POLLING ERROR IN THE 2015 UK GENERAL ELECTION: AN ANALYSIS OF YOUGOV'S PRE AND POST-ELECTION POLLS

DOUGLAS RIVERS AND ANTHONY WELLS

1. INTRODUCTION

The polls released by YouGov and all other pollsters prior to the 2015 UK general election significantly overestimated Labour support and underestimated Conservative support. For example, YouGov's final pre-election poll, fielded 4–6 May and released on Wednesday afternoon on the eve of the election, had Labour and Conservatives tied at 34%. In fact, Conservatives won with 37.7% of the vote, well ahead of Labour at 31.2%. The purpose of this report is to understand why we were wrong and what could be done to fix it.

We have divided our analysis into three types of error. There are, of course, many potential explanations for what happened, but they fall into three broad categories:

1. **Late swing.** There was nothing wrong with the pre-election polls. The samples were representative of the population and the results were accurate at the time the interviews were conducted, but people changed their minds after the polling was completed.
2. **Misreporting.** The samples were representative of the population, but respondents did not accurately report if or how they would vote.
3. **Sample composition.** The samples were unrepresentative of the British population and the weighting used was inadequate.

These are difficult analyses to perform. Some of the variables are difficult or impossible to measure. Voting is secret, so we can never definitively say whether someone has reported accurately how they voted. Further, although the size of the polling error—predicting a tossup when, in fact, Conservatives won by almost seven points—seems large, it is, in fact, a 3.7% error in the estimate of Conservative vote share. With sample sizes in the thousands, an error of this magnitude should be rare, but 3% or 4% effects are inherently difficult to analyze, even with quite large sample sizes. Nonetheless, we have attempted to do so and can summarize our main conclusions here:

1. There may have been a small late swing toward the Conservatives—perhaps one percent in magnitude, probably less—but it was much too small to explain the polling error. That is, something was definitely wrong in the pre-election polls.
2. YouGov's pre-election polls substantially overstated turnout, due to a combination of overreporting by respondents and samples with too many voters. We propose some improvements in our sample selection procedures and calibration of our likely voter models. However, overestimation of voter turnout was *not* the cause of the error in our pre-election polls: turnout was equally overstated for Conservative and Labour voters.
3. It is difficult to find either support for or rule out the possibility that respondents said they had voted Labour when in fact they voted Conservative. Even in the exit poll, which was much closer

to the actual outcome than the preelection polls, there appears to be a consistent underestimate of Conservative voting. The size is not large (about 1.4%) and this argument remains quite speculative.

4. We believe that the primary source of the polling error in the 2015 UK General Election was due to skews in sample composition. Most important was the distribution of respondents within the younger age groups, which skewed towards those with higher levels of interest in politics. The younger age group is more likely to vote Labour than Conservative, but as a whole is less likely to vote at all; if there are too many young people in the sample with high levels of political interest, the sample tends to have too many Labour voters. Second, YouGov's sample contained too few people over 70 years old, and those people vote disproportionately Conservative. Correcting for skews in age and political interest accounts for about 2.3% error of the 3.7% error in our pre-election poll. Adjusting for the other small skews in the sample would have increased the proportion voting Conservative by about half a percent. These skews account for most of the error in the pre-election poll.
5. The problem was probably exacerbated by the decision to sample from respondents who had completed a baseline political survey January–February. This decision was made in order to reduce sample volatility, but it had the side effect of excluding some panelists from our pre-election polls with less interest in politics than frequent panel responders.
6. There are other sample skews and methodological problems that also appear to have caused YouGov's pre-election polls to underestimate the Conservative vote: too few women, too many 2010 Labour voters, and inadequate regional weighting. Adjusting for each of these would have pushed the sample 0.2% to 0.7% in the Conservative direction, though the combined impact of these adjustments is likely to total less than 1%.

The panel methodology used by YouGov relies upon the collection of a large number of profile variables which can be used to select and weight samples to be representative of a population of interest. Conventional methods for weighting a sample on demographics and past vote would have eliminated the skews in these variables, but it is preferable to select a sample balanced on these factors instead of trying to repair the sample by weighting after the data are collected.

The presence of too many or too few persons in the panel of a particular type (*e.g.*, too many party members or too few people over age 70) does not prevent a representative subsample from being selected if the panel is large and diverse enough. With enough panelists of any type, these factors can be used for selection of a representative subsample. Most of the variables described above were already collected as profile data by YouGov and could have been used, but were not. The most important variable that was not part of our sampling methodology during the 2015 elections was interest in politics. Our profile data already contain two separate questions on interest in politics, but we have traditionally relied instead on a measure of newspaper readership that may be obsolete.

Many panelists join YouGov because of the high visibility of its election polls. To reduce sample volatility, respondents for the daily pre-election polls were chosen from those who had completed a political survey in January-February. This decision may have further increased the portion of these samples with high levels of political interest. Over the past year, YouGov has instituted new programs to attract panelists interested in topics other than politics to reduce panel imbalance. It will still be necessary to measure panelists interest in politics and to include this as part of our sample selection methodology in the future.

2. DATA

In this report, we have utilized four different datasets: (1) a collection of daily polls conducted between 30 March and 7 May on YouGov's UK panel, which were reported contemporaneosly; (2) a large recontact study of panelists after the election; (3) a vote validation study to verify the turnout of panelists; and (4) the British Election Study post-election survey, an area probability sample with a 56% response rate. More details about each data source are provided below.

2.1. **YouGov Daily Pre-election Polls.** Over the course of the campaign, YouGov interviewed approximately 2,000 persons per day. These panelists were selected from a pool of approximately 34,000 panelists who participated in political surveys conducted during January and February 2015. The January-February samples were selected using the procedures standard for YouGov political surveys in the past with quotas on age, region, gender, ethnicity (in London), education (for respondents under 25), newspaper readership, and party identification. The January-February pool of respondents was divided into ten groups who were interviewed on a rotating basis every ten days. The sample for each days was composed of those who had responded in the past 24 hours regardless of which group they belonged to or when they had been invited, so the daily samples contain a mixture of fast and slow responders.

A total of 63,233 interviews with 31,851 distinct panelists were conducted between 30 March and 7 May 2015 with this group of panelists. The daily samples of responding panelists were raked to age $\times$ gender, region, social class, newspaper readership baseline and voting intention (measured in January/February).

Respondents were asked how likely they were to vote on a scale ranging from one to ten. The scale value was divided by ten to convert it into an estimate of their probability of voting. The raking weight (described above) was multiplied the estimated turnout probability to obtain a likely voter weight.

2.2. **YouGov Post-election Recontacts.** After the election, YouGov interviewed 123,486 panelists and asked them how they voted, their interest in the campaign, general interest in politics, party identification, and similar questions. 27,315 of the 31,851 panelists in the daily pre-election surveys completed a recontact survey as part of this effort. These interviews were conducted between 8 May and 18 May.

2.3. **YouGov Vote Validation.** After the election, YouGov staff examined the electoral registers in 36 local councils to verify turnout of a subset of panelists. The councils were chosen partially on grounds of convenience (staff went to their own local councils), but was intended to be broadly representative. Outsiders were hired to examine registers in the North and Midlands. A list was compiled with the names and addresses of all YouGov panelists living within each council area who had completed the post-election survey. The canvaser identified each panelist's address on the register and recorded whether the name person was on the register at that address and, if so, whether they were recorded as having voted or not. Vote validation was performed for 5,284 panelists. Of these, 3,530 voted, 466 were found in the electoral register but did not vote, 307 were not registered at the address they provided, and 91 were not eligible to vote in the election.

2.4. **2015 UK Exit Poll.** The exit poll has the advantage of only including voters. The design samples one or two precincts from each of 133 constituencies. The sample of constituencies is not random (insofar as possible, the same precincts are selected in each election) and precincts are not selected at random from within constituencies. Further, the poll includes only a single question—how did you vote?—so it contains no demographic data. However, in recent elections, the poll has been quite accurate.

2.5. **BES Face-to-face Post-election Survey.** From 8 May to 13 September, the British Election Study conducted face-to-face interviews of eligible voters. BES used an addressed-based cluster sample design in order to be representative of all eligible voters residing in Great Britain aged 18 or older. Constituencies were stratified by county (and region within England). Within each country/region, constituencies were classified by party competition, defined as a combination of winning party and party competition from the 2010 election. The final stage of stratification was to sort the constituencies within each cell from the least to the most marginal. The constituencies were then selected with probability proportional to population size. The survey consists of 2,987 completed interviews with an overall response rate of 59.9%.

