ESOMAR | Events

Are We Getting Worse at Political Polling?

Jon Puleston

ESOMAR
Office address:
Atlas Arena, Azië Gebouw
Hoogoorddreef 5
1101 BA Amsterdam
Phone: +31 20 664 21 41
Fax: +31 20 664 29 22

Email: customerservice@esomar.org
Website: www.esomar.org

Publication Date: September 2017
ESOMAR Publication Series Volume Congress 2017
ISBN 92-831-0296-7

Copyright

Published by ESOMAR, Amsterdam,
The Netherlands
Edited by: Deborah S. Fellows

About ESOMAR

ESOMAR is the global voice of the data research and insights community, representing a network of 35,000 data professionals.

With more than 4,900 members from over 130 countries, ESOMAR's aim is to promote the value of market and opinion research in illuminating real issues and bringing about effective decision-making.

To facilitate this ongoing dialogue, ESOMAR creates and manages a comprehensive programme of industry specific and thematic events, publications and communications, as well as actively advocating self-regulation and the worldwide code of practice.

ESOMAR was founded in 1948.

About ESOMAR Membership

ESOMAR is open to everyone, all over the world, who believes that high quality research improves the way businesses make decisions. Our members are active in a wide range of industries and come from a variety of professional backgrounds, including research, marketing, advertising and media.

Membership benefits include the right to be listed in the ESOMAR Directories of Research Organisations and to use the ESOMAR Membership mark, plus access to a range of publications (either free of charge or with discount) and registration to all standard events, including the Annual Congress, at preferential Members' rates.

Members have the opportunity to attend and speak at conferences or take part in workshops. At all events the emphasis is on exchanging ideas, learning about latest developments and best practice and networking with other professionals in marketing, advertising and research. CONGRESS is our flagship event, attracting over 1,000 people, with a full programme of original papers and keynote speakers, plus a highly successful trade exhibition. Full details on latest membership are available online at www.esomar.org.

Contact us

ESOMAR

ESOMAR Office:
Atlas Arena, Azië Gebouw
Hoogoorddreef 5
1101 BA Amsterdam
The Netherlands
Tel.: +31 20 589 7800

Email: customerservice@esomar.org
Website: www.esomar.org

Are We Getting Worse at Political Polling?

Jon Puleston

Introduction

This paper is a summary of analysis conducted of an international database of 31,310 polls from 473 elections and voting events across 40 countries around the world from 1936 through to 2017 compiled by Kantar.

When examined at a global level polls are generally very accurate, the average error of polls conducted within seven days before an election is +/-2.5%.

The aim of this paper, though, is to help you understand why there can be differences in the size of errors seen in different types of election in different countries around the world and to educate on the things you should be looking out for when evaluating any political poll.

Polling companies around the world are judged, fairly or not, by how accurately their polling projections predict the outcome of election results. Polling companies can be easily criticised for not predicting accurately enough the end results of elections and vote, but how do we judge what accurate actually is? The polling industry has been subject to some severe criticism recently for what were seen as mistakes in not predicting the outcome in three recent prominent voting events: the 2015 UK general election, the Brexit referendum and the 2016 US general election. But how big were these "mistakes"? Where they indeed mistakes at all, were they bigger or smaller than you would expect from similar elections?

It perhaps depends on how you calculate it and what you compare it against and partly the confusion and misconceptions arise because there are no established international standards for comparing the accuracy of polls. So the wider aim of this paper is to identify the best method to compare international polling accuracy and establish some international polling accuracy benchmarks.

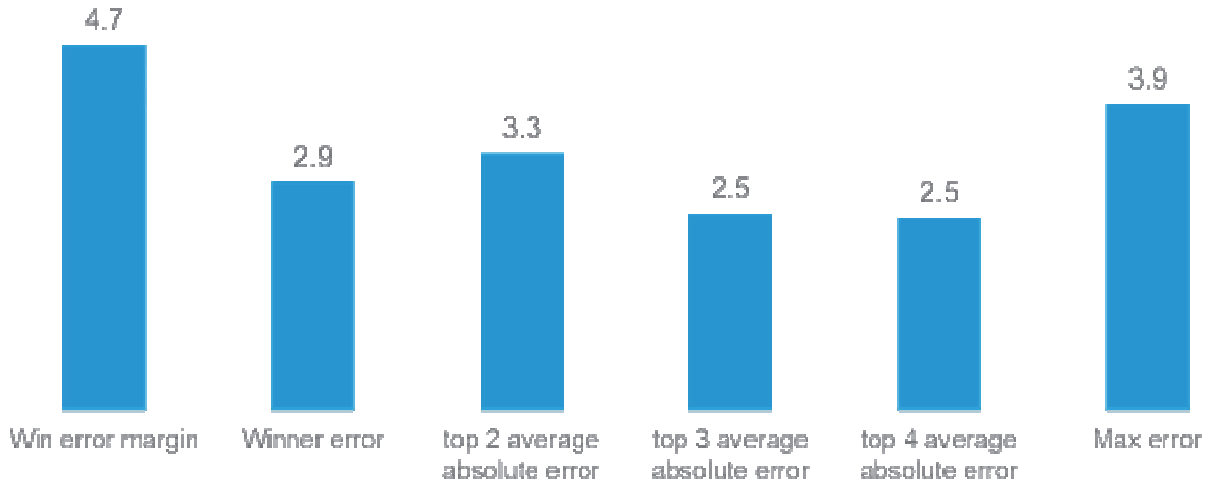
Measuring the accuracy of polls – the different ways in which this can be done

When making international comparisons about the accuracy of voter opinion polls the first question to tackle is what measure you should use to compare results? There are various means by which polling accuracy can be assessed, and lots of debate about which is the best and fairest measure to use. Some of the methods are a lot more complicated than others. The details of all the main established techniques and how they are calculated are outlined in the appendix, but in essence these are the six common ways in which polling accuracy tends to be evaluated.

1. *Win Error*: The error in measuring the share of vote of the winning party or candidate, e.g. if the polls estimated the winner would get 55% of the vote and the end result was 57%, the win error would be 2%.
2. *Win Margin Error*: The error in forecasting the size of the gap between the first and second place parties in the vote compared to the poll. e.g. If the poll estimated that the winner would get 55% and the second place would get 45% but in the vote the result was 57% and 43% the win error would still be 2% but the win margin error would be larger at 4%.
3. *Average Absolute Error*: the average size of error (be it positive or negative) of the top parties' or candidates' votes. This could be the top two, three or four parties or all parties gaining over a certain share of vote. So taking the above example if the error of the first party is +2% and the error of the second party is -2% if we ignore the + or – the average absolute error is 2%
4. *Average Absolute Marginal Error*: The absolute size of margin error between each of the main parties or candidates which is then averaged. This method is a little more complicated to work out: Imagine there were three parties and the poll forecast a vote share of 50%, 30%, and 20% but the vote ended up 47%, 32% and 21% the margin error between the 1st and 2nd party is 5%, between the 1st & 3rd is 4% and between the 2nd & 3rd is -1% . The absolute average is 3.33%

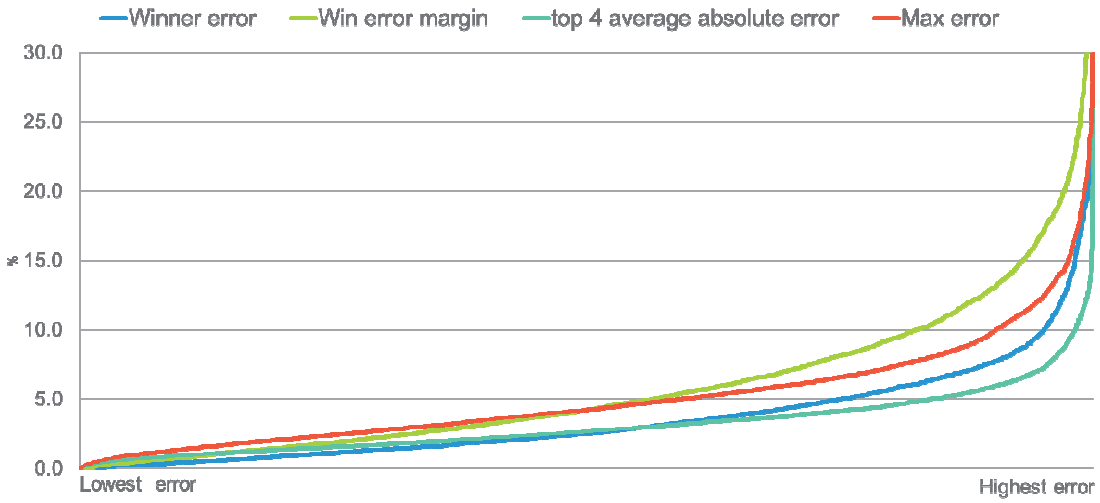
5. *Max Error:* The largest single measured error of any party or candidate
6. *Outcome Error*: Simply whether or not the poll predicted the outcome

Figure 1. Different types of error measurement = Different scales of errors for different elections



Base: Average error for all polls conducted within seven days before an election

Figure 2. Distribution of error



Which method is best to use for international comparisons?

Different countries tend to rely on different systems of error measurement based on the character of elections being measured.

For example in the USA where they have presidential elections with two main candidates, the focus tends to be on the Win Margin Errors. Countries with multi-party legislative elections tend use the Average Absolute Error for the main parties contesting the election and the Average Marginal Error between parties.

The number of parties or candidates included in these averages varies too depending on how many parties typically contest an election or gain a significant voter share. In first past the post elections, where the candidate or party with the largest number of votes wins, there tends to be fewer contesting parties to affect the outcome, the error relating to the top three or four parties are commonly aggregated, but in proportional representation elections where many more smaller parties (sometimes in numbers reaching double digits) gain a measurable vote share, they may average out the errors of all the parties above a threshold minimum voter share, perhaps 5%.

In proportional representation elections where the vote share of every party may be important there is a sometimes a preference for identifying the Max Error, the single biggest error that impacted on the outcome.

The difficulty of cross-comparing methods from different types of votes

The problem with cross-comparing results across different types of election is that no one measure tells the whole story or is perfect for every type of election.

To illustrate this, compare the two election scenarios illustrated in table 1. The first is an imaginary American election with two lead candidates and the second a European election with many parties. The win error for both elections as you can see is exactly the same, but the Average Absolute Error comparison for the European multi-party election comes out lower because the smaller error from the smaller parties reduces the overall average. So using average error like this might not be seen as a totally fair comparison.

