Predicting elections

A 'Wisdom of Crowds' approach

Martin Boon
ICM Research

Opinion polls are the currency of politics. They are used by media organisations to evaluate the performance of governments, and by governments and political parties to test the policies that shape manifestos and reform agendas. But opinion polls all rely on one thing – asking people how they themselves intend to vote – and, too often, classical opinion research techniques fail to confront the issues that underpin inaccuracy. In the UK and in many other countries around the world, their performance over the past 20 years has ranged from excellent to disastrous.[1] The 'Wisdom of Crowds' concept turns conventional predictions on their head. It assumes that any crowd that conforms to a core set of principles is capable of delivering a more accurate prediction than the smartest people within it. This paper tests this proposition within the context of actual elections in the UK, showing that the Wisdom of Crowds approach used by ICM Research at the 2010 general election would have produced the most accurate final pre-election prediction. It also shows that a Wisdom approach to regular vote intention tracking produces an interesting complement to classically conducted vote intention polls. Or, if one were to be bold, a competitor to them.

Introduction

The 2010 general election featured as much diversity in political polling data collection methods as the collective insight of the research industry could possibly muster. All three quantitative interviewing methods were utilised, with online (51 polls from the day the election was called to election day itself) and telephone (33 polls) featuring extensively, and even the, nowadays rarely employed, face-to-face interviewing technique enjoying a handful of outings (four polls). But that diversity still depended

[1] In the 1992 UK general election, for example, most of the published, final polls predicted the wrong winner.

Received (in revised form): 20 March 2012

DOI: 10.2501/IJMR-54-4-465-483

on one obvious concept: asking people directly how they themselves planned to vote.

The iconic vote intention question may have subtle wording variations over time, or between different polling organisations, but there can be little doubt that all polls intend to measure individual and actual voting behaviour either in a hypothetical 'election tomorrow', or in the one that might actually be taking place on the day of the poll's publication. Using a combination of classic sampling methods, tried and tested (but varied, and often controversial) analytical techniques, and perhaps relying on a little bit of old-fashioned luck, pollsters add together the voting intentions of 1,000 or more people each time, and come up with four political party share percentages that, in the case of a 'final prediction' poll at least, are supposed to give an accurate prediction of the actual outcome that is announced little more than 24 hours later.

However, political opinion polling is a rather tricky business, at least if the example of recent British general elections is anything to go by. In 2010, the most accurate poll had an average error of 1.25%,[2] but the worst polling average error was 3.25%, a figure outside standard statistical tolerances.[3] While 2005 was considered to be a success for the pollsters, the 2001 and 1997 elections featured mediocre final poll assessments, and the very mention of the 1992 election brings shudders to the spine of the pollsters who remember it, most of whom predicted the wrong winner.

The causes of polling error are discussed elsewhere in considerable depth. For our purposes, all we need to highlight is the simple observation that, with honourable exceptions, the recent record of classically conducted vote intention polls in the UK is variable and mixed. Maybe the voting public has become too sophisticated – or too reluctant – to provide reliable individual answers to a voting question that requires a deeply personal response. As Earls and Kearon (2009) attested, people are generally unreliable witnesses about themselves; maybe then, within the polling context, there is an alternative, or even better, way to assess voting behaviours than to ask people about their own.

The Wisdom of Crowds

The Wisdom of Crowds is a concept now well known to market researchers, the idea being that the collective estimation of a random

[2] 'Average error' is defined as the average of the differences between the estimated percentage for each party and the actual result.

[3] British Polling Council. http://www.britishpollingcouncil.org/press100508.html.

crowd is superior to even the smartest people within it, so long as a small number of conditions are met. Surowiecki (2004) defined the conditions as follows.

1. Diversity of opinion: each person has some information, even if it is an eccentric interpretation of known facts.
2. Independence: opinions are not influenced by anyone else contributing to the crowd's response (which cannot be the case within an opinion polling setting).
3. De-centralisation: people are able to specialise and draw on local knowledge.
4. Aggregation: some method exists for turning private judgements into a collective decision; for us, any opinion polling methodology is appropriate.

When such conditions are met, the errors within sets of predictions cancel one another out to produce an accurate estimate. Surowiecki (2004) evidenced his concept with numerous examples, notably a crowd perfectly predicting the weight of an ox and, more pertinently, a few hundred amateur financial traders in Iowa outperforming Gallup polls in predicting American election results. This led ICM to ask itself a simple question: would our own randomly selected sample (a crowd, couched in an orthodox quasi-random telephone survey guise) do better at predicting the election result via an averaging out of their own guesses, than they would by feeding their own voting intentions through the ICM polling methodology?