3. Late Swing

There was very little volatility in the daily polling estimates over the course of the campaign. Excluding undecideds, the percentage saying they intended to vote Conservative ranged from a low of 32.4% (on 26 April) to a high of 35.3% (on 10 April), but never reached the actual Conservative 37.7% of vote cast. Similarly, the poll estimates of Labour vote ranged from a high of 36.1% (on 13 April) to a low of 33.3% (on 5 May), but never fell to the 31.2% of vote actually received by Labour. There was hardly any trend and in nearly every daily sample the Conservative and Labour vote share were within a percentage point of one another.[1]

The pattern in polls conducted by other organizations was similar (and similarly mistaken). Fieldwork for YouGov's final poll finished during the day on Wednesday, with fieldwork for the last polls from ICM and MORI continuing to Wednesday evening. Late swing covers the possibility that these polls were accurate at the time, but people changed their minds between answering the final polls and actually voting, perhaps even changing their minds in the polling booth itself.

If there had genuinely been a late swing we might expect to see evidence of movement towards the Conservatives in the fieldwork conducted on the last day of the campaign, and should certainly expect to see change in fieldwork conducted on the day itself or evidence of people switching their vote in post-election recontact surveys of those people who were interviewed before the election.

Looking at late polling the British Election Study Continuous Monitoring Survey (also conducted by YouGov) showed Conservative leads on the final two days of fieldwork, but the sample size for the BES was smaller (about 1,000 per day) and it produced volatile day-to-day numbers so this is not conclusive. ICM have released a day-by-day breakdown of the data from their final poll which shows no movement towards the Conservatives, nor did comparison between Populus's Tuesday and Wednesday data.

[1]The lack of day-to-day volatility is due in part to the design of these surveys, since the group of panelists invited each day is a stratified subsample of the initial group interviewed in January-February, with the same distribution of demographics and party identification.

YouGov's post-election recontact surveys do show a small (about one percent) increase in Conservative support. Comparing respondents' stated vote intention in daily pre-election polls with how they said they actually voted after the election finds most—about six out of seven respondents—saying they had voted the same way as the told us they intended in their pre-election interview. However, as shown in Table 1, almost no one shifted from Labour to Conservative and this was almost equally offset by a small number of Conservative voters switching to Labour. The bulk of the movement comes from people moving from one of the other parties, from being undecided, or changing their likelihood of voting. The average net increase in the Tory vote is just above one percent and most of this comes from switching in pre-election samples before the last week.

In summary, there is little evidence of a substantial late swing to the Conservatives and certainly nothing like a seven percent swing. Late swing can, at best, account for a small proportion of the overall polling error in this election.

4. MISREPORTING

The second broad category of potential error is people giving pollsters answers that do not accurately reflect their voting intentions. This is distinct from the first category—late swing is people accurately reporting their current responses at the time but then changing their minds at a later time. Misreporting is people giving an answer that is inaccurate when it was given.

Colloquially misreporting has often been referred to as the problem of 'shy Tories', referring to social desirability bias: people saying they are voting Labour to signal that they are a socially responsible person. This risks being a misleading description. While shyness—reluctance to admit a voting intention that may be seen as unfashionable or selfish—is one possibility, it may be more a case of denial (of respondents not even being willing to admit their intentions to themselves), of answering the wrong question (for example, a respondent giving their first preference when they are actually voting tactically for someone else), rather than social desirability bias. Secondly it does not necessarily affect only the Conservatives. It could be as much a case of over-enthusiastic Labour as shy Tories. Thirdly the 'shy Tory' adjustment adopted by ICM after the 1992 election dealt only with people saying don't know or refusing to answer the question, the potential issue is not limited to people who don't give a voting intention, and could include people giving a false voting intention.

Misreporting can involve either turnout or party preference. The first is false reporting of people's likelihood to vote—that is, whether they will vote or not, as opposed to who they vote for. The second type of misreporting concerns which party people will vote for. We consider each in turn.

4.1. **Overreporting of Turnout.**

Nearly all surveys overestimate voter turnout, often by a large amount. Nearly all survey organizations utilize some kind of likely voter model. This can be as simple as asking voters how likely they are to vote and to exclude respondents from the sample who say they are unlikely to vote (a 'likely voter screen'). In the UK, YouGov uses 'likely voter weighting.' Respondents place themselves on a zero to ten scale representing their likelihood of voting. The scale value selected is divided by ten to produce an estimated probability of voting for

Date	Stable	Labour-Conservative	Conservative-Labour	Other-Conservative	Other-Labour	Conservative-Other	Labour-Other	Net Change
6 April	81.0	0.4	0.4	5.6	6.8	2.4	3.4	-0.3
7 April	83.7	0.3	0.8	4.8	4.6	2.5	3.4	0.6
8 April	78.8	0.2	0.4	5.9	6.3	4.0	4.5	-0.0
9 April	79.1	1.0	0.8	5.9	5.3	2.8	5.1	3.0
10 April	82.5	0.7	0.5	5.4	4.4	2.4	4.1	2.9
11 April	80.4	0.7	0.7	6.5	4.4	3.4	3.9	2.7
12 April	81.9	0.9	0.3	5.7	4.6	3.2	3.5	2.0
13 April	82.6	0.4	0.1	5.4	4.4	3.1	3.9	2.1
14 April	82.3	0.6	0.5	4.8	5.3	2.9	3.6	0.2
15 April	84.1	0.5	0.2	5.9	4.2	2.2	2.9	2.8
16 April	82.9	0.5	0.4	5.9	5.0	2.3	3.0	1.8
17 April	84.6	0.5	0.1	4.6	4.0	2.8	3.3	1.7
18 April	82.2	0.4	0.6	5.2	4.2	3.1	4.3	1.9
19 April	81.6	0.6	0.5	4.9	5.5	3.1	3.7	-0.0
20 April	84.3	0.3	0.6	4.5	3.5	3.4	3.4	0.6
21 April	82.9	0.6	0.7	5.7	4.4	2.7	3.1	1.7
22 April	84.2	0.8	0.6	4.1	4.7	2.2	3.4	0.7
23 April	85.5	0.3	0.6	5.6	3.6	1.3	2.9	3.3
24 April	86.7	0.5	0.2	4.4	4.3	1.5	2.4	1.2
25 April	86.2	0.7	0.2	4.6	4.6	1.6	2.2	1.2
26 April	86.2	0.5	0.6	3.9	5.2	2.0	1.7	-1.6
27 April	87.3	0.5	0.3	3.6	4.4	1.9	2.0	-0.5
28 April	85.4	0.6	0.6	4.0	4.4	2.5	2.6	-0.3
29 April	86.1	0.7	0.4	3.9	4.0	2.2	2.6	0.5
30 April	87.4	0.3	0.5	3.4	4.6	1.7	2.0	-1.0
1 May	87.5	0.5	0.2	3.7	3.8	1.5	2.7	1.5
2 May	87.0	0.3	0.4	3.7	4.2	2.1	2.3	-0.4
3 May	89.5	0.4	0.3	3.9	2.5	1.5	1.9	1.8
4 May	88.7	0.4	0.2	3.2	3.8	1.2	2.4	0.8
Average	84.2	0.5	0.4	4.8	4.5	2.4	3.1	1.1

TABLE 1. The table displays the percentage of respondents in each day's pre-election poll who gave the same answer in their pre- and post-election interviews (stable) or switched from their pre-election voting intention. There was a small net movement (about 1.1%) toward Conservatives in the post-election recontact. survey. Conservative-Labour switching was relatively rare, but there was quite a bit of switching from 'other' (which includes non-voters, parties other than Labour or Conservatives, and (in the pre-election interview) undecideds.

each respondent and the sample is weighted by this quantity.[2] Unfortunately, the implied turnout rate that comes from this procedure (the mean of the estimated turnout probabilities) was 89.7%, while actual turnout in the election was just 66.1%.

[2]To be precise, the post-stratification weight (which attempts to correct for sample selection bias) is multiplied by the estimated probability of voting to obtain the likely voter weight.

It is unclear whether the exaggeration of turnout is due to misreporting (people saying they are certain to vote when there is a chance they won't), inadequacy of the procedure for estimating turnout probabilities (some form of calibration other than just dividing by ten could be used), or selection bias (the sample is composed of too many voters). We will show that over-reporting appears to be relatively modest in the UK.[3] Some improvement in calibration of turnout probabilities is feasible, but the overwhelming problem is that the samples contained too many actual voters.