Table 1. Two election scenarios

Scenario 1: US Presidential election	Poll	Vote	Abs error
Candidate 1	51	55	4
Candidate 2	46	42	4
		Winner error	4.0
		Average absolute error	4.0

Scenario 2: Euro Multi-party election	Poll	Vote	Abs error
Party 1	35	39	4
Party 2	30	26	4
Party 3	18	20	2
Party 4	11	9	2
Party 5	5	3	2
Party 6	1	3	2
		Winner error	4.0
		Average absolute error	3.0

You could potentially resolve this by including in the data from the two fringe candidates in the US election (scenario 1a below), but as they only tend to pick up 1 or 2% of the vote, this then now somewhat artificially reduces the average error of the US poll to below the average error from the multi-party election which again does not seem like a totally fair comparison either.

Table 2. Two election scenarios: 1a and 2

Scenario 1a: US presidential election	Poll	Vote	Abs error
Candidate 1	51	55	4
Candidate 2	46	42	4
Candidate 3	2	1	1
Candidate 4	1	2	1
		Winner error	4.0
		Average absolute error	2.5

Scenario 2: Euro Multi-party election	Poll	Vote	Abs error
Party 1	35	39	4
Party 2	30	26	4
Party 3	18	20	2
Party 4	11	9	2
Party 5	5	3	2
Party 6	1	3	2
		Winner error	4.0
		Average absolute error	3.0

This might lead you to conclude that Winner Error might be a better solution for cross comparing different elections? But the challenge with focusing on Winner Error comes in comparing different types of multi-party elections (which make up the vast majority of international votes). Over 80% of international polls involve three or more candidates or parties. For these polling events the errors of the smaller parties can be far more relevant and important than the Winner error per se. Take for example Scenario 3 below where a 3rd party candidate in the polls ends up winning. In this situation looking simply at Winner Error fails to tell the full story and so presents a misleading picture.

Table 3. Two election scenarios: 2 and 3

Scenario 2: Euro Multi-party election	Poll	Vote	Abs error
Party 1	35	39	4
Party 2	30	26	4
Party 3	18	20	2
Party 4	11	9	2
Party 5	5	3	2
Party 6	1	3	2
		Winner error	4.0
		Average absolute error	3.0

Scenario 3: Euro Multi-party election	Poll	Vote	Abs error
Party 1	30	29	1
Party 2	28	29	1
Party 3	26	30	4
Party 4	11	4	7
Party 5	5	2	3
Party 6	1	5	4
		Winner error	1.0
		Average absolute error	3.3

Looking at how different polling error measures correlate

Examining the correlation between different types of error measurement techniques across the body of all 30,000+ polls, it turns out that at an aggregated level there is in fact fairly high correlation between all the different error measurement techniques. Whilst there is no one perfect method for comparing polling data on an international scale, the scenarios cited above are actually extreme theoretical situations that don't often present themselves in real elections.

On average the different error measurement techniques cross-correlate at c0.8. This is perhaps not unsurprising: if there is a large error in the measuring the winning party you would expect there would be a correspondingly large error seen in measuring the smaller parties as the two are linked.

The case for the Top Four Average Absolute Error

Of all the different methods of measuring error though some do correlate slightly more consistently than others - see table 4. The measure that appears to have highest level of cross-correlation with the other error measurement techniques is the Top Four Average Absolute Error. This correlates at c0.8 using Winner Error, c0.9 using Win Margin Error and c0.93 with the Max Error.

Table 4.

	Winner error	Win error margin	Top 2 average absolute error	Top 4 average absolute error	Max error	**Average**
Winner error		0.83	0.73	0.84	0.84	**0.81**
Win error margin	0.83		0.76	0.87	0.86	**0.83**
top 2 average absolute error	0.7	0.8		0.8	0.77	**0.76**
top 4 average absolute error	0.8	0.9	0.8		0.93	**0.86**
Max error	0.84	0.86	0.77	0.93		**0.85**

This would point to this being the best measure to use for international comparison.

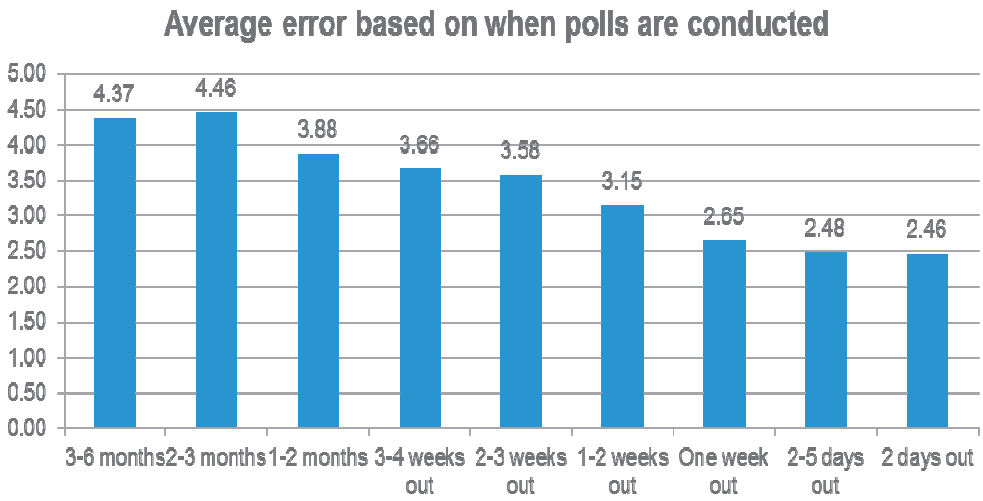
So for these reasons, the main polling accuracy figures quoted in this paper are based on the Top Four Average Absolute Error but we also are providing the average error ranges for some of the main alternative error measurement techniques for comparison purposes. These details are at the end of this paper.

What period is best to use for performance comparisons?

The next important question when comparing the accuracy of any polling event is what polling periods should be included and excluded based on when the polls were conducted. Generally speaking the accuracy of polls will steadily

increase the closer you get to the election event. Figure 4, based on time period analysis of all national level polls from the database, illustrates this. Polls conducted three months before an election will have an average error roughly twice the size of poll conducted in the last week of the election.

Figure 4. Average error based on when polls are conducted



The reasons for this probably don't need too much explaining, many of us make up our minds over the course of an election campaign and our opinions can change based upon campaigning activities.

For comparative purpose, we would therefore want to exclude event based factors that might mask the accuracy of the polling techniques themselves, so ideally you would want to include polls that were conducted as close to the election as possible when as many people as possible will have made up their minds.

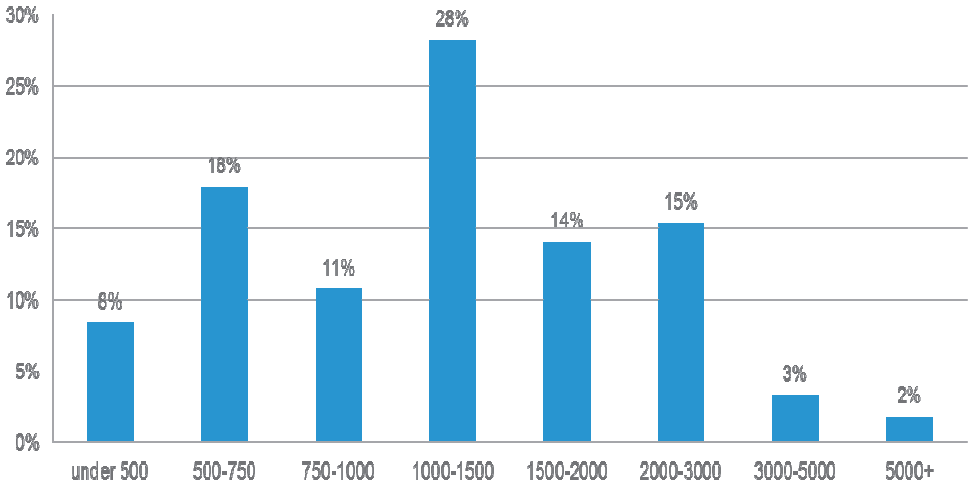
This decision though is complicated by two factors. Firstly the period it takes to gather polling data, as some polling methods can take several days to sample and so there is a limit to how close to an election you can conduct some type of polling. The second issue is the fact that in some markets polling in the few days immediately before an election is not allowed.

So for fair comparison at an international level we have chosen to base most of our comparisons on polls conducted within a seven day period before an election, of which a total of around 1,700 polls exist in the Kantar database.

What is a minimum acceptable sample size for a political poll?

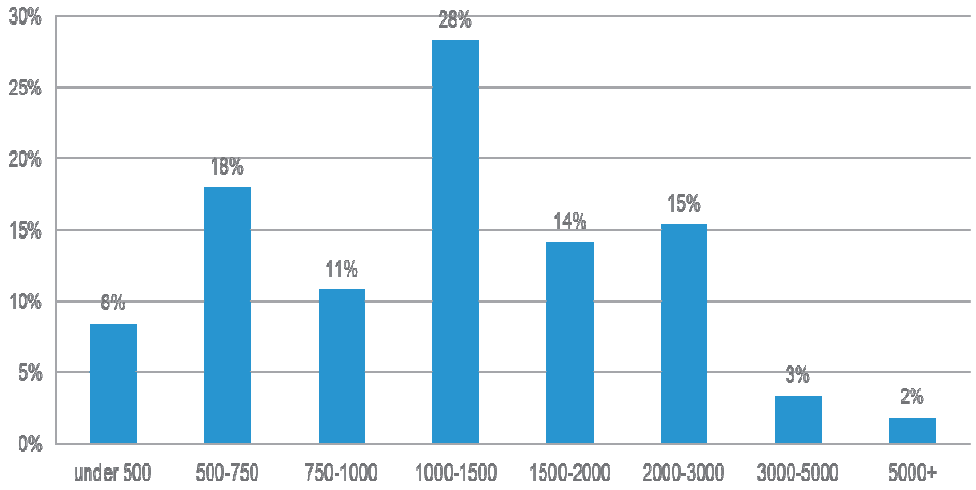
Whilst there is no official standard for sample size based on analysis of the database, the average sample size for a national opinion polls is almost exactly 1,000 respondents.

Figure 5. Average sample size of polls in the database



If you examine variation in poll error based on sample size error levels you will understand why polling accuracy based on sample starts stabilise at a point somewhere between 750-1,000 responses, so sampling below this figure will introduce sampling errors.