Market research is rarely measured against a reality outcome. Consumer and business-to-business research assumes accuracy within standardised margins of error so long as standard tenets of probability theory are complied with; there is rarely an election 'tomorrow' to prove the efficacy of monthly, weekly or even daily output of vote intention polls. Only when an election does occur tomorrow does a political opinion poll morph into a highly anticipated prediction poll: a test of credentials of both polling and pollster. The 2010 UK general election, then, was a once-in-a-political-cycle opportunity to test the merit of ICM's (and others') polling skills. It also provided the perfect opportunity to test whether the prediction of the crowd was better than the prediction of the poll.

The 2010 election test

Our final poll for the *Guardian* newspaper interviewed a random sample of 2,022 adults aged 18+ by telephone on 3–4 May, for publication on

election day (6 May 2010). The poll was conducted using exactly the same questions that each and every ICM political poll has employed for many years: a turnout estimate, vote intention question, and recall of 2005 voting behaviour questions.

The poll also used the standard suite of ICM political polling techniques, as follows.

- Data were weighted to the profile of all adults aged 18+ (including non-telephone-owning households). Data were weighted by sex, age, social class, household tenure, work status and region. Targets for the weighted data were derived from the National Readership Survey, a random probability survey comprising 36,000 random face-to-face interviews conducted annually.

- Data were weighted to turnout to ensure that people's certainty to vote was properly reflected within the data. On a scale of likely turnout, where 10 implies that someone was 'certain' to vote, and 1 implies they were 'certain not to vote', the score given directly reflected that potential voter's importance within the data. Where someone stated they were 10/10 certain to vote, they were allocated a weight of 1. When only 9/10 certain, they were allocated a weight of 0.9, and so on. Where someone provided a likely to vote score, but did not vote in the previous general election in 2005, their turnout weight was halved. For example, if someone said they were 7/10 certain to vote, they were allocated a weight of 0.35.

- Data were past vote weighted (targets set 80% to the 2005 general election result and 20% to the average recall of the previous 25 ICM polls to account for faulty recall among a small number of voters). Past vote weighting is an analysis technique that seeks to ensure that a sample is not only demographically representative, but politically representative. The simplest way of doing this is to tie the recall of voting in the previous general election back to the actual result, give or take an allowance for misremembering.

- Finally, data were manually adjusted by reallocating half of the people who told us how they voted in 2005, but refused or didn't know how they would vote in 2010 (partial refusers), back to the party they voted for in 2005.

The poll produced the most accurate pre-election prediction of the actual general election result, producing an average error of 1.25%.[4] This published average error figure is the benchmark by which the Wisdom of Crowds prediction is comparatively judged. For the Wisdom of Crowd concept to have predictive credibility, its average error needed to compare favourably against the best poll prediction, which happened to be the very poll on which the Wisdom suite of questions were also included.

The Wisdom of Crowds questions were placed at the end of the telephone interview, i.e. after the completion of the standard vote intention suite and a small number of other political questions. This was deliberate; the primary function of the poll was to provide the *Guardian* with its poll prediction, and as such respondents had to approach the standard suite of polling questions without influence. While it could be argued that respondents had been influenced by the standard polling questions before they were asked the Wisdom questions, it is also true that the object of the test was to see if the same set of respondents could improve upon their own 'poll' prediction via the Wisdom concept.

At the end of the interview, ICM asked two Wisdom questions: first, for an unprompted guess at the actual outcome and, second, for a prompted guess once the party shares in 2005 had been revealed to each respondent. There can be little doubt that the core of the Wisdom of Crowds premise is contained within the spontaneous question. There was no information given to respondents, little in the way of preamble, and respondents were simply asked to supply their estimated share for each of the three main parties and the net share for minor parties. The second question is less pure, giving away vital information about recent, relevant electoral history. In many ways, this prompting was the equivalent of orthodox polls' past vote weighting mechanisms, mooring the poll to contemporary political reality, but at the interview stage rather than via data-processing applications.[5]

The exact wording of the two questions was as follows. For each question, respondents' answers were written in by the telephone interviewer, and the Computer Assisted Telephone Interview system totalled the four sets of responses. When percentages totalled more or less than 100%, respondents were advised to revise the share of the votes accordingly.

[4] http://www.britishpollingcouncil.org/press100508.html.

[5] On this occasion, and all subsequent occasions (excluding referendum tests) when the Wisdom questions were asked, both questions were put to respondents, spontaneous question first, prompted question second.

Q1: As a bit of fun, please tell me what percentage share of the vote you think the …… party will win in the forthcoming general election?

The Labour Party
The Conservative Party
The Liberal Democrats
Other parties

Q2: At the last general election in 2005, the Labour Party won 36% share of the vote, the Conservatives had 33% share, the Liberal Democrats 23% share and smaller parties combined had 8% share of the vote. Knowing this, please tell me what percentage share of the vote you think the …… party will win in the forthcoming general election?

