example, an over-representation in the sample of Labour supporters at the expense of non-voters – precisely the sample bias we suspect in 2015), so producing an accurate poll projection. Then a 2015 increase in Labour exaggeration to the same level as the Conservative exaggeration would have left us with a single, uncorrected bias instead of two errors that compensated for each other, which would also explain what we found. This would unite the two elements of the investigation, the sample bias detected by the BPC enquiry and the changing turnout responses we saw in our polls, into a single explanation. The BPC report concluded (Sturgis et al., 2016, pp. 46–47) that there was no significant differential in turnout misreporting in 2015, which would be more consistent with this second possibility.

Figure 1: “Certain to vote” by party in Ipsos MORI’s monthly polls 2008-15



Source: Worcester et al (2016, 225)

Why should the propensity of Labour supporters to exaggerate their turnout likelihood have increased? Here we can only speculate. In our regular monthly polls, we noted a very abrupt increase in the proportion of Labour supporters saying they were certain to vote quite early in the parliament, which was then maintained until the election (see Figure 1). We took this at the time to indicate that the advent of a Conservative-led government had increased the determination of Labour supporters to use their vote against it. Possibly it only increased their sense of social obligation to say that they were going to do so, without being powerful enough to affect their behaviour. Or perhaps a group of the public who are particularly prone to exaggerating their future turnout had switched their support to Labour from other parties. Or perhaps the increase in those intending to vote Labour was real almost until the last minute, but Labour’s poor campaign and the Conservatives’ effective negative messaging finally weakened their determination to turn out even though they did not admit this to pollsters or, probably, to themselves⁸. If so, the polls may have been more nearly right throughout the election than is now generally taken to be the case.

How can we put it right?

We have concluded, therefore, that the inaccuracy in our 2015 forecast requires a two-fold response, improving both sample representativeness and the projection of voting behaviour from the data, and

⁸ Previous research has found evidence of negative campaigning leading to a depressed turnout in the supporters of the candidate or party targeted in certain cases; see Kahn and Kenney (1999).

we have already begun to apply some of our conclusions to our regular “peacetime” polling, although this remains a work in progress.

Firstly, we needed to tackle the under-representation of the politically disengaged in our polls. The BPC report is right to identify this problem, but offers no blueprint for solving it. Probability sampling is not the answer for election polls (as the BPC report accepts): to do it properly would cost more than media clients are prepared to pay, and it is far too slow for this purpose in any case (even though it is the ‘gold-standard’ methodology for research projects where it is suitable.) The under-representation of the disengaged presumably reflects a response bias, as Taylor suggests – those who are less interested in politics are, unsurprisingly, less willing to take part in political polls and are also correspondingly less likely to vote. Our initial approach has been to tackle it indirectly, controlling population characteristics that we know are related to engagement. Two which can be clearly identified, with up-to-date independent population profiles against which our samples can be tested, are newspaper readership and educational attainment. On investigation, we have found that on both criteria the samples that we were drawing were somewhat skewed. We could simply have added both variables to our weighting scheme, but after experiments we found that this tended to cause an unacceptably high design effect (the reduction in effective sample size when substantial weights are applied to correct under-representation of groups in the raw sample). We have therefore also redesigned our quotas to take account of education, so that in our initial sample we interview enough members of the public who have no formal educational qualifications, and reduce our previous over-representation of graduates. This has consequences in terms of the efficiency of interviewing (it tends to take more calls to achieve a given sample size conforming to the quota, and therefore involves a higher cost per interview achieved), but – combined with the addition of readership and education weights – has had a significant effect on our sample profile and on our voting intention measurements which, had it applied in 2015, would have gone a substantial way towards making our final poll match the election result.⁹

⁹ We cannot directly test the effect of our new weights on the 2015 election poll, as it did not include all the questions needed to apply the weighting. However, we can test the effect on our aggregated monthly polls

We are nevertheless continuing to experiment, and may add other sample controls in due course. As always, the biggest obstacle to sophisticated weighting schemes is the availability of reliable data from which to set the weighting targets – and this is especially true when weighting for attitudinal rather than for demographic factors, for example by controlling directly for interest in politics, where population totals might easily change dramatically and unpredictably.