Figure 6. 7 day out polling error by sample size

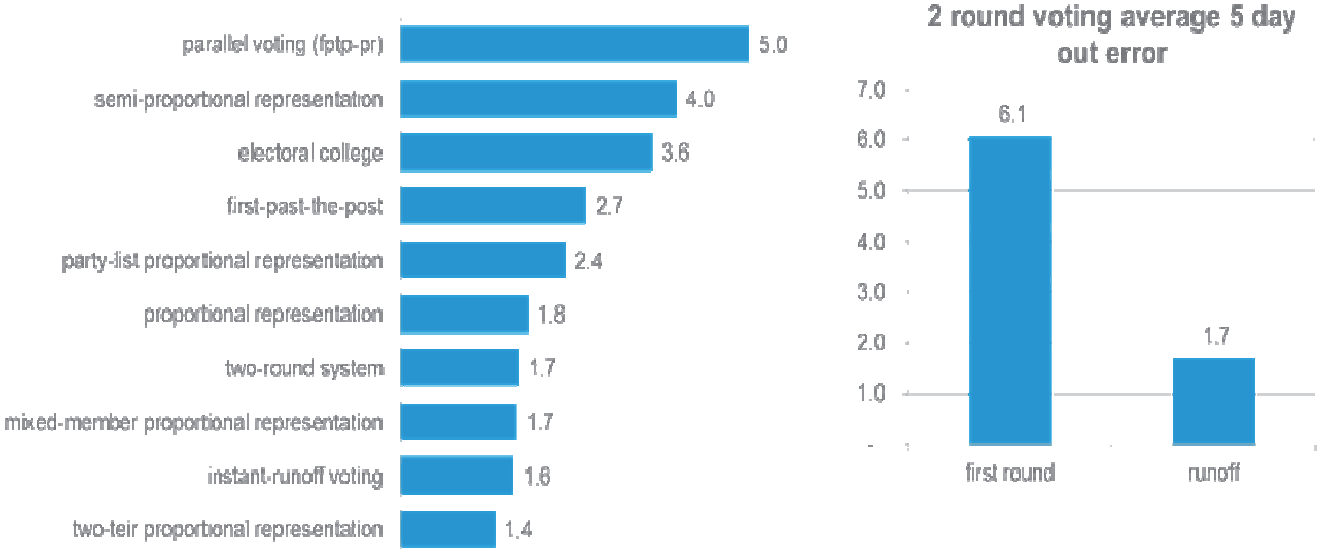


Key things to look out for when cross comparing polls and election events

The accuracy of polls will differ by type of election and by country
The overall accuracy of polls varies considerably depending on the type of election and the country the polls are being conducted in.

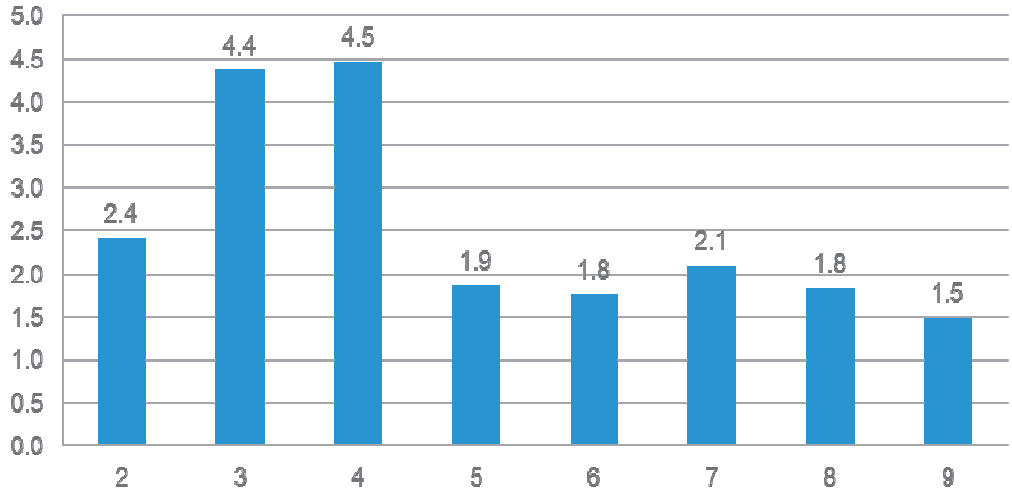
Generally speaking proportional representation elections are associated with lower average polling errors than first past the post elections. Two round systems are ultimately more accurate than single round, though first round votes have much larger errors than run-off votes.

Figure 7. 7 day out polling error by voting method



But there will also be natural differences as a result of the different style of elections and the numbers of candidates or parties contesting each type of vote. As exemplified above, two candidate events will tend to have lower errors than multi-candidate events, and the more parties contesting an election the smaller the proportion of votes for any one party, which will result in lower average errors.

Figure 8. Average error by number of candidates or parties contesting the election



But even if you take account these factors, comparing votes with the same number of parties for example. there can still be significant differences. Polling measures from proportional representation style votes still tend to be more accurate than first past the post votes for example and in some countries polls with the same number of candidates and the same election method are more consistently accurate than in others.

The reasons for these differences are complex. There are a number of factors that will contribute to different scales of error, even before you think about methodological differences used by different survey companies:

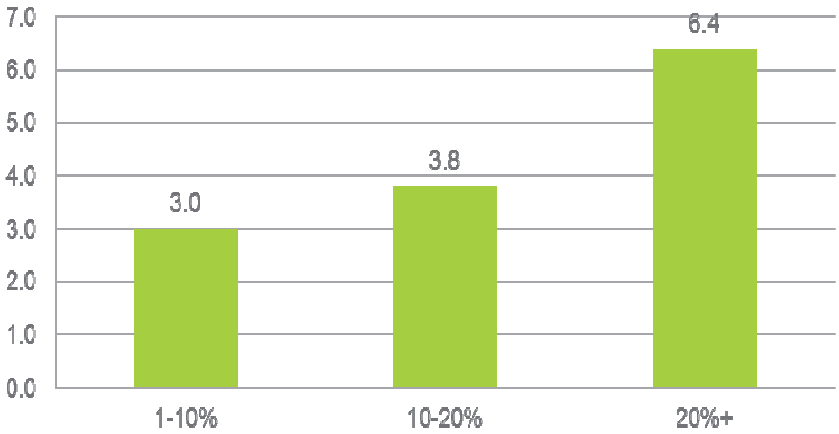
Key factors that impact on the scale of polling error

1. The number of undecided voters in any voting event
2. The number of people who vote in an election and how much this varies
3. The character of the parties, candidates or choices involved in an election
4. How much historical voting data there is available
5. How well mapped out the voter population of the country might be
6. How easy it is to reach a balanced representative sample

Varying numbers of undecided voters

As you can imagine the more undecided voters there are in any one election the harder it will be for a poll to predict the outcome. Analysis of the 438, 7 day out polls in the database where undecided voter data has been published, albeit a small statistical base, reveals a strong relationship between the number of undecided voters and average size of polling errors.

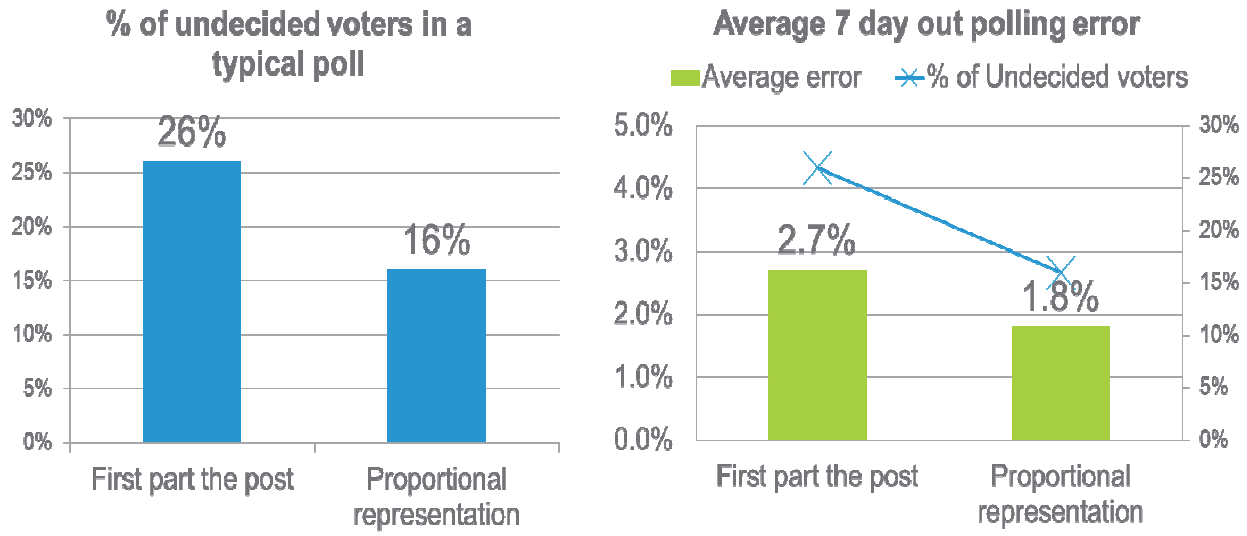
Figure 9. 6 day out accuracy of polls based on the % of undecided voters



One of the main reasons different types of elections have different scales of error is because they tend to have different proportions of undecided voters.

Proportional Representation elections tend to offer a wider outlet for people with differing opinions so there is less of a need to tactically vote, so there tends to be fewer undecided voters compared to first past the post elections. Analysis of our voting database suggests Proportional Representation votes have on average around one third fewer undecided voters which happens to correspond to one third lower levels of polling error.

Figure 10.



Polling errors are also related to the proportion of people who vote and variation in the number of voters in any one election

It is a lot harder to undertake accurate polling in markets with low voter turnout or fluctuating numbers of voters. Polling companies in these circumstances not only have to work out who people will vote for but also predict whether or not they will vote, which is an extra cause of error.

As a result elections in countries or election events with lower average voter shares tend to have larger polling errors.