The Labour Party
The Conservative Party
The Liberal Democrats
Other parties

Question wording is an imprecise science, and a number of alternative wordings were considered. The introduction to the initial question was couched as 'a bit of fun' – unorthodox, but with perceived merit for two reasons. First, it should be remembered that respondents had already been asked how they themselves planned to vote, and some mental separation was needed between the two sets of questions, which, at first glance, might seem very similar. Second, survey respondents do not react well to the thought of giving up their volunteered time to be unexpectedly tested. There was a conscious need to avoid intimidating randomly selected people, many of whom will not have thought much about politics, despite the imminent general election. Presenting the questions as 'fun', seemed the best way of avoiding such pitfalls.

The value of the Wisdom method, however, hinged on the accuracy of the predictions it produced. Table 1 shows the actual election result in 2010, measured against the ICM prediction in the *Guardian* newspaper and the two Wisdom question predictions. It should be remembered that all the individual share estimates were aggregated, and averaged out at total level. This was the only analysis technique that needed to be applied in production of vote share predictions. Each of the three predictions was measured in terms of average error, and based on data weighted only to

Table 1 2010 general election result and predictions

	Election result	ICM final prediction	Wisdom spontaneous prediction	Wisdom prompted prediction
Labour (%)	29.6	28	34	31
Conservative (%)	36.8	36	36	35
Liberal Democrat (%)	23.8	26	21	24
Others (%)	9.8	10	9	10
Average error	**–**	**1.2**	**2.2**	**0.9**

standard demographics. A comparison with past vote weighted data is discussed in a subsequent section.

There are a number of immediately striking observations about the outcomes. First, the ICM poll prediction – itself being the closest pre-election prediction – was outperformed by the prompted Wisdom outcome, with the average error figure down to 0.9%. However, the purist version of the Wisdom question fared noticeably worse, producing an average error of 2.2%. That said, the average error across all final polls conducted by British Polling Council members averaged 2.2%, which makes even the purist Wisdom question no better or worse than the polls as a whole.[6]

Second, the Wisdom predictions replicated the most serious polling error of the past 20 years: overstating the eventual Labour share of the vote (while most of the final prediction polls did not). Moreover, both Wisdom questions understated the Conservatives (although marginally), another phenomenon with which British polling observers are well acquainted.

Third, and perhaps most dramatically, the Liberal Democrat share was precisely accurate for prompted prediction, but three points *lower* on the purist, spontaneous prediction. The great polling mystery of the 2010 election was, of course, why all the pre-election polls overstated the Liberal Democrat surge after Nick Clegg's bravura leader debate performances. With these sets of predictions, however, the error is on the reverse side of the coin. Part of the reason why the ICM poll overstated the Liberal Democrats was because of the impact of past vote weighting, which pushed up the Liberal Democrat score by two points (Curtice & Sparrow 2010), but the more relevant question is why the Wisdom questions differed from all classically conducted poll outcomes by erring on the underside of the Liberal Democrat share spontaneously, or getting

[6] The published ICM poll average error of 1.2% would have been improved to 1.0% had we done nothing more than demographically weight the sample, rather than employing all ICM standard techniques.

it spot on under prompted conditions. If the Wisdom theory holds, it may well have been because the crowd saw through the hyperbolic reaction to Clegg's debate performances and didn't believe that fellow electorate members would abandon their tribal political instincts. In other words, the crowd was smart.

Perhaps the great attribute of the Wisdom methodology is its simplicity. The method liberates pollsters from all the sampling constraints and theoretical difficulties that have dramatically affected polls over the past 20 years. It circumvents all the issues that surround conventional polling because it asks people about what 'others' will do, rather than addressing the behaviour of the respondent. It should be noted that no other questioning, modelling or analysis was required for the provision of a predictive set of numbers from these questions. Unlike classic voting intention polls that correct for things like turnout issues, recall of past voting behaviour error, and adjustment of 'refusers' and 'don't knows', the Wisdom questions need no further work. All that was required was the averaging out of all guesses.

Furthermore, unlike our final poll, the views – guesses – of everybody counted. ICM's final telephone poll of 2,022 adults actually counted the answers of only 1,527 respondents once non-voters and 'total refusers' had been factored out – a loss of one-quarter of the contacted sample. This is an important problem for conventional polls: not only is a significant chunk of a poll sample excluded from the process, but if the non-response is differential by party, then the vote intention of the excluded might have real (but unknown and, as such, unmeasurable) implications for the vote share prediction.