4.1.1. *Post election recontact.* Comparing people's answers in pre-election surveys to whether they reported voting in post-election recontact surveys does not find large number of people saying they ended up not voting and does not suggest it had a disproportionate impact on either party. Labour voters were not disproportionately likely to report that they didn't end up voting. Looking at YouGov's final poll around 2% of both Labour and Conservative voters said they ended up not voting when they were re-contacted after the election.

The question about likelihood of voting from the pre-election interview is highly correlated with respondents self-reported voting behaviour. 99% of respondents who said they were 10/10 certain to vote reporting having done do after the election, compared to 83% of those who said they were 7/10 likely to vote, 51% of those who said they were 5/10 likely to vote and 14% of those who said they were 0/10 certain not to vote. There is some nonlinearity in the voting rates (not captured by the simple 'divide by ten' rule), but the benefit of making this adjustment would be relatively small.

4.1.2. *Vote validation study.* For a subset of respondents we have tested their reported post-election voting behaviour against the marked electoral register to see whether respondents are falsely claiming to have voted when they did not. This work is ongoing, and will eventually be supplanted by a much larger validation exercise from the British Election Study. Early results are indicative though. Based on vote validation of 5,284 respondents in 36 constituencies, we found that roughly 10% of those who said they voted in a post-election interview did not actually vote: about 5% were registered at the address they gave us, but were not marked as voting, or were marked as ineligible to vote, and a further 5% were not registered at the address they provided, though it is possible that some of these respondents had moved or had registered and voted at an alternative address.

	Reported vote (post-election interview)		
Validated turnout	Conservative	Labour	Other party
Voted	88.0%	89.4%	90.8%
Did not vote	6.3%	5.7%	4.6%
Not registered	5.7%	4.9%	4.6%
n	761	893	852

TABLE 2. There is not much difference in the validated turnout rate for self-reported Conservative and Labour voters. In fact, over-reporting is slightly *lower* for those not voting Conservative.

[3] In contrast, careful studies in the US find larger amounts of over-reporting than we found in the UK.

Furthermore, misreporting rates are not correlated with *how* respondents say they voted. Conservative voters are slightly more likely to report having voted when they did not, compared to Labour or other party voters, but the difference is small (and, in any event, would have caused the survey to overestimate Conservative vote, not underestimate it).

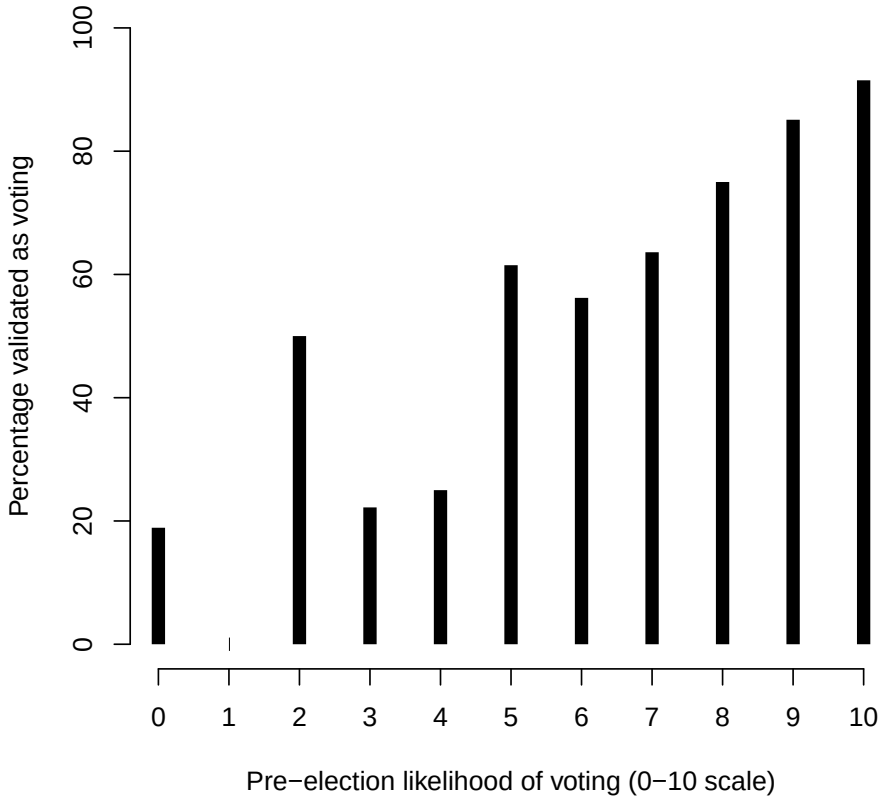


FIGURE 1. Results of voter validation for 1,004 respondents in YouGov pre-election polls for whom we had a valid address. Although there is a strong correlation between respondents' self-assessed likelihood of voting in pre-election polls, validated turnout rates are below those assumed in likely voter weighting at the upper end of the scale.

4.1.3. *Calibration of likely voter models.* We conducted a further calibration of the likely voter question using validated turnout. The results are shown in Figure 1. For scale values between 6 and 10, there is a roughly linear relationship between the respondents reported likelihood of voting and their validated voting rate. The sample sizes become quite small at the lower end of the scale (only 27 respondents in the validation study who said their likelihood of voting was between 1 and 4), so more data would need to be collected to estimate an adequate turnout model.

Finally, the reliability of the likelihood of voting question increases as the election nears. As shown in Figure 2, respondents with low expressed likelihood of voting before 15 April were more

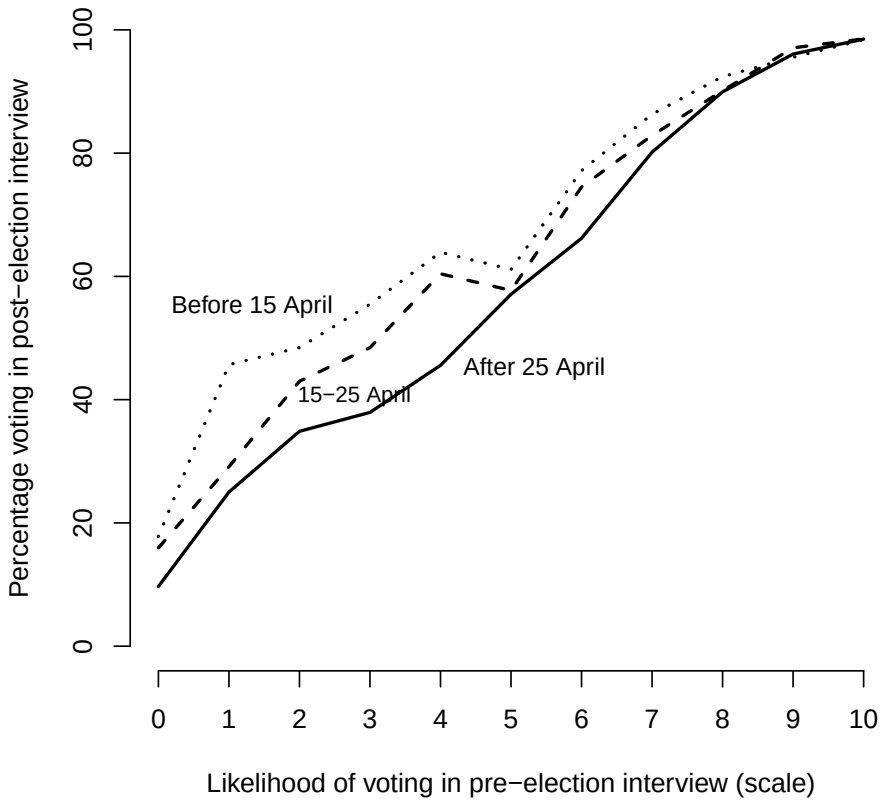


FIGURE 2. As the election approaches, likelihood of voting is a better predictor of actual turnout. Earlier in the campaign, respondents with low self-assessed likelihood of voting turn out at moderate rates. The phenomenon doesn't occur for respondents who say they are very likely to vote.

likely to report voting in their post-election interview than respondents with the same scale value interviewed later. It would be advisable to adjust the likely voting procedure depending on how near the election is.

4.1.4. *The impact of overestimating turnout.* Overestimating turnout will not necessarily result in inaccurate predictions of vote share. In a poll that attempts to correctly measure how people will vote, a failure to correctly represent those people who are not going to vote anyway may be irrelevant. This explains why this problem, which may have got worse but is certainly not new, has not necessarily lead to inaccurate results in the past.

The primary risk is if the overstatement is not linear, with turnout being overestimating to a different extent amongst different political and demographic groups. For example, if a sample contained the correct proportion of older people who voted, but overrepresented young people who voted, the final sample of likely voters would be skewed towards the young. If a sample

underrepresented the sort of young people who don't bother to vote, but weighted up the sort of young people who do in order to replace them, the final sample would be skewed.