Figure 11. Polling error vs. turnout

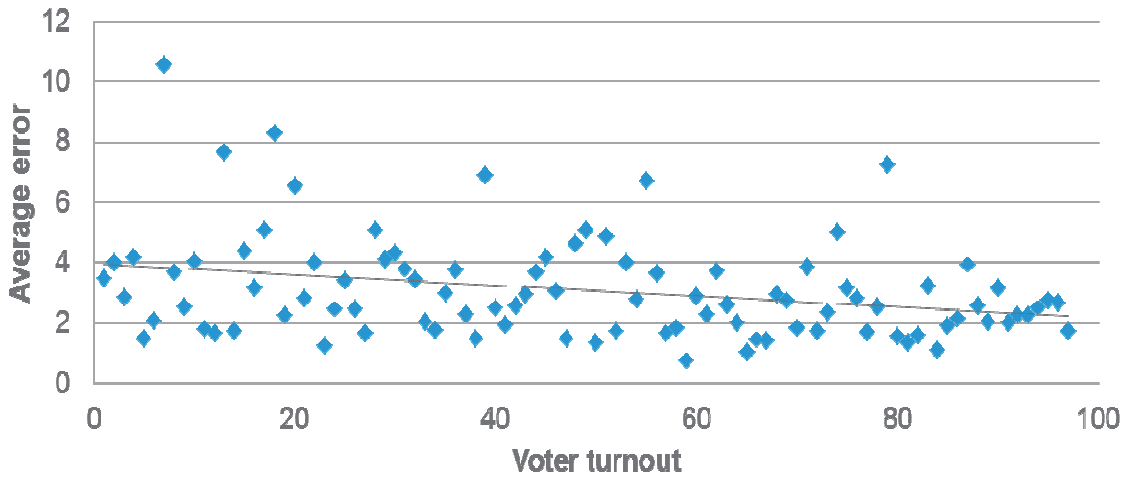
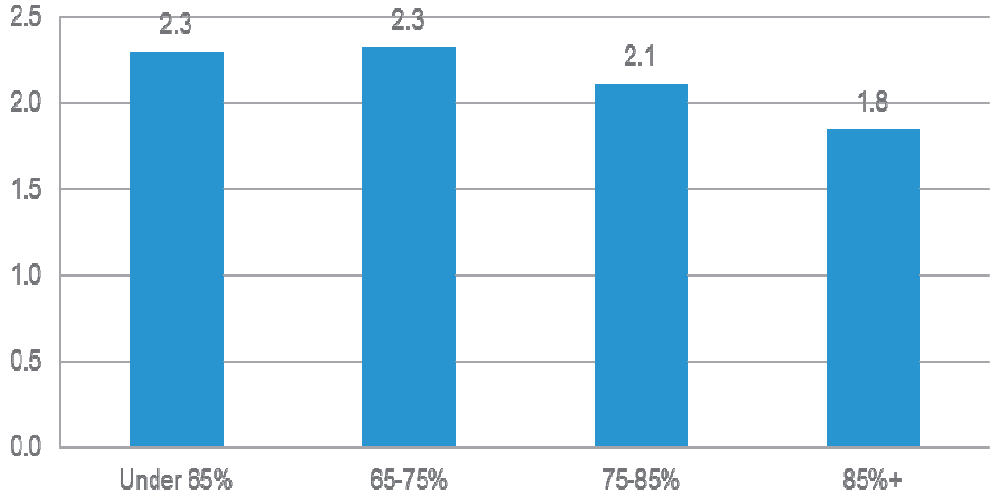


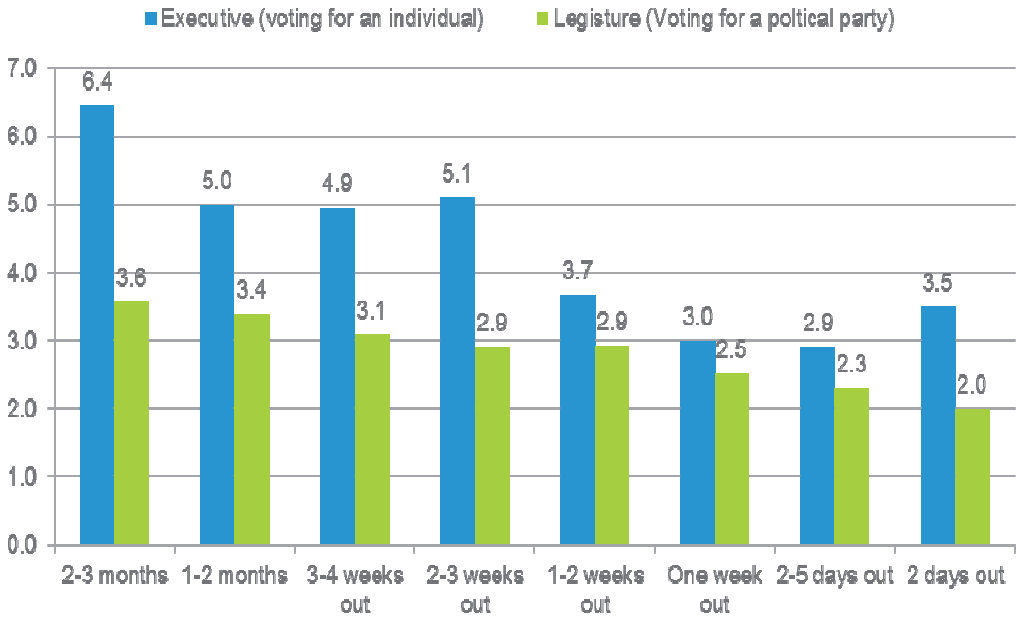
Figure 12. 7 day out accuracy of polls by voter turnout



People vs. Party votes

There are also differences in average polling errors between Legislative votes, where people vote for parties, and executive style votes e.g. presidential elections where people vote for individuals. Legislative polls tend to be more accurate.

Figure 13.



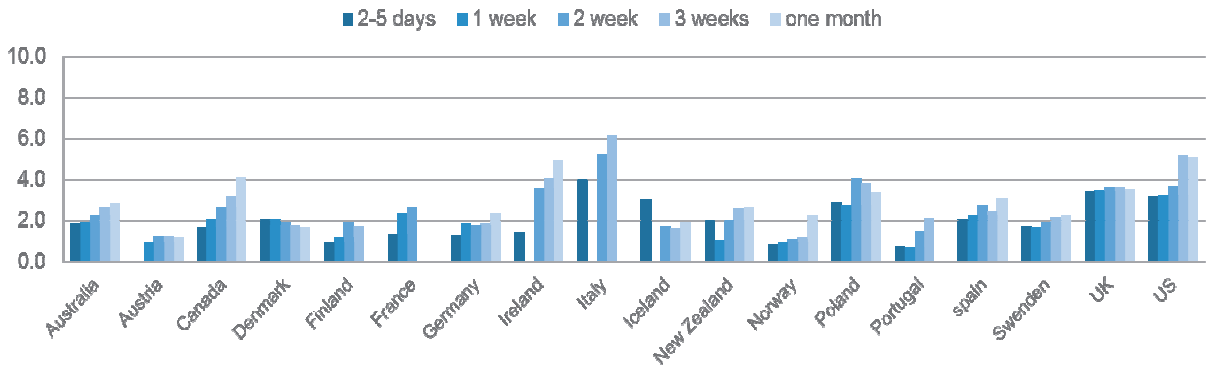
The reason for this is that party allegiances tend to be more stable and predictable than individual allegiances. Executive elections often involve new entrants and one-off candidates with no historical voting history to measure against. You could liken this to the comparative difficulty of predicting the market share for an established brand compared to a new entrant brand in consumer research.

Comparing polling error country to country is difficult

All these factors make it very difficult to compare polling accuracy country to country. As you can see from figure 14, average polling errors can vary dramatically when you examine them at a country level. These differences are not just because of the different style of elections and the number of parties contesting elections in each country, the main reason is simply down to the very small and varying number of voting events in any one country to make a fair

statistical comparison. Figure 14 is shown for anecdotal purpose only, no comparisons or conclusions should be directly drawn from this chart – it is shown simply to illustrate the levels of differences that can be observed.

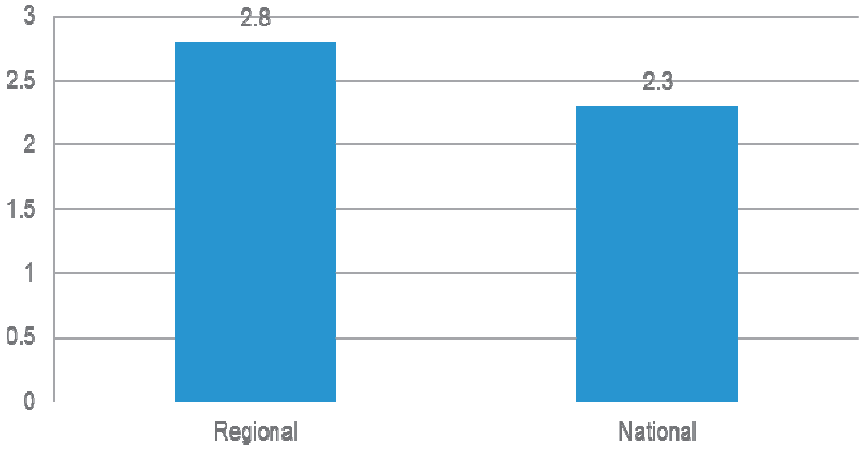
Figure 14. Examples of how polling errors vary by country



Regional v national polls

There is more data to compare regional and national election statistically. Regional polls tend to be associated with 20% higher average polling errors than national based polls.

Figure 15. Regional v national poll accuracy

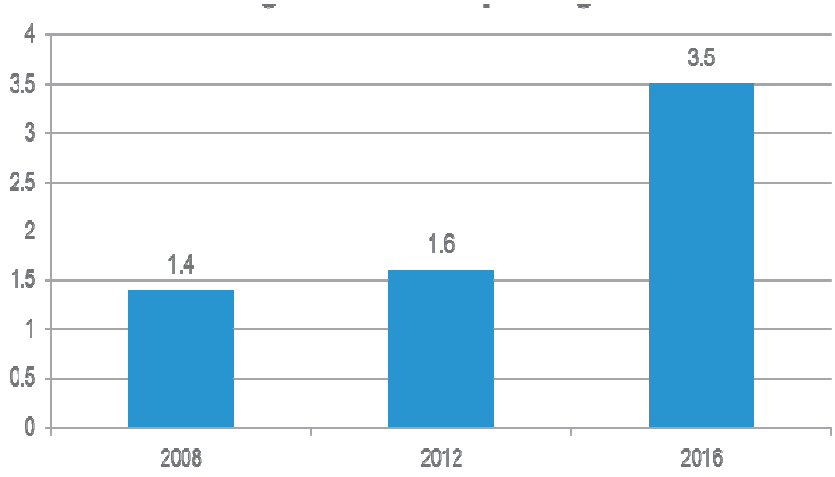


There are several factors at play here. The key general factor that all polling companies have to contend with is that voter population based on regions is often less well mapped out than at a national level. As such this approach is subject to more sample error due to micro regional fluctuations in the demographic make-up of the population.

Samples can also be harder to reach using certain sampling methods at a regional level.