We have also considered how we might improve our turnout questions or the way we model turnout from them. There is no straightforward and obviously correct way of screening out the non-voters from a polling sample. As previously explained, our approach in 2015 and for several general elections before that was to ask “How likely will you be to vote... on a scale of 1 to 10...?”, and to treat only those who said they were “absolutely certain to vote”, 10 out of 10, as voters. Many other pollsters, in this country and in other countries, handle the task differently. Some use a similar question to ours, but apply it as a graduated scale rather than as a simple cut-off: for example, those rating their likelihood of voting at ‘9’, ‘8’ or ‘7’ might also be included in the final numbers but at a steadily decreasing weight (on the basis that some ‘7’s as well as some ‘10’s will vote, but that a smaller proportion of them will do so).

Another possibility is to include other questions that measure different aspects of political engagement and so offer an indirect indication of the likelihood of voting. We might ask, for example, “How interested are you are in politics?” or “How important, if at all, is the result of this election to you personally?” These answers could then be combined into a composite score for each respondent, and again this could then be applied as a simple cut-off or on a graduated basis with a lower engagement score being interpreted as implying a lower likelihood of voting, or even used in combination with an independent estimate of the probable turnout to pick out a pre-determined proportion of respondents. It is possible, too, to go beyond purely attitudinal questions, and incorporate factual questions about past voting behaviour (“Did you manage to vote in the most recent local or mayoral elections in your

between January and May 2016 (with a total sample size just over 5,000), which drew their samples in the same way. (After May 2016 the revised quotas were introduced.) Calculating voting intentions as we did in the final pre-election poll, applying the new education and readership weights on top of the existing weighting scheme increased Conservative voting intention share by 1.7 percentage points and decreased Labour share by 2.6 points.

area?") or something between the two ("Which of the following best describes how often you vote in General elections: I never vote, I rarely vote, I sometimes vote, I usually vote, I always vote or It depends?"). If answers to these questions are less volatile, or less subject to exaggeration for social desirability reasons than the simple likelihood of voting question, and still closely related to real future likelihood of voting, they might give us a more accurate turnout filter.

But there is no *a priori* likelihood that one of these models will be better than another, they can only be justified empirically: we can test which filters would have resulted in the best forecasts at past elections, or, better, we can draw on evidence from surveys such as the British Election Study which include validated turnout at respondent level, allowing the power of individual responses as predictors of voting to be tested directly. But this, of course, depends on the questions of interest having been included in past surveys: radically new approaches would initially be based entirely upon educated guesswork. Here we run into a perennial problem with British election polling, that experimentation is difficult because we can only adequately test the relationship between responses to survey questions and general election voting behaviour at a general election.

Many pollsters in other countries go further, and use likely voter models that impose a pre-determined pattern on turnout: for example, the relative turnout levels of different age groups, or even of those identifying with different parties, may be fixed on the basis of past experience rather than drawing on answers in current polls. Very different though these various approaches may seem, they share a common weakness: all rely on the assumption that the future will resemble the past, whether in detailed turnout patterns or in the relationship between survey answers and behaviour. This can become perilous in the face of any dramatic changes in political circumstances or voter psychology.

Our polling in the referendum on EU membership, in which we attempted to apply the lessons from our post-election investigation, suggests that we have largely solved the problem of drawing a representative sample but may still face difficulties in how best to allow for turnout. Our final pre-

referendum projection¹⁰ of the vote was 52% to remain in the EU (Murphy and Cecil, 2016), against a result of 48%. However, this projection was based on a new method of accounting for turnout, incorporating respondents' reports of their past voting behaviour and how important they believed the result would be; had we projected instead on the basis of those saying they were absolutely certain to vote, as at all recent general elections, we would have predicted 49% for "Remain", a satisfactory forecast.

But this merely emphasises the extra difficulties that pollsters are faced with in referendums. This referendum has no recent precedent¹¹ for either pollsters or voters, meaning that there was no past evidence to guide the pollsters in interpreting respondents' reports of their likelihood of voting. We chose to partly discount their declared determination to vote in favour of other measurements. That had the effect of making the demographic pattern of our projected turnout resemble the known pattern of past general election voting more closely. In particular, it depressed the projected turnout of young working class voters, essentially the group whose turnout we think we overestimated at the general election.