In order to tell if turnout is skewing results we need reliable figures for actual turnout amongst different demographic groups. There are no official figures on turnout and sources drawn from other faulty opinion polls (such as the widely used Ipsos MORI aggregates) risk containing the same errors that caused the inaccuracy. There are two other potential sources—the first is to use the few genuine stratified random sample surveys (primarily academic surveys like the BSA and the face to face element of the BES), the other is to use inferred evidence from election results themselves.

The British Election Study (BES) and British Social Attitudes Survey (BSA) both use high-quality random, stratified face to face samples and both find reported levels of turnout that are far closer to reality than regular opinion polling.

In the pre-election wave of the 2010 BES 53% of people said they were 10/10 certain to vote and the implied turnout from the weighted 0-10 scale was 75%. In their post-election wave 77% of people reported having voted. In a validation exercise using the market electoral register 67% were found to have actually voted at the address given. In the 2010 wave of the BSA 69% of respondents reported having voted in the general election. Actual turnout in the 2010 general election was 65%, suggesting that in terms of election participation the BES and BSA samples are broadly representative.

Looking at age the BSA and BES have both consistently shown a strong correlation between age and likelihood to vote. While this pattern was also present in our final poll, it was not nearly so marked. In BES and BSA data the proportion of retired people who vote is around thirty percentage points higher than the proportion of under 30s who vote, the difference in our data is only ten percent.

The same pattern is present in other demographic breaks. In BSA and BES data unemployed respondents are around twenty percent less likely than employed respondents to report having voted at the preceding general election. In YouGov's data unemployed respondents are only six percent less likely to report having voted in the 2015 election. The pattern of younger, less affluent voters being less likely to vote which is present in random samples and implied by actual election results is far less stark within our figures.

Given that our validation data suggests only a relatively small proportion of this is from people falsely reporting whether they voted or not this suggests that young people and people from less affluent backgrounds in our polls are not representative and are too likely to turnout to vote.

4.2. **Misreporting of party support.** YouGov's recontact survey showed only evidence of a small change from pre-election voting intentions. Therefore, we must assume that either there was little misreporting of voting intentions, or people who gave false answers before the election continued to give the same false answers after the election. Without such direct evidence of false reporting, we need to look at other indirect evidence that may provide answers.

Several hypotheses around false reporting have been presented. One is that people voted Conservative despite not being a supporter of the party because they trusted them more on leadership and the economy, feared a Miliband minority Labour government or a Labour government propped

	Age						
Survey	15-24	25-34	35-44	45-54	55-64	65-74	75+
British Social Attitudes Survey							
2001	42.3	54.8	64.7	77.4	73.8	81.1	82.7
2005	39.8	55.5	66.4	76.2	80.4	84.3	86.4
2010	45.5	49.0	67.6	74.6	85.1	90.1	87.2
British Electoral Study (self reported)							
2001	53.5	55.4	67.5	79.1	77.8	86.3	86.3
2005	48.3	55.3	72.6	74.3	84.7	85.1	88.7
2010	60.8	67.2	75.6	79.3	82.0	85.6	90.9
British Electoral Study (verified)							
2001	45.2	53.2	64.0	75.3	76.4	83.3	82.6
2005	35.4	47.8	63.1	66.3	74.9	73.4	74.2
2010	48.1	53.3	66.5	74.4	72.0	77.5	82.2
Average	46.5	54.6	67.6	75.2	69.5	83.0	84.6
YG 2015 final call	84.6	84.3	88.1	91.1	93.4	95.8	95.8
YG 2015 post election	87.2	86.5	90.0	92.4	94.8	96.7	96.3

TABLE 3. Entries are percentage turnout rates by age group in different surveys and different years. There is a weaker correlation between age and turnout in YouGov's samples than in other surveys.

up by the SNP, yet would not admit this to pollsters. This hypothesis is largely based upon anecdotal evidence and other polling questions giving answers that were far more favourable to the Conservatives than topline voting intention questions were.

A second is that that polls did not, in some way, properly reflect the voting experience because they were too dissimilar to ballot papers. According to this hypothesis including candidate names in the polls would have picked up local tactical considerations and personal votes that the polls themselves missed and, therefore, produced different figures at the national level. Most notably this has been suggested by Jim Messina, the former-Obama pollster who worked with David Cameron in the 2015 election.

4.2.1. *Reluctance to admit voting Tory.* Throughout the 2010-2015 parliament David Cameron maintained a strong lead over Ed Miliband as respondents' preferred Prime Minister. Once the economy returned to growth in the Conservative party also held a consistent lead on which party people most trusted to run the economy.

The apparent contradiction between the Conservative lead on leadership and economic trust and the Labour party's lead in voting intention did not go unnoticed. It was widely commented on and analysed by pollsters and other political commentators. Many pollsters and commentators highlighted this contradiction as a reason to expect movement towards the Conservative party as we headed towards the general election—movement that, in the event, never happened. Peter Kellner

repeatedly pointed out how unprecedented it would be for a party to win when it trailed on both these measures.

With the benefit of hindsight it is tempting to point to this contradiction as a signpost that the polls were wrong, but had the result been different the discrepancy could have been realistically explained by Labour's better ratings on terms of party image—they were consistently seen as being more in touch, while the Tories continued to be seen as the party of the rich. The apparent contradiction could also be explained through the unusual politics of coalition—typically people who say they would vote for a party also pick its leader as 'best Prime Minister', but under the coalition Liberal Democrat voters tended to tell pollsters that David Cameron would make the best Prime Minister, rather than Nick Clegg.

All polling companies currently take stated party preference at face value. There are no assumptions that someone who says they will vote Labour might, actually, vote Tory (nor can we see how such an approach could ever be viable). Unlike the marked register there is no independent way of verifying which party individual poll respondents actually ended up voting for.

The exit poll model performed much better than the pre-election polls. The exit poll methodology is based upon estimating swings in precinct vote between pairs of elections. The sample consists of one or two precincts in each of 133 constituencies which are selected somewhat haphazardly, so it's not an ideal design for estimating vote proportions in a single election. However, the data shown in Table 3 below is at least suggestive that Conservative vote was also underestimated in the exit poll, albeit by a smaller amount than in the pre-election polls. A special computation (provided by Professors John Curtice and Stephen Fisher) shows the error in the exit poll estimate of Conservative vote in 2015 and in the swing in Conservative vote between 2010 and 2015. Both the swing between 2010 and 2015 and the 2015 vote was underestimated by about 1.4%. This happened in six out of seven regions, which suggests that the effect is not due to the choice of precincts or constituencies within regions.

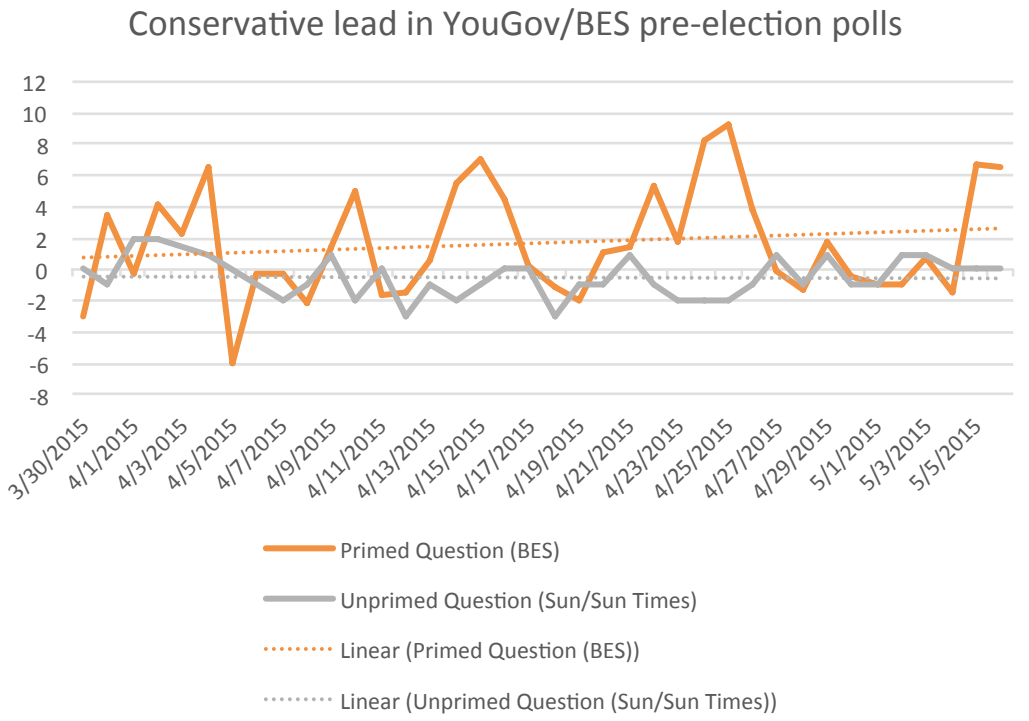
	Actual Conservative Vote			Exit poll Conservative Vote		
Region	2015 Vote	2010-2015 Swing	Number of Constituencies	2015 Vote	2010-2015 Swing	Number of Constituencies
London	30.1%	0.3%	73	26.9%	-1.4%	16
South	49.5%	2.4%	138	49.2%	1.3%	39
East	49.2%	2.0%	60	48.3%	2.5%	13
Midlands	38.9%	1.4%	157	38.1%	0.1%	21
North	30.2%	0.2%	105	28.7%	-3.1%	29
Scotland	14.9%	-1.8%	59	14.5%	-2.3%	10
Wales	27.2%	1.2%	40	21.0%	-4.6%	5
National	37.7%		632	36.3%	-1.4%	133