There is a particular emerging difficulty with telephone polling at a state level in the US, for example as a result of the more widespread use of mobile phones. In the past it was possible to geo-target representative geographical regions using localised phone number digits, but with a mobile phone that physically could be anywhere in the US this is not possible. This may have contributed to the disproportionally larger errors seen in the state level polls in the 2016 election that was not apparent at a national poll level.

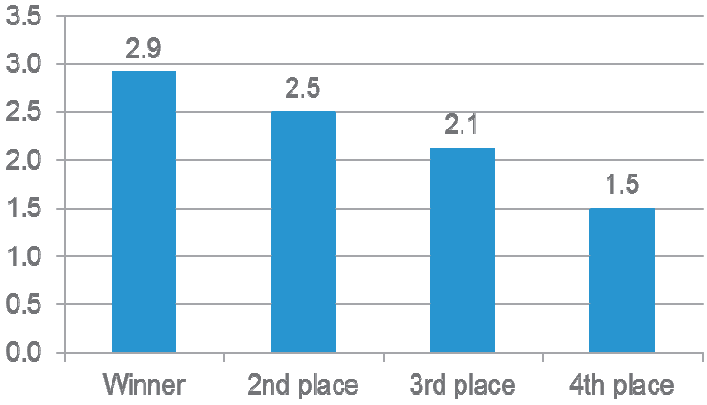
Figure 16. US average state level polling level



Differences in error based on position in the polls

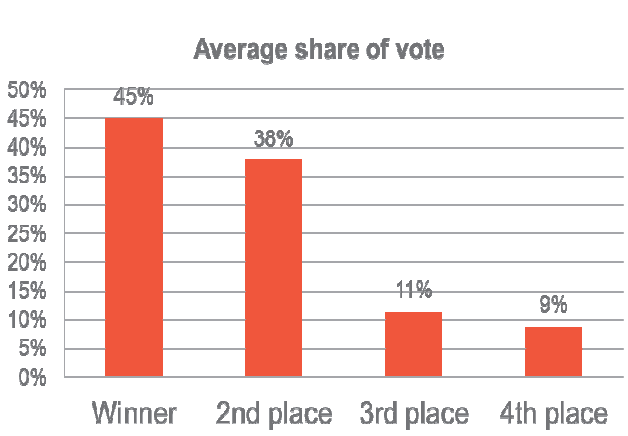
There are differences in the average size of error to expect based on the winning position too. The largest party also tends to have the largest polling error too.

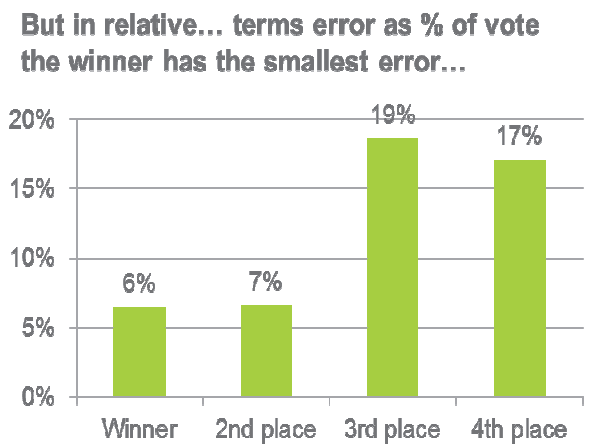
Figure 17. Absolute error by position



Why? Well, the absolute size of error is inherently linked to the total number of people voting for any choice. If you transpose the error to the percentage of the vote you will see a different picture…

Figure 18.



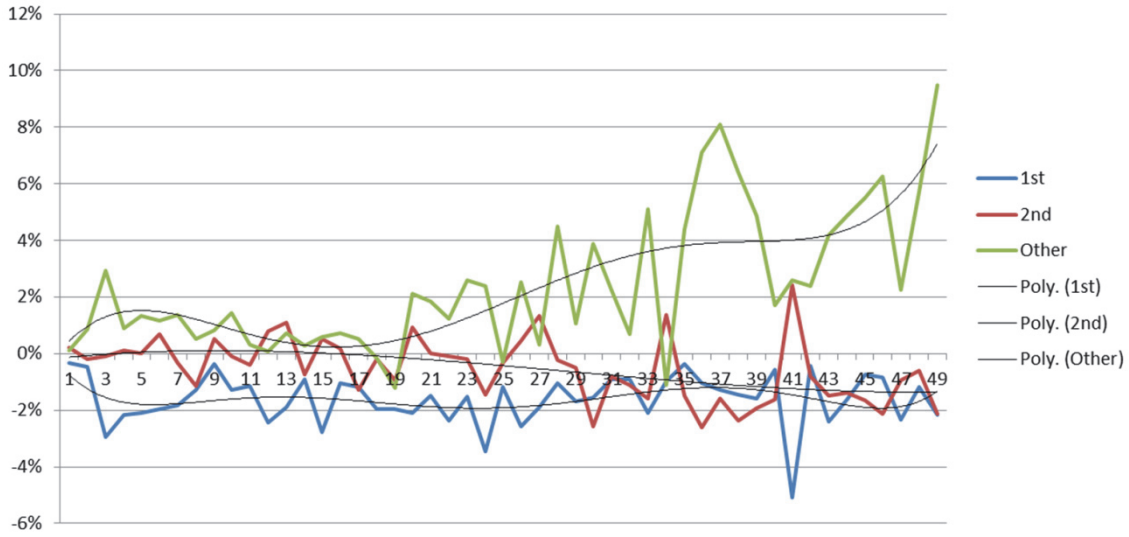


Why? Well, the absolute size of error is inherently linked to the total number of people voting for any choice. If you transpose the error to the percentage of the vote you will see a different picture…

How levels of error vary in the run-up to an election for different parties

It is apparent when examining time sequence comparisons that voting intention poll measures for smaller parties is subject to much higher levels of error in the run-up period to an election than for the main parties.

Figure 19.



You can also observe from this analysis that in the final phase of elections the winning parties vote share tends to be slightly underestimated and the smaller parties slightly overestimated. Another way to look at this, in the last few days of elections there tends to be shifts in voter opinion, moving from the smaller parties to one of the leading parties. This is quite a well known trend, certainly in first-past-the-post British elections where the phenomenon is known as the "third party squeeze". Voters abandoning their latent loyalties to one party or candidate often to vote for a more pragmatic choice of parties to try to ensure that another party does not win the vote.

Win errors appear directionally misbalanced

Large scale analysis of this polling database indicated that the winner error is as a result directionally misbalanced. 62% of the time the winner's margin of error is under-estimated by an average of 2%, mostly made up of vote choice switches from tertiary parties.

Figure 20. 7 day polling error of leading party or candidate

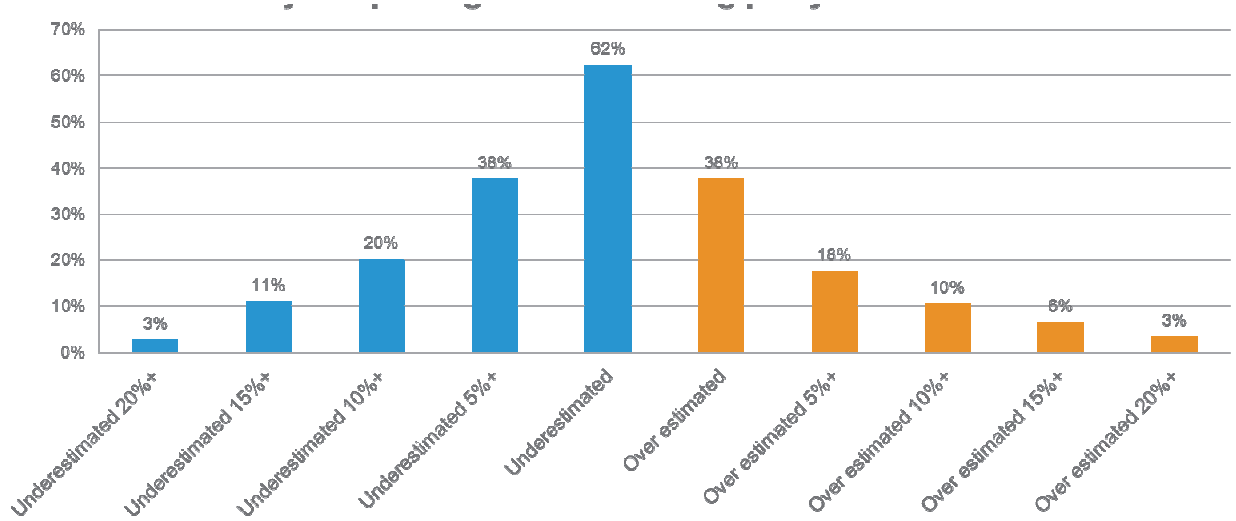


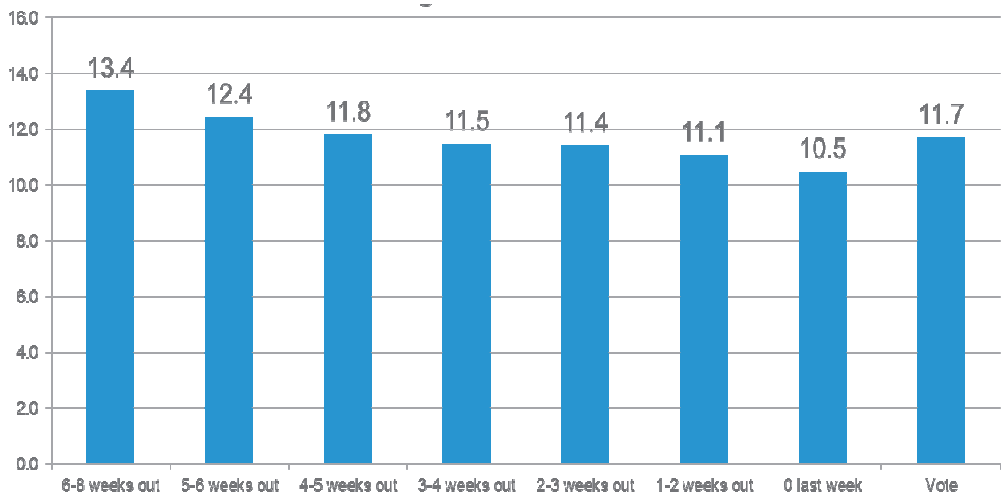
Figure 20 illustrates the percentage of elections with different size of either positive of negative errors.