We now suspect this was the root cause of our error at the referendum, and that this group turned out more strongly than is their wont. Are we therefore also wrong to apply this revised turnout model at the next general election (and to our "peacetime" voting intention polls until then¹²)? It failed us in the referendum, but is that because it is fundamentally unsound or because general elections and referendums are different? This is a question that will be exercising us in the coming months.

¹⁰ We also made a further prediction, based partly on on-the-day telephone polling of voters after they had voted, which was not released until the polls had closed. As its methodology was necessarily somewhat different from our pre-election polling, we exclude it from consideration here.

¹¹ The UK held a national referendum in 2011 about the parliamentary voting system, but turnout in that referendum was 42%, whereas in the referendum on EU membership the turnout was 72%.

¹² The turnout filter we are currently using in our "peacetime" polling includes those who are '9' or '10' certain to vote, but only if they also say they "always" or "usually" vote at general elections or that "it depends". (In the EU referendum we added a further step, that they must consider the referendum outcome to be "very" or "fairly" important.) On the same January-to-May 2016 dataset referred to in note 9 above, using the revised weighting scheme, the impact of our new turnout filter as compared to the one used at the election was to increase Conservative share of voting intention by 1.5 percentage points and to decrease Labour share by 0.6 points. The cumulative effect of the weighting and turnout filter changes was therefore to increase the Conservative lead over Labour by 6.4 points, slightly over-compensating for our 5.5-point under-estimate at the election.

Final remarks

We are very grateful to Patrick Sturgis and his team for all their hard work, which has told us a good deal we did not know about why the polls had problems in 2015. Sadly, but unsurprisingly, they were unable to give us the one thing we sought the most, an easy way to stop it reoccurring. All diagnoses of poll error are to some extent tentative, because the evidence is never complete. Moreover, in prescribing a remedy we are once more faced with the question of whether the future will resemble the past. The analysis of polling methods, like polling itself, involves judgment as well as science. At Ipsos MORI we are working hard at finding a solution, as no doubt are other polling companies, but unfortunately we won't know for certain if we have succeeded until the next general election proves us right or wrong.

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.
- Kahn, K. F., & Kenney, P. J. (1999). Do negative campaigns mobilize or suppress turnout? Clarifying the relationship between negativity and participation. *American Political Science Review*, 93(4), 877-889.
- Market Research Society. (1994). *The Opinion Polls and the 1992 General Election: A Report to the Market Research Society*. London: Market Research Society. Retrieve from: <http://www.ncrm.ac.uk/polling/documents/The%20Opinion%20Polls%20and%20the%201992%20General%20Election.pdf> [accessed 22nd May 2016].
- Mouncey, P. (2016). Research methodology under the microscope: the interim findings of the BPC/MRS enquiry. *International Journal of Market Research*, 58(2), 164–168.
- Murphy, J. and Cecil, N. (2016). Brexit poll: Remain takes narrow lead in final EU referendum poll, *London Evening Standard*, 23 June. Retrieve from: <http://www.standard.co.uk/news/politics/brexit-poll-remain-secure-narrow-lead-in-final-eu-referendum-poll-a3278996.html> [accessed 30th August 2016].

- Noelle-Neumann, E. (1993) *The Spiral of Silence: Public Opinion - Our Social Skin*, 2nd ed. (Chicago: University of Chicago Press).
- Smith, T. M. F. (1983). On the validity of inferences from non-random sample. *Journal of the Royal Statistical Society. Series A (General)*, 146(4), 394-403.
- Sturgis, P., Baker, N., Callegaro, M., Fisher, S., Green, J., Jennings, W., Kuha, J., Lauderdale, B. and Smith, P. (2016). *Report of the Inquiry into the 2015 British General Election Opinion Polls*. London: MRS/British Polling Council/NCRM. Retrieve from: http://eprints.soton.ac.uk/390588/1/Report_final_revised.pdf [accessed 22nd May 2016].
- Taylor, H. (2016). Explaining Cameron's comeback (Book review), *International Journal of Market Research* 58 (3), 486-8.
- Worcester, R. (1996). Political polling: 95% expertise and 5% luck. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 5-20.
- Worcester, R., Mortimore, R., Baines, P., & Gill, M. (2016). *Explaining Cameron's Comeback*. London: IndieBooks.