TABLE 4. The exit poll also under estimated the Conservative vote in every region. The largest underestimates occurs in Wales, but this is based on data from only five precincts.

On the other hand, the location of the underestimate is different from that in YouGov's pre-election polls, where the largest misses were in the South. Nonetheless, the pattern is suggestive that there might be some reluctance to admit voting Tory, though not enough to account for all or even most of the pre-election polling error.

The BES face-to-face survey slightly over-estimated the Conservative margin. These data were fielded over a long period with the sample proportion of Conservative vote rising toward the end of the field period. This could be due to differential response speed or recall bias. If the latter, it suggests that Tories were becoming less shy over time.

4.2.2. *Priming respondents.* In the aftermath of the general election several claims were made about private polling for the political parties having shown more accurate results. Most notably James Morris of Greenberg Quinlan Rosner claimed that their private polling for the Labour party had shown the Conservatives with a bigger lead than published polls because they took a different approach, asking people other political trackers before the voting intention question, such as what the most important issue facing the country is. The claims about Labour's figures has been met with some scepticism from other figures within the Labour party, and it appears it may actually have referring to GQR's private polling for Labour earlier in the Parliament. Nevertheless it is worth looking at the potential for changing the question order.



Traditionally voting intention is asked first to remove any possibility of preceding questions influencing people's answers, viewing any impact as inevitably negative. The alternative view is that politically neutral questions that precede the voting intention question may have a positive effect by making people think more carefully about who they would actually vote for at a general election, rather than asking them 'cold'. It is also possible that respondents' desire to appear consistent in

a survey could counteract their reluctance to admitting they would vote Conservative—*i.e.* having already said they prefer the Conservative party's leader and trust them more on major issues of the day, people are less reluctant about saying they'd vote for them.

We can test the potential effect of priming using BES data and YouGov data from the BES pre-election waves, the parallel CMS and YouGov daily polling, and YouGov post-election experiments.

In the first three waves of the BES, conducted in 2014, the placement of the voting intention question was randomized, appearing at either the beginning or part way through the survey. In their analysis of these tests Jon Mellon and Chris Prosser show that asking the question later in the survey slightly increased the level of Conservative support, but not enough to be statistically significant.

In the BES online campaign survey the voting intention question was asked part of the way through the survey, following questions on the most important issue facing the country, which party the respondent would trust to handle it, level of interest in the election, whether parties would do a good or bad job in government, whether respondents like or dislike the party leaders. This data was collected from the YouGov panel so can be directly compared to the figures from the daily YouGov voting intention figures for the Sun and Sunday Times, where voting intention was asked first. The daily figures from the BES survey have a smaller sample size so are more erratic, but on average they are again slightly more Conservative than the YouGov daily polling. Over the whole of the short campaign period the BES primed data on average shows a Conservative lead two points higher than the YouGov unprimed data.

Finally since the general election YouGov has experimented with asking voting intention before and after priming questions on alternate days. This took place over around twenty surveys to give robust findings, and again found putting priming questions before voting intention produced a slightly larger Conservative lead, though by only a single point.

	Vote Intention				
Question Order	Conservative	Labour	Lib Dem	UKIP	Other
Vote question first	40.4	30.1	6.0	13.6	10.1
Priming questions first	40.4	29.2	6.2	13.6	10.3
Difference	0.0	-0.9	0.2	0.0	0.2

TABLE 5. Entries are the percentages intending to vote for each party in surveys with different question order. The priming effects are negligible: vote intention is the same whether asked first or after the priming questions.

All of these experiments produced only minor effects and taken alone the differences would not be significant. However across all three experiments the pattern was the same, with priming consistently producing figures that are around one percentage point better for the Conservatives. By itself, priming cannot explain the polling error.

4.2.3. *Tactical Voting.* One possibility other than 'shyness' is people falsely reporting tactical or local voting decisions. For example, someone who has a preference for the Labour party but votes

tactically for the Liberal Democrats given the position in their own constituency. When answering a poll, do those people answer Labour as the party they really support, or Liberal Democrat as the party they will actually cast a vote for? Both seem plausible, and either could be given by a respondent seeking to give a genuine and honest answer, but unsure which question the pollster really wanted them to answer.

In terms of correctly predicting national vote shares in a general election the desired answer is the party respondents actually intend to cast a vote for, even if this is different from their true preference, but which question are respondents actually answering?

Lord Ashcroft's widespread polling of constituencies over the last two years highlighted a technique of asking two voting intention questions, a general one and one asking specifically about people's own constituencies (a technique originally used in YouGov polls for PoliticsHome in 2008 and 2009). This technique was intended to pick up tactical votes and personal votes that it was thought the standard question may miss.

YouGov's campaign polling included this "locally prompted" question allowing us to compare results. While the locally prompted question produced significantly different results in areas with a strong Lib Dem presence, at a national level it made little difference. On the core issue of measuring Labour and Conservative support the two different wordings produced identical results, and on Liberal Democrat support the locally prompted result was marginally less accurate:

	Vote Intention (percent)				
Question wording	Conservative	Labour	Lib Dem	UKIP	Green
Standard wording	34	34	8	12	5
Constituency prompt	34	34	10	12	4
Difference	0	0	2	0	-1

TABLE 6. Entries are the percentage voting for each party in a survey with either standard wording or the constituency prompt. The standard wording and constituency prompt produce identical estimates of the Conservative and Labour vote shares.

An alternative approach to this issue has been to include candidate names in the question. This was cited by the Liberal Democrats during the election campaign as a reason why Lord Ashcroft's opinion polls may be underestimating their position (in hindsight, of course, they were in fact overestimating Liberal Democrat support!). Jim Messina, the US pollster working for the Conservative party also cited not mentioning candidate names as a potential source of the error. Clearly it is impossible to go back and ask pre-election questions using candidate names if they were not included at the time, but there are some examples of other polling companies using candidates' names within opinion polls.

In Survation's three final polls for the Daily Mirror and the Mail on Sunday they asked three versions of the voting intention to the same sample: a standard question, a constituency prompted question, and a mock ballot paper including the names of the candidates standing in the respondent's own constituency. The effect of the alternative wording and the prompting by individual candidate names was again minimal, and where it did make a difference, it made the topline results

less accurate, not more accurate. Survation also asked a final election poll that only prompted by candidate names which produced a result far closer to general election, but this used an atypical sampling technique (a telephone poll sampled by taking numbers from lifestyle databases rather than RDD) so it is impossible to confidently conclude that the different result was down to prompting rather than the sampling.

Survey	Conservative	Labour	Lib Dem	UKIP	Green
Final *Mirror/Survation* Poll					
Normal prompt	33	33	9	16	4
Candidate prompt	31	31	10	16	5
Difference	-2	-2	1	0	1
Penultimate *Mirror/Survation* Poll					
Normal prompt	33	34	9	16	4
Candidate prompt	31	32	10	15	5
Difference	-2	-2	1	-1	1
Final *Mail on Sunday/Survation* Poll					
Normal prompt	31	34	8	17	4
Candidate prompt	29	33	9	17	6
Difference	-2	-1	1	0	2
Average difference	-2.0	-1.7	1.0	-0.3	1.3

TABLE 7. Entries are the percentage voting for each party in surveys by different organizations with different question wordings. Other surveys show somewhat larger effects of the candidate prompt.