Lead margin movements

This is a bit of a statistical optical illusion however, as the above analysis is based on looking back from the eventual winner and so is subject to hindsight bias. The winner wins by definition because they have picked up the most amount of votes on the way. But if you look at it from the starting positions of parties you see a slightly different picture.

Generally speaking the party that is in the lead in the polls at the start of a campaign sees their lead eroded over the course of the campaign. This is perhaps because the leading party is subject to more scrutiny. Typically leads erode back by around 3% over the course of a typical eight week campaign but then on election day bounce back slightly by around 1%.

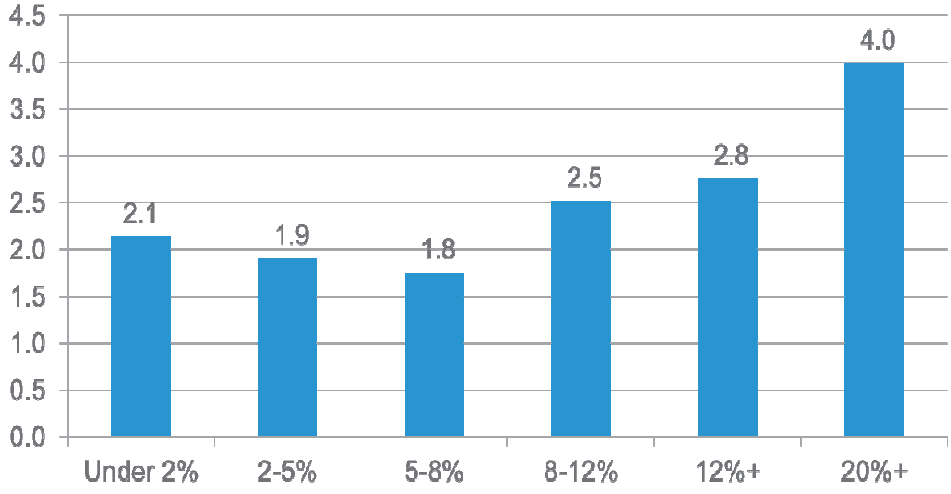
Figure 21. Lead margin movement



Size of polling errors based on margin of victory

There are some interesting differences in the average size of Winner Error based on the how big a lead a party has in the poll. It is a non-linear relationship. In tight elections where the lead margin is under 2%, there are slightly higher levels of election errors that elections than where there is a clearer lead. But as the lead increases to a more unassailable margin there tends to be a much bigger error. Perhaps in these situations voters are actually less concerned about whom to vote for from a tactical perspective and choose to vote more from personal preference.

Figure 22. Error based on margin of victory



Assessing the overall accuracy of polls

The average polling error in the database is 2.5% which is well within the average quoted margin of error of polls in this database (3.4%). Analysing all the polls in this database we found that 88% of polls were within their own declared margin of error and 85% correctly predicted the outcome of the vote they were forecasting.

Figure 23. The accuracy of polls conducted within seven days before an election

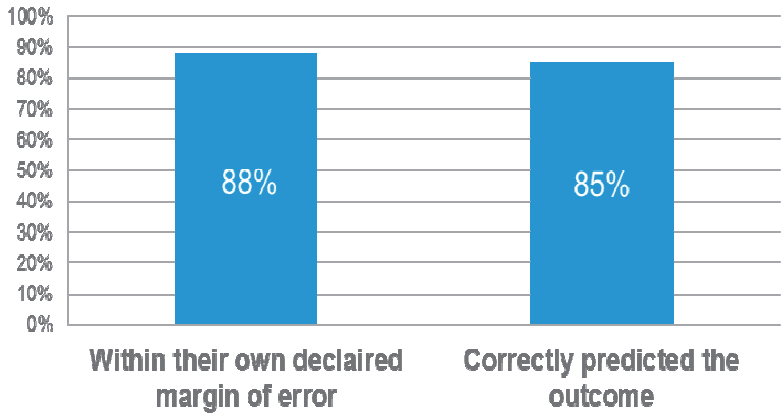


Table 5 illustrates the percentile distribution of error for various different error measurement methods.

Table 5. Percentile distribution of errors by polling measurement technique

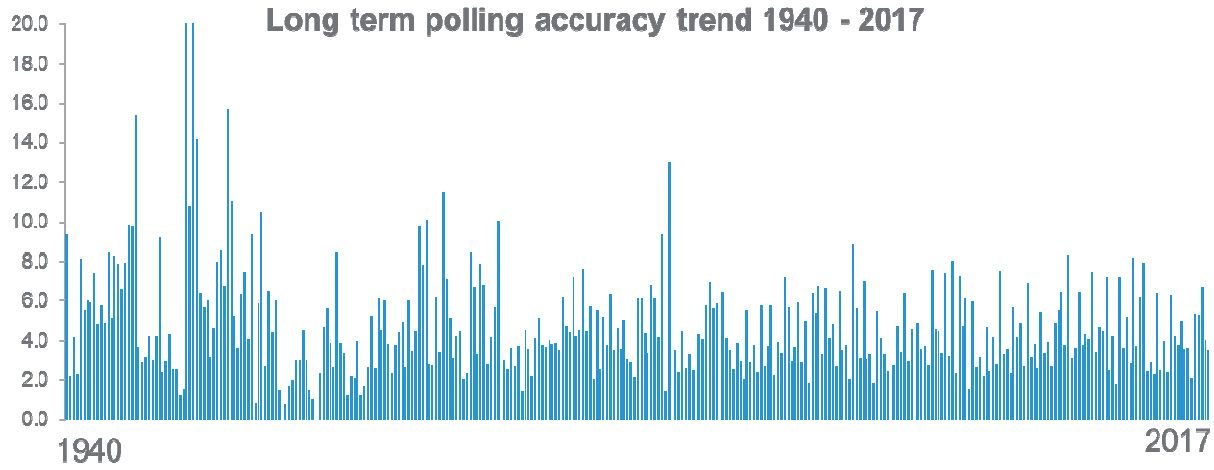
	Percentile	Winner error	Win error margin	top 4 average absolute error	Max error	Mid campaign top 4 error
Smallest error	0-5%	0.1	0.2	0.4	0.6	0.5
	5-10%	0.3	0.6	0.8	1.2	0.9
	10-15%	0.5	1.0	1.0	1.6	1.1
	15-20%	0.8	1.3	1.2	1.9	1.4
	20-25%	1.0	1.7	1.5	2.3	1.6
	25-30%	1.3	2.1	1.7	2.6	1.8
	30-35%	1.5	2.5	1.9	3.0	2.0
	35-40%	1.8	2.9	2.1	3.3	2.3
	40-45%	2.1	3.5	2.3	3.7	2.5
	45-50%	2.4	4.0	2.5	4.0	2.8
	50-55%	2.7	4.6	2.8	4.4	3.1
	55-60%	3.2	5.2	3.1	4.9	3.4
	60-65%	3.6	5.9	3.4	5.3	3.6
	65-70%	4.1	6.7	3.7	5.8	3.9
	70-75%	4.7	7.6	4.0	6.3	4.2
	75-80%	5.3	8.7	4.3	6.9	4.5
	80-85%	6.1	10.0	4.8	7.8	5.0
	85-90%	7.1	11.8	5.5	8.9	5.7
	90-95%	8.6	14.7	6.5	11.0	6.8
Largest error	95-100%	14.5	23.2	9.8	16.4	10.0

Are we getting worse at polling?

This paper has partly come about because of the three notorious polling “miss calls” from the 2015 UK election, the 2016 Brexit Referendum vote and the 2016 US Presidential election. All three prominent votes failed to predict the outcome of the events they were measuring. This has led to some speculation about the potential decline in the accuracy of polling.

It can be concluded that examining polling error data over the history of polling there is no evidence of any long term decline in the accuracy of polls.

Figure 24. Long term polling accuracy trend 1940 - 2017



The paradox of phantom polling errors

It is worth examining these three polling events as error case studies

The UK 2015 General Election

The first election to compare is the 2015 UK general election, generally seen as being a major polling miss call, failing to spot the size of Conservative party's winning margin, out by 4.1%.

….And yes compared to other elections, this election did have a higher than Average Win Error and Win Error Margin, both in the 65th error percentile. But note, the Average Absolute Error was actually below average in this election and the max error was no bigger or smaller than on average. You would in fact expect at least one party's polling measurement in a typical election to be out by around 4%. It just so happened in this case to be the winning party, but in pure statistical terms this was a pretty average election polling result. This election was classified as "miss called" because the polls were suggesting a hung parliament, where in fact the Conservative party won an outright majority which was a surprise but it highlights the fact that size of errors in pure statistical terms are neither here or there if the polls do not predict the outcome.

Table 6. UK 2015 election

	Percentile	Winner error	Win error margin	top 4 average absolute error	Max error	Mid campaign top 4 error
Smallest error	0-5%	0.1	0.2	0.4	0.6	0.5
	5-10%	0.3	0.6	0.8	1.2	0.9
	10-15%	0.5	1.0	1.0	1.6	1.1
	15-20%	0.8	1.3	1.2	1.9	1.4
	20-25%	1.0	1.7	1.5	2.3	1.6
	25-30%	1.3	2.1	1.7	2.6	1.8
	30-35%	1.5	2.5	1.9	3.0	2.0
	35-40%	1.8	2.9	2.1	3.3	2.3
	40-45%	2.1	3.5	2.3	3.7	2.5
	45-50%	2.4	4.0	2.5	4.0	2.8
	50-55%	2.7	4.6	2.8	4.4	3.1
	55-60%	3.2	5.2	3.1	4.9	3.4
	60-65%	3.6	5.9	3.4	5.3	3.6
	65-70%	4.1	6.7	3.7	5.8	3.9
	70-75%	4.7	7.6	4.0	6.3	4.2
	75-80%	5.3	8.7	4.3	6.9	4.5
	80-85%	6.1	10.0	4.8	7.8	5.0
	85-90%	7.1	11.8	5.5	8.9	5.7
	90-95%	8.6	14.7	6.5	11.0	6.8
Largest error	95-100%	14.5	23.2	9.8	16.4	10.0

The Brexit

The next election to compare is the Brexit vote in 2016. This was a "shock" result as the polls did not predict the outcome of the vote, but in pure statistical terms actually the polls were slightly better than average. The four point win margin error might seem large but in a binary choice vote situation this marginal error is in fact completely normal. The only evidence to explain the shock was the larger than normal mid-campaign polling error which set clear expectations of a Remain victory.

Table 7. Brexit

	Percentile	Winner error	Win error margin	top 4 average absolute error	Max error	Mid campaign top 4 error
Smallest error	0-5%	0.1	0.2	0.4	0.6	0.5
	5-10%	0.3	0.6	0.8	1.2	0.9
	10-15%	0.5	1.0	1.0	1.6	1.1
	15-20%	0.8	1.3	1.2	1.9	1.4
	20-25%	1.0	1.7	1.5	2.3	1.6
	25-30%	1.3	2.1	1.7	2.6	1.8
	30-35%	1.5	2.5	1.9	3.0	2.0
	35-40%	1.8	2.9	2.1	3.3	2.3
	40-45%	2.1	3.5	2.3	3.7	2.5
	45-50%	2.4	4.0	2.5	4.0	2.8
	50-55%	2.7	4.6	2.8	4.4	3.1
	55-60%	3.2	5.2	3.1	4.9	3.4
	60-65%	3.6	5.9	3.4	5.3	3.6
	65-70%	4.1	6.7	3.7	5.8	3.9
	70-75%	4.7	7.6	4.0	6.3	4.2
	75-80%	5.3	8.7	4.3	6.9	4.5
	80-85%	6.1	10.0	4.8	7.8	5.0
	85-90%	7.1	11.8	5.5	8.9	5.7
	90-95%	8.6	14.7	6.5	11.0	6.8
Largest error	95-100%	14.5	23.2	9.8	16.4	10.0

2016 US Election

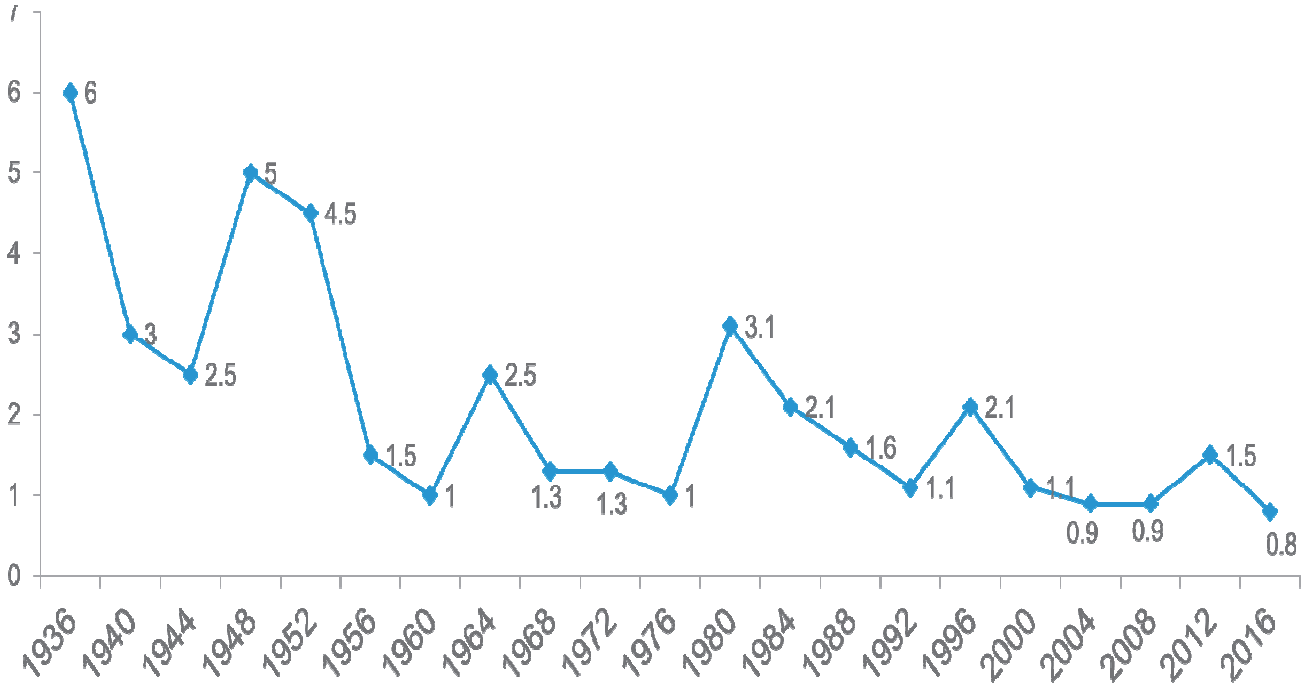
The third event is the 2016 US Presidential election. This again was a shock election result and yes, as highlighted above there were some larger than expected state level polling errors, but at a national share of vote level, the poll predictions were extremely accurate.

Table 8. US 2016 Election

	Percentile	Winner error	Win error margin	top 4 average absolute error	Max error	Mid campaign top 4 error
Smallest error	0-5%	0.1	0.2	0.4	0.6	0.5
	5-10%	0.3	0.6	0.8	1.2	0.9
	10-15%	0.5	1.0	1.0	1.6	1.1
	15-20%	0.8	1.3	1.2	1.9	1.4
	20-25%	1.0	1.7	1.5	2.3	1.6
	25-30%	1.3	2.1	1.7	2.6	1.8
	30-35%	1.5	2.5	1.9	3.0	2.0
	35-40%	1.8	2.9	2.1	3.3	2.3
	40-45%	2.1	3.5	2.3	3.7	2.5
	45-50%	2.4	4.0	2.5	4.0	2.8
	50-55%	2.7	4.6	2.8	4.4	3.1
	55-60%	3.2	5.2	3.1	4.9	3.4
	60-65%	3.6	5.9	3.4	5.3	3.6
	65-70%	4.1	6.7	3.7	5.8	3.9
	70-75%	4.7	7.6	4.0	6.3	4.2
	75-80%	5.3	8.7	4.3	6.9	4.5
	80-85%	6.1	10.0	4.8	7.8	5.0
	85-90%	7.1	11.8	5.5	8.9	5.7
	90-95%	8.6	14.7	6.5	11.0	6.8
Largest error	95-100%	14.5	23.2	9.8	16.4	10.0

In fact 2016 turned out at a national level to be the most accurate Presidential election national voter share forecast in the history of polling.

Figure 25. Winner error US presidential elections 1936 - 2016



Again what may have contributed to shock of the result was the mid-term polling data which was used as the basis for many prediction protocols which was suggested a strong margin of victory for the candidate that eventually lost. The error was also directionally misbalanced, i.e. all the polls were pointing towards a different end result. But as highlighted above this is quite common too, in 62% of polls the error is a directionally misbalanced by an average of 2% so again this direction error is lower than normal. You also have to take account of the peculiarities of the US electoral college system where the national vote winner, Hillary Clinton, wasn't the actual winner.

In conclusion

The polling industry has been subject to some fierce criticism over the last 18 months for failing to predict the outcome of these three recent UK & US elections. If polls do not predict the outcome it is fairly natural to question the polls as a miss called poll occurs only once every six or seven elections. So three miss calls occurring in close succession like this is perhaps no wonder the level of concern, the chances of this happening by chance has been calculated to be less than 1 in 100.

It does appear though from the objective evaluation of the polling errors data from thousands of polls from around the world that in many respects this criticism is undue.

What was just as unusual was how close all three of these elections/votes were, all with victory margins around 4% or less, the chances of three close elections like this happening in a row is equally rare occurring once every 100 elections). As we have seen, the evidence in this paper suggests close elections like these are naturally harder to forecast. They were also conducted in two markets with lower than average voter turnout and higher than average numbers of undecided voters, both factors as explained in this paper primary causes of increased polling error. So when assessed in this wider context of international polling data the scale of these errors do not appear by any means to be that abnormal, overall you even go as far as suggesting that in many respects these results were more accurate that might have be expected under the circumstances.

What this perhaps underlines is how little the size of errors actually matters when the outcome prediction is wrong. It is hoped that this paper will provide a framework to enable the fair evaluation of polling accuracy in the future.

Acknowledgements

With special thanks to David Gundry from York University who helped compile and prepare the Kantar polling database, and preparing the below appendix notes and source references; as well as to Finn Raben, President of ESOMAR, Kathy Frankovic & Enrique Domingo from the ESOMAR Public Standard Committee for their intellectual contributions to this white paper.

References

Arzheimer, K. & Evans, J., 2014. A new multinomial accuracy measure for polling bias. *Political Analysis*, 22(1), pp.31–44.

Martin, E., Traugott, M. & Kennedy, C., 2005. A Review and Proposal for a New Measure of Poll Accuracy. *Public Opinion Quarterly*, 69(3), pp.342–369. Available at: http://poq.oxfordjournals.org/content/69/3/342.short.

Mitofsky, W.J., 1998. Was 1996 a Worse Year for Polls Than 1948? *The Public Opinion Quarterly*, 62(2), pp.230–249.

Mosteller, Frederick. 1949. The Pre-election Polls of 1948: The Report to the Committee on Analysis of Pre-election Polls and Forecasts. Vol. 60. Social Science Research Council

The Author

Jon Puleston, Lightspeed Research, United Kingdom.

Appendix: About the polling database

This analysis is based on data compiled by Kantar on 31,310 polls from 473 elections and voting events across 40 countries around the world from publicly available data sources online, spanning the period of 1936 through to 2017. The database, made up of 160 national voting events & 313 regional, state and mayoral level voting events.

Given that the poll data we currently have is unevenly distributed across multiple markets, it must be noted that any conclusions drawn from the entire data set are subject to availability bias. For historical and international comparisons the summary figures quoted are based on national legislative and executive elections only and exclude non-national based voting events.

Table 1 shows the distribution of polls and votes in the database by country. The most represented countries in terms of poll numbers are the United States (26%), Germany (16%), and the United Kingdom 16%). This is unsurprising due to the popularity of polling in these markets.

The voting events correspond to elections for national and municipal government and referenda. The majority (81%) of voting events recorded are for party-polling for national legislative bodies such as parliaments, followed by national executive bodies (i.e. President).

Different voting systems are represented. The most common is some form of proportional representation (36%), followed by first-past-the-post (22%). (See figure 2.)