Theories based upon constituency or candidate prompting do not assume a dishonest respondent, but a respondent answering a slightly different question from the one the pollster intends. This has been suggested as a possible reason for polling error by Mark Textor, the partner of Lynton Crosby — that people gave pollsters their first preference of Liberal Democrat or Labour, but then actually voted tactically to ensure a Conservative majority. It should also be possible to test by asking respondents questions on what they meant by their response. Post-election YouGov experimented with asking respondents how they voted, and then asking them to specify if they voted for that party as their first choice, voted for that party tactically despite actually supporting a different party, or supported that party but actually voted tactically for someone else.

The poll did indeed find a significant proportion of people saying they voted tactically - 16% of respondents said the party they voted for was different from the party they actually supported (whether for tactical or candidate related reasons). However, in reporting how they voted at the general election the overwhelming majority of these people gave the party they physically cast a vote for, not the party they actually supported. Only 1% of respondents said they had answered with a party they supported, but didn't actually vote for it.

In conclusion, it appears as if people who are voting tactically or on local factors are correctly reporting these voting intentions to pollsters. Alternative approaches to asking the voting intention question itself by prompting people to think of their own constituency or including candidate

	Party respondent said they voted for (percent)					
	Conservative	Labour	Lib Dem	UKIP	Green	Total
Reason for voting for party						
As first choice	83	78	60	91	78	81
As tactical vote or for candidate	13	19	33	7	9	15
Total vote for	97	97	93	98	87	96
Supported this party but didn't vote for them						
Voted tactically for a different party	1	1	1	1	2	1
Didn't actually vote for them	1	0	0	0	0	0
Total supporting but not voting for	2	1	1	1	2	1
None of these	1	2	6	2	11	3

TABLE 8. One to two percent of respondents, when asked, report voting tactically or otherwise not voting for the party they preferred. About the same nummber of Labour and Conservative voters said that they had voted for a party other than their first preference.

names make little difference at a national level, and if anything appear to make the data less accurate. Prompting by candidate would, anyway, not be feasible to adopt outside the three weeks immediately running up to an election.

5. SAMPLE COMPOSITION

Neither late swing nor misreporting provide a satisfactory explanation for the polling error in 2015. We now turn to the third category of explanation: lack of representativeness in our pre-election polling samples. We normally select and weight our polling samples to be 'nationally representative' in terms of age, gender, social grade, newspaper readership, and party ID. Demographic targets are obtained from Office of National Statistics (ONS) and newspaper readership and party ID from surveys which produced accurate results in previous elections. The distributions are for the population of all adults, not voters.

The conventional approach to election polling, used by YouGov and other survey organizations, is to first select and/or weight a sample to known demographic distributions of the population and then either to identify likely voters as either a subset of that sample (using a likely voter screen) or to weight the sample based on estimated turnout probabilities (likely voter weighting). However, since the pre-election poll samples contain too many voters (as demonstrated by the vote validation study), weighting the sample to general population targets is not justified. The problem is that weighting a sample of (mostly) voters to adult population targets (only two-thirds of whom are actually voters) may distort the composition of the voter subsample.

For the purposes of election polling, the main concern is that the subsample of voters be representative of all voters. If nonvoters are under-represented in the sample, this is not a serious

problem, since nonvoters are effectively discarded by likely voter screens or weighting. It is, however, critical that *voters* in the sample be representative of the population of *voters* in the population (at least after sample selection and weighting).

In the UK, limited information about the characteristics of voters is available. The UK exit poll, for example, contains no information about voters other than the constituency in which they live and which party they voted for.[4] Thus, we are forced to rely upon other surveys, to evaluate the sample composition of voters. For present purposes, we compare the YouGov daily election samples to the BES face-to-face survey. The BES design is not ideal: sample size is moderate, the field period was quite long, and it overstates actual turnout, though not by as much as the daily pre-election polls.

All comparisons that follow use pre-election vote intention, but voters are identified using their post-election self-report of having voted (not their pre-election likelihood of voting). Similar results would be obtained if we weighted by the pre-election turnout probability or used the post-election reported party vote.

5.1. **Demographic Composition of the Sample and Electorate.** Voters in the YouGov sample tend to be younger than those in the BES sample. This is especially evident in the top age group (75 or older) which contains 12.2% of the BES voters, but only 3.6% of YouGov voters. (Voters are identified according to their self-reported turnout in the post-election recontact interview.) Since in both polls, the oldest voters have the highest level of support for Conservatives, this tends to depress the Conservative vote in the YouGov sample. If we reweighted YouGov voters to match the age distribution in the BES, the Conservative lead over Labour increases from 0.2% to 2.0%. In other words, almost half of the discrepancy between the YouGov poll and the election outcome would disappear if voters in the sample were weighted to this target.

We can do a similar calculation for gender. The difference between the gender distribution in YouGov's preelection polls and the BES is relatively small. YouGov had slightly more male voters than female voters, while BES had the reverse. However, there was almost no gender gap in the YouGov data (men were 0.2% more likely to vote Conservative than women), so reweighting by gender has virtually no impact on the party vote estimate.

Education is another variable where there are noticeable differences between the YouGov voter sample and the BES voter sample. YouGov's panel (like most online panels) is short of people with the lowest level of educational attainment and the YouGov voter sample has only about a third as many voters with no qualifications as BES. At the other extreme, YouGov has substantially more respondents with a university degree. However, the impact of reweighting voters to match the BES education distribution has only a small impact and actually would have the effect of reducing the (weighted) sample proportion of Conservative voters.

In previous election surveys, the regional distribution of YouGov's samples did not appear to be skewed, so no effort was made to weight to regional targets. The regional distribution of the pre-election samples in 2015 does not differ markedly from that in the BES. However, regional variations in voting patterns are substantial, so reweighting the sample to match the BES regional distribution results in a 0.7% increase in the Conservative vote estimate. In small regions, the BES

[4]In contrast, in the US, the exit poll and the Current population survey provide detailed information about the composition of the electorate.

	Sample Distribution		Conservative Lead	
Age	YouGov	BES	YouGov	BES
18-24	10.7	9.1	-12.1	-16.1
25-34	13.6	13.3	-5.8	-18.6
35-44	16.1	14.3	-6.6	-9.1
45-54	16.7	20.0	-5.6	16.6
55-64	21.9	16.0	2.0	17.1
65-74	17.3	15.1	17.1	25.6
75+	3.6	12.2	25.1	27.8
Unadjusted lead			0.2	8.1
Adjusted lead			2.0	
Difference			1.8	

TABLE 9. Entries in the first two columns are the weighted percentage of sample voters in each age category in the YouGov daily polls and the BES face-to-face post-election survey. Entries in the last two columns are the weighted percentage voting Conservative in each age group in each survey. 'Unadjusted lead' is the Conservative lead over Labour using the standard weights for each survey. 'Adjusted lead' reweights the YouGov poll using the distribution of age in BES. The YouGov sample has substantially fewer respondents over 75 years old, compared to the BES sample. Reweighting the YouGov sample to have the same age distribution as BES would increase the Conservative margin by 1.8%.

	Sample Distribution		Conservative Lead	
Gender	YouGov	BES	YouGov	BES
Male	49.6	48.2	0.3	6.7
Female	50.4	51.8	0.1	9.3
Unadjusted lead			0.2	8.1
Adjusted lead			0.2	
Difference			-0.0	

TABLE 10. The YouGov sample has 1.4% more women than BES, but post-stratifying on gender makes has almost no effect, due to the lack of a gender gap in the YouGov data. There is a small gender gap in the BES data.

cluster design will have high variances, which probably accounts for the anomalous result in the Northeast.

In summary, YouGov's election samples could be improved somewhat if demographic targets were available for likely voters. The introduction of demographic targets to likely voter weighting requires reliance on other surveys or statistical modeling. We recommend that both approaches be investigated.

	Sample Distribution		Conservative Lead	
Education	YouGov	BES	YouGov	BES
No qualifications	7.3	21.5	-3.5	4.4
GCSE	27.2	25.7	4.3	13.1
A-level	24.5	26.3	-0.7	9.3
Univesity degree	41.0	26.5	-0.4	5.0
Unadjusted lead			0.6	8.1
Adjusted lead			0.1	
Difference			-0.5	

TABLE 11. The YouGov sample has fewer respondents with no qualifications or GCSE, but there is only a weak correlation between education and 2015 vote, so adjusting for education has little effect.

5.2. **Non-demographic Variables.** Next, we consider some non-demographic characteristics of our voter samples in 2015: 2010 vote and interest in politics.