Table 1. Counts and percentages of polls and votes represented in the database by country

Country	Number	polls	% age
Australia	11	1,219	4%
Austria	3	174	1%
Belgium	2	18	0%
Brazil	70	370	1%
Canada	4	1,181	4%
Columbia	2	47	0%
Croatia	3	102	0%
Czech	6	387	1%
Denmark	2	497	2%
Egypt	1	33	0%
Estonia	1	54	0%
Finland	3	105	0%
France	32	1,041	3%
Germany	15	4,924	16%
Greece	2	114	0%
Iceland	3	162	1%
Iran	1	12	0%
Ireland	1	201	1%
Israel	2	223	1%
Italy	4	1,330	4%
Japan	2	39	0%
Kenya	1	17	0%
Luxembourg	1	1	0%
Mexico	3	97	0%
Netherlands	2	25	0%
New Zealand	4	486	2%
Norway	2	91	0%
Peru	1	54	0%
Philippines	3	98	0%
Poland	7	1,041	3%
Portugal	3	290	1%
Romania	2	48	0%
Russia	1	140	0%
Slovakia	2	25	0%
South Korea	3	106	0%
Spain	13	1,899	6%
Sweden	5	1,126	4%
Taiwan	2	380	1%
United Kingdom	43	5,003	16%
United States	205	8,150	26%
Total	**473**	**31,310**	**100%**

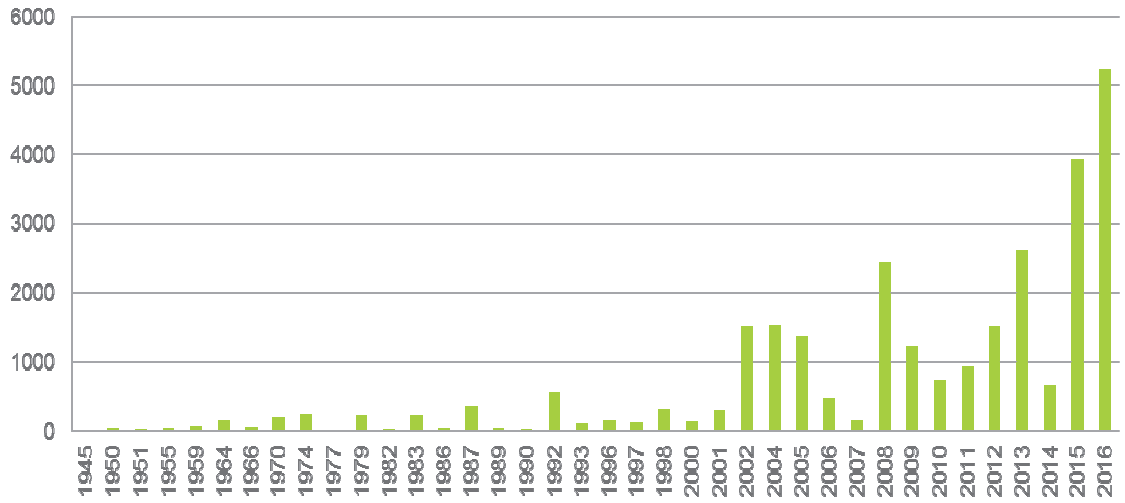
Table 2. Counts of countries, votes and polls in the database.

	Polls	%	Votes	%
Total	31,310	100	473	100
CONSTITUTIONAL BODY				
Executive	9549	30%	226	54
Gubernatorial	149	0%	41	11
Legislature	19429	62%	111	25
Referendum	453	1%	9	2
Senatorial	115	0%	27	7
VOTING SYSTEM				
Electoral College	8096	26%	192	51
First-Past-The-Post	5574	18%	48	13
Instant-Runoff Voting	1219	4%	11	3
Mixed-Member Proportional Representation	5410	17%	19	5
Parallel Voting (Fptp-Pr)	83	0%	4	1
Party-List Proportional Representation	421	1%	8	2
Proportional Representation	4699	15%	26	7
Semi-Proportional Representation	857	3%	1	0
Two-Round System	542	2%	47	13
Two-Tier Proportional Representation	154	0%	2	1

The table is broken down into some of the categories included in the database. Not all data is fully categorised so percentages do not necessarily add up to 100%.

More recent votes are much more highly represented in the database in terms of polling. 89% of polling data in the database is post-2000, and 54% is post-2010. In part this is due to availability bias in the construction of the database. More recent polling data is easier to locate and in a more accessible format, hence the dominance of markets like the United States, where polling data is readily available. However, it also reflects the fact that opinion polling has become much more common in recent years.

Figure 1. Polls contained in the database grouped by year



Poll accuracy and bias

An archive of polling data can be used for understanding sources of bias and error in opinion polling. For example, Crespi (1988) found that the closer a final opinion poll was to the election, the more accurate it was. In a surprising result, Crespi (1988) and subsequently Lau (1994) do not find a connection between sample size and poll accuracy, while DeSart and Holbrook (2003) do. Many more effects on accuracy have been discovered, such as the length of a poll's fieldwork (Lau, 1994), whether the poll is conducted during the week or at weekends (Salmond, 2009), or whether likely voters or registered voters are sampled (DeSart and Holbrook, 2003).

Polling accuracy, or inversely polling error, is a term used to describe the predictive accuracy of an opinion poll relative to a vote. Poll bias is a related concept, and describes the extent to which the distribution of votes as predicted by a poll deviates from the distribution of votes in the election. Because it is concerned with the distribution of votes, poll bias is not affected by whether a poll reports undecided voters. Measures of polling error, however, require the allocation or exclusion of undecided voters.

In a situation where more than two parties are polled, a further distinction can be drawn between measures of accuracy single party, and measures of accuracy for the whole poll. The former are directed concepts (a party can be biased towards or against, a prediction error can be up or down), whereas the latter are not directed concepts (a poll can be unbiased, but it cannot have negative bias). Directed measures of poll bias actually describe bias for a party, which is a suitable surrogate only in two-party polls.

Measures of accuracy for a poll are useful when analysing effects over the entire poll, for example, does a particular methodology lead to increased error? Measures of accuracy for a particular party are useful for analysing effects within polls such as house bias. Obviously, both bias and error towards or against a particular party will be largely determined by historical factors in addition to any methodological factors.

Ways of measuring party poll accuracy

An obvious measure determining the error in a poll with respect to a single party is the difference between a vote outcome and the corresponding poll prediction for that party. This is intuitively straightforward, as it gives the number of percentage points that the poll was out; the raw directed error. This value will be positive where the polling is over-optimistic about a party's success and negative where the party exceeds the poll's predictions. However, by this measure, polls that report undecided voters will generally underestimate party support as these undecided voters will be distributed among the parties in the final vote. Methods to mitigate this include distributing undecided voters among the parties and recalculating the percentages to exclude undecided voters.

Arzheimer and Evan (2014) developed a measure of per-party poll bias, *A'*, for multi-party polls related to an earlier, two-party measure developed by Martin et al. (2005), *A*. As Martin et al's *A* is for two parties, it can serve (in two-party polls) also as a per-party measure of accuracy.

Martin et al. (2005) propose an odds-based measure of polling accuracy, A, comparing the relative poll bias between two parties. They define accuracy as the natural log of the ratio between two ratios. The first ratio is between the poll values (*P*) for the first two parties. The second ratio is between the vote results (*V*) for the first two parties.

$$A = \ln\left(\frac{\frac{P_1}{P_2}}{\frac{V_1}{V_2}}\right)$$

The measure *A*, describes the bias (and error) in terms of an odds ratio for the two winning parties. An odds ratio can be interpreted as in betting odds. A poll with an odds ratio of 1.1 in favour of party gave that party has a higher chance of winning than was justified by the final result. A poll with an odds ratio of 0.9 for a party was biased against them.

Arzheimer and Evans extend *A*, proposing A'_p, which is a measure of party polling bias for each party *p* in a poll. This is still expressed as an odds ratio, however A'_p describes the bias for that party relative to all of the other parties in the poll. With two parties, this corresponds exactly to Martin et al's *A.*

$$A'_p = \ln\left(\frac{\frac{P_p}{\sum_{i=0}^{n} P_i}}{\frac{V_p}{\sum_{i=0}^{n} V_i}}\right) for\ i \neq p$$

Like *A*, *A'* is the natural log of a ratio of two odds ratios. The first odds ratio is the poll share for a given party against the summed poll shares for the other parties. The second odds ratio is the vote share for that same party against the sum vote shares for the other parties. Measures of A' can be calculated relative to a different number of parties and this will give different results.

Whole poll accuracy

The accuracy of a poll considered as a whole is a single value that represents the overall accuracy of a poll. However, even in cases with only two parties, there is no obvious formula to produce such a value.

Mosteller et al. (1949) present eight methods for calculating the overall accuracy of a poll. Some of these are evaluated in detail in Mitofsky (1998). Of these, Mitofsky prefers measures 3 and 5. He ultimately selected 5 as the only measure that it comparable between both two- and multi-party votes. Unfortunately this is because it only includes the first two parties.

Mosteller et al. (1949) define M_3 as "The average (without regard to sign) of the percentage point deviation for each candidate between his/her estimate and the actual vote." They define M_5 as "The difference between two differences, where the first difference is the estimate of the vote for the two leading candidate from a poll and the second difference is the election result for the same two candidates".

$$M_3 = \frac{\sum_{i=0}^{n} |P_i - V_i|)}{n}$$

$$M_5 = \frac{V_1 - V_2}{P_1 - P_2}$$

M_3 is easy to interpret as it is expressed in percentage points. For each of the significant parties, M_3 gives the average to which the poll miscalled their vote share. M_5 represents the amount to which the lead of the leading party was overestimated compared to the second party. A high value of M_5 indicates the poll was biased against the leading party, and a fractional value of M_5 indicates the poll was biased towards the leading party. A value for M_5 of 1 indicates the poll was accurate with respect to the two leading parties.

Measure 5 (M_5) is comparable across elections, but unsatisfactory for elections with more than two parties. Measure 3 (M_3) incorporates all 'significant' parties, leaving the decision of how many parties to include up to the analyst. Adding additional parties artificially reduces the error, making it not reliable for comparison between two-party and multi-party polls. However, for our summary analysis, will include up to the top four parties in the calculation of M_3. We will express this measure as $M_3(4)$. While not methodologically sound, we feel this is a suitable compromise between two-party and multi-party polls for the purposes of our discussion, as it permits the use of this more intuitive measure.

Arzheimer and Evans' (2014) *B* is a multi-party development of Marin et al's (2005) *A*. Arzheimer and Evans' A' was described in an earlier section as the odds ratio describing the relative bias towards or against one party relative to the other parties. *B* is calculated as the unweighted average of the absolute values of A'_p, for all parties *p*.

$$B = \frac{\sum_{p=0}^{n} A'_p}{n}$$

However, as they admit, this measure of bias is inflated by the inclusion of additional parties. As the average of party bias A'_p is unweighted, biases in the measurement of small party votes have the same effect as biases in the measurement of large parties' votes, even though the biases in the large party votes will likely have a much more significant impact on poll interpretation. This is a problem for our analysis, as polls vary between reporting only a few, and very many party results. Unfortunately, there is no overall measure of whole-poll accuracy as yet that is reliable for comparing polls with different numbers of parties.