5.2.1. *2010 vote.* The distribution of 2010 vote is known for the general population in 2010 from election returns. However, some 2010 voters have died and others were too young to vote in 2010 but have become eligible in the meantime. Further, we do not know how many 2010 voters did not vote in 2015 or how many 2010 nonvoters did vote in 2015. We have relied on the BES face-to-face sample, where 2010 turnout and vote is based upon the respondents recall in 2015.

	Sample Distribution		Conservative Lead	
2010 Vote	YouGov	BES	YouGov	BES
Conservative	31.7	34.2	69.5	81.8
Labour	25.5	28.9	-68.4	-60.3
Liberal Democrat	20.9	11.8	-14.8	-6.6
UKIP	2.3	1.7	9.5	7.0
SNP	1.8	2.0	-3.2	-4.2
Plaid Cymru	0.4	0.4	-4.8	12.4
Other	2.8	2.0	-8.9	-5.3
Nonvoter	7.9	12.6	-7.3	-4.3
Too young	6.9	6.4	-8.9	-18.6
Unadjusted lead			0.2	8.1
Adjusted lead			0.7	
Difference			0.5	

TABLE 12. The BES sample contained almost half as many 2010 Liberal Democrats and more 2010 non-voters, though this may be due to using recall rather than contemporaneous reports of 2010 behavior.

5.2.2. *Political interest.* All polling depends upon some degree of voluntary cooperation, so that it is not surprising that the people who take election polls have above average interest in politics. The proportion of people on YouGov's panel who are members of political parties is higher than the actual figures for party membership. The proportion who claim to have watched political events like the leadership debates during election periods is higher than the official viewing figures.

Panel skews in political interest are well-known. In UK political surveys, it has been standard practice to weight samples by newspaper readership, which includes a category for respondents who read no newspapers. These, presumably, were people who paid little attention to politics. Over time, patterns of news consumption have shifted and this variable is now only weakly correlated with other measures of political interest. Approximately a quarter of those who say their attention to politics rates as a 10 on a 0-10 scale report not reading *any* newspaper. At the other end of the scale, a larger proportion do not read any newspaper, but it is still only 42%.

There are obviously no official measurement of levels of political interest, but there are some questions that have been consistently asked on the BSA and BES random face-to-face surveys. The BSA includes a five point verbal scale question on interested respondents are in politics, the BES repeats this question and also includes a 0-10 scale on how much attention respondents. We will focus on the 0-10 attention to politics scale found in the BES.

As would be expected, levels of interest and attention towards politics before the election do strongly correlate with turnout; samples that over-represent people with a high level of political interest are also likely to overstate turnout, and that correcting this over-representation may help address turnout issues.

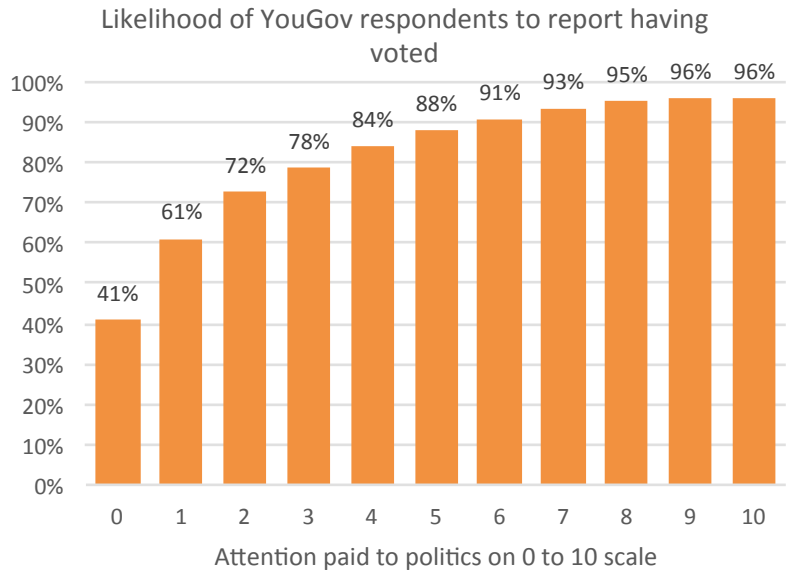


FIGURE 3.

Crucially levels of political attention also seem to be relevant to how people voted at the general election beyond the impact on turnout. Breaking down our data based upon respondents' party identification and the levels of attention they pay to politics shows that party identifiers with a high

level of political engagement switched their votes differently than those with a low level of political engagement.

Amongst Conservative identifiers vote switching was broadly similar across different levels of political engagement. Amongst Labour identifiers, Labour ID respondents with low political attention were twice as likely to switch to voting UKIP in 2015 as respondents with Labour ID & high political attention. The impact amongst Liberal Democrat identifiers was the starkest: Lib Dem respondents who paid a high amount of attention to politics were most likely to switch their vote to Labour, Lib Dem respondents who paid a low amount of attention to politics were more likely to switch their vote to the Conservatives.

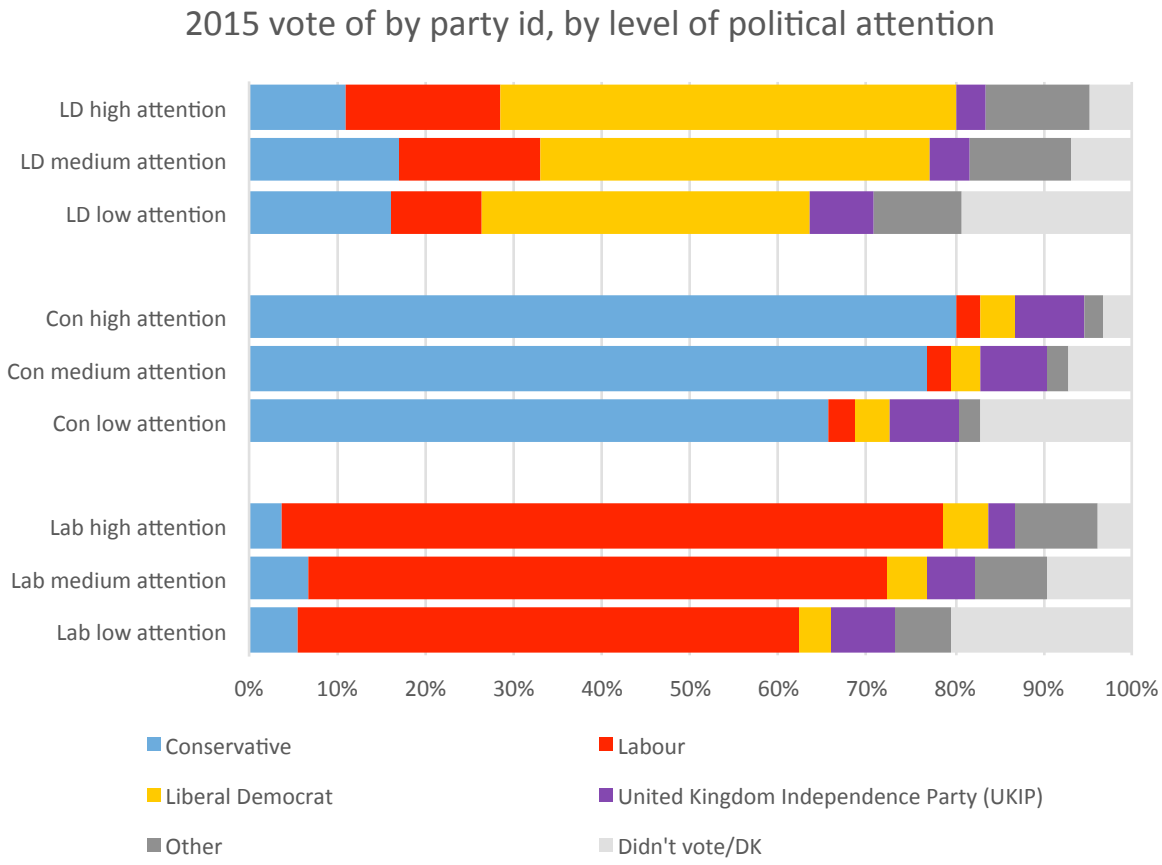


FIGURE 4.

After weighting, only 7% of the voters in YouGov's 2015 daily polling samples were between 0 and 3 on the ten point BES attention to politics scale, compared to 21.2% in the BES face-to-face survey. At the other end of the scale, YouGov's daily sample of voters had more than twice as many voters in the top two categories (9-10) of the scale, compared with BES voters. (See Table 13.) Our sample of voters was heavily skewed toward those who pay close attention to politics, even after weighting. However, the effect of reweighting by this measure of political interest is not large—it would make the sample only 0.7% more Conservative—since both the top and bottom ends of the scale tend to support Labour with the middle support the Tories.

	Sample Distribution		Conservative Lead	
Attention to Politics	YouGov	BES	YouGov	BES
0	0.9	3.2	-13.8	-18.1
1	1.0	2.6	-10.1	-32.8
2	2.2	5.4	4.2	-5.8
3	2.9	10.0	6.0	-0.2
4	3.4	7.6	-1.1	2.2
5	8.6	14.6	1.1	12.5
6	10.5	14.5	1.4	18.3
7	18.6	17.4	4.2	19.9
8	22.3	13.0	4.5	14.4
9	13.2	5.4	-3.2	11.0
10	16.4	6.3	-7.8	-9.6
Unadjusted lead			0.3	8.2
Adjusted lead			1.0	
Difference			0.7	

TABLE 13. There are large differences between the YouGov and BES samples on attention to politics (measured on a 0 to 10 scale). However, adjusting for this variable alone only increases the estimate of Conservative vote by 0.7%, due to the curvilinearity of the relationship between attention and Conservative voting.

Are skews in political interest relatively harmless? The answer is 'no', since there is an important interaction between political interest and age. The panel overrepresents young (18-24 year old) voters with high attention to politics by a factor of almost four; the degree of over-representation of high interest respondents in other age groups is smaller. Similarly, the underrepresentation of young voters with low attention to politics is more severe than in any other age group. Skews in age and political interest are both present and the worst case is young voters with low levels of interest.

The good news, as shown in Table 14, is that within categories of age and political interest, Conservative voting rates were similar in the BES and YouGov samples. This means that if a sample was selected or weighted on *both* age and political interest, that the vote estimates would tend to be similar. And, indeed this is what happens when the YouGov daily polling samples are weighted to the combination of age and attention to politics, as shown in Table 15. The table shows the effect of reweighting the YouGov daily sample to have the same proportion of voters in the 12 categories formed from crossing age (18-24/30-44/46-64/over 65) by attention to politics (low/medium/high). Just reweighting by the combination of these two variables raises the Conservative lead from 0.3% to 2.5%—correcting about two-thirds of the total error.

5.2.3. *Panel effect and panel quality.* If poll respondents are too engaged, is this a function of the sort of people who join the YouGov panel, or it is a result of them being panellists? Do we recruit people who are abnormally interested in politics to begin with, or does the repeated participation in surveys lead to initially normal panellists becoming too interested in politics?

	Attention to politics					
	YouGov			BES FTF		
Age	Low	Medium	High	Low	Medium	High
18-29	25.0	25.8	23.8	21.0	25.3	27.6
30-44	33.5	30.8	31.4	30.3	37.3	25.0
45-64	30.2	36.8	29.8	35.1	47.0	37.0
65+	38.2	47.9	42.5	40.4	51.9	46.3

TABLE 14. Entries are the percentage of persons voting Conservative in either YouGov's daily polling samples or the BES face-to-face post-election survey. Attention is measured on a 0-10 scale which was divided into three categories: low (0-4), medium (5-7), and high (8-10). Within age groups, the relationship between attention to politics and Conservative voting is similar across the two samples.

		Sample Distribution		Conservative Lead	
Age	Attention to Politics	YouGov	BES	YouGov	BES
18-24	Low	1.7	6.5	-12.8	-31.6
	Medium	5.5	6.4	-12.8	-15.4
	High	8.7	2.2	-9.7	-4.2
30-44	Low	2.7	5.8	-1.3	-17.0
	Medium	10.2	11.0	-6.1	3.1
	High	11.6	4.8	-5.9	-28.7
45-64	Low	4.3	10.3	0.8	5.2
	Medium	14.8	15.9	4.6	27.8
	High	19.7	9.8	-5.9	11.2
Over 65	Low	1.9	6.3	13.0	14.0
	Medium	7.2	13.2	23.5	31.5
	High	11.8	7.9	16.2	28.3
Unadjusted lead				0.3	8.2
Adjusted lead				2.5	
Difference				2.3	

TABLE 15. Adjusting for the combination of age and attention to politics increases the estimated Conservative vote share by 2.3%. This is about two-thirds of the polling error in the 2015 YouGov daily polls.

We have data already recorded on when panellists joined the YouGov panel and on how they were recruited — either organically through the YouGov website, through being referred by existing YouGov panellists or through paid recruitment (largely paid online advertisements). If panel-effect was a significant problem we should expect to find people who had been on the panel for a long period of time to have become too politically engaged and to return lower quality data.

Another potential concern is that people who have sought out the YouGov panel and joined under their own motivation may be too political, and be joining for party political motivations, particularly given the often political nature of the YouGov website.

In breaking down our final call sample though, neither of these seem to be case. In fact it reveals the opposite pattern — respondents who had been members of the YouGov panel for several years were no more politically engaged and provided data that better reflected the actual election result. Panellists who were recruited organically through the YouGov website were less political, and provided higher quality data than those who had been recruited through paid advertising.

Dividing our final call sample into two based on how long they had been on the panel and weighting each half separately, amongst those panellists who had been recruited more recently Labour had a two point lead, among those who had been on the panel for several years the Conservatives had a three point lead.

	Conservative	Labour	Lib Dem	UKIP
Newer recruits only	33.6	35.4	8.7	11.5
Long standing panellists only	35.9	33.2	8.1	12.5
Difference	2.3	-2.2	-0.6	1.0

TABLE 16. Entries are the percentage of panelists voting for each party.

Comparing levels of political attention again does not produce any evidence of panel effect. The pattern of response is almost identical between older and newer panellists, except for a slightly higher proportion of new respondents reporting an extremely high level of interest in politics.

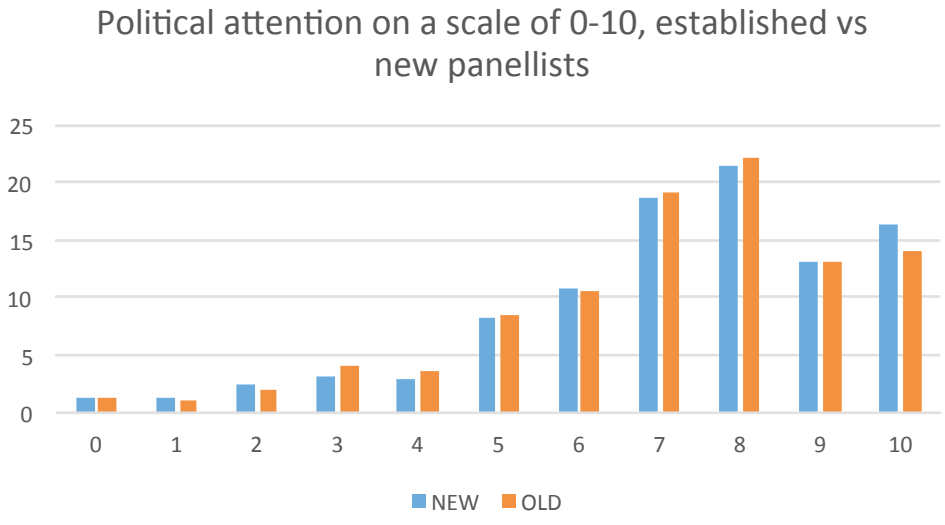


FIGURE 5. Established and new panelists do not differ significantly in terms of attention to politics.

This evidence suggests that panel effect is not a factor in the error (quite the opposite!), but raises a different issue of recruitment quality — that the sort of people who are joining the YouGov poll are producing lower quality, less accurate data. This is supported by the observations of the YouGov panel teams — long established YouGov respondents and respondents who have been organically recruited provide better quality data on other measures too, such as being willing to tick more boxes on BrandIndex surveys.

It is unclear why more established respondents would have produced a lower level of Labour support. One possibility is that that they are simply answering more honestly — some straws in the wind suggest that their answers are more accurate and internally consistent. For example, among long-standing respondents 75% of those who during the campaign said they trusted the Conservatives the most on the economy voted Conservative, among new respondents only 69% of those who trusted the Conservatives the most on the economy voted Conservative. Looking at the validated vote check, 93% of long-standing respondents who said they voted actually did so, 88% of newer respondents — the sample size is presently too small for this to be significant, but it warrants rechecking when we have more data.

This also corresponds with some of the wider explanations for the failure of the polls in 2015 — panel effect always seemed an unlikely cause for the polling error as telephone pollsters, immune to panel effect, had exactly the same error. But if the ultimate cause of the error was the population becoming less willing to take part in polls — something evidenced by the falling response rates and growing difficulties telephone pollsters face in getting people to agree to interviews — then this could affect both telephone and online polls.