For the Wisdom predictions, however, all 2,022 people counted irrespective of whether they voted or not. Indeed, the concept actively values the inclusion of the less politically engaged and less informed: it is a theoretically necessary condition of accuracy (although the exclusion of 'don't knows' and refusers from the Wisdom prediction set (2010 general election data) had no substantive impact on the predictions).[7]

Vote intention tracking tests

With the prospect of the Wisdom concept being a credible source of polling information, logically it followed to test vote intentions at the same point

[7] Average error improved slightly to 2.0% on the spontaneous question, but worsened fractionally to 1.0% on the prompted question.

in time that ICM produced regular vote intention tracking polls on behalf of the *Guardian* newspaper. Every month since 1989, ICM has produced vote intention numbers for the newspaper, most usually around the third weekend of every month.

However, the central point to the Wisdom concept is that any diverse, independent and decentralised crowd can produce accurate predictions so long as some form of aggregation exists. ICM is able to aggregate data via telephone or online data collection methods, but hitherto has used only telephone data collection methods because it has serious issues with the efficacy of online polls (Sparrow 2007). When it comes to the Wisdom context, however, these concerns can be put to one side, being largely irrelevant. Even though random online samples attempt to be fully representative by demographics via the application of standard demographic weighting procedures, it should not matter to the Wisdom concept if the crowd is demographically, or perhaps even politically, compromised or biased.

For this reason we chose our online omnibus service as the vehicle to track Wisdom predictions, which we now call the Wisdom Index. The online omnibus is conducted twice weekly and comprises a minimum of 2,000 adults aged 18+. On the four separate occasions that we placed the questions on it, fieldwork closely approximated the independent telephone tracker poll for the *Guardian*. The objective of the repeat tests was simple: to understand whether or not the Wisdom Index produced credible voting intention numbers on an ongoing basis.

However, there is a complication surrounding the definition of 'credible'. ICM polls are seen to be credible because they have consistently been more accurate than most in election prediction scenarios. However, ICM's monthly 2011 polls have, for example, higher Liberal Democrat shares than any other polling firm. There is no way of proving which set of vote intentions is right and which is wrong; whether something is credible or not largely depends on the views and prejudices of the poll user. However, for the purposes of our test we assumed that ICM's regular monthly telephone polls were the standard by which the Wisdom questions could be objectively measured.

Table 2 shows the comparisons between the telephone poll, conducted and weighted using the same standard ICM techniques described above (N = 1,000+ each time), and the four online Wisdom Index tests (N = 2,000+ each time) that ICM has conducted since the 2010 general election. The online Wisdom Index predictions are weighted only to standard demographics, not to past vote recall.

Table 2 Comparing telephone poll vote shares against online Wisdom Index vote predictions over time

	Conservative (%)	Labour (%)	Liberal Democrats (%)	Other (%)
21–23 January 2011				
Guardian telephone poll	35	39	15	11
Wisdom Index spontaneous	35	38	16	11
Wisdom prompted	33	36	18	13
18–20 February 2011				
Guardian telephone poll	35	38	18	9
Wisdom Index spontaneous	33	37	17	13
Wisdom prompt	32	36	19	13
4–6 May 2011				
Guardian telephone poll	36	37	15	11
Wisdom Index spontaneous	35	36	17	13
Wisdom prompted	34	35	18	13
22–23 June 2011				
Guardian telephone poll	37	39	12	12
Wisdom Index spontaneous[a]	35	36	15	14
Wisdom prompted[b]	34	35	17	15

[a] Question wording for every post-election tracker test: 'As a bit of fun, please tell me what percentage share of the vote you think the following would win in a general election if it were held tomorrow?'
[b] Question wording for every post-election tracker test: 'At the last general election in May 2010, the Conservatives won 37% share of the vote, Labour won 30% share, the Liberal Democrats 24 % share and smaller parties had 10% share of the vote. Knowing this, please tell me what percentage share of the vote you think the following would win if there were a general election tomorrow?'

Once again, there are a number of characteristics that are immediately striking, as follows.

- The lead between the Labour Party and the Conservatives in the telephone poll is replicated in direction on every occasion despite its relative narrowness, and in size to within no more than a one-point variation. This implies that the Wisdom Index is more than capable of reflecting the same central standing of the parties in ICM's regular polling series.

- The spontaneous Wisdom Index numbers are generally one or two points lower for both the Conservatives and Labour on each occasion. The prompted Wisdom Index numbers shave no more than a further point from both parties.

- The Liberal Democrats' share improves as a result, as does that of others – this is consistent with the pre-election test. The Liberal Democrats are the great polling mystery of the current political cycle, with their share of the vote overstated by all pre-election prediction polls in 2010, but not by the 2010 Wisdom Index prediction. Now, however, with regular online polls suggesting they stand as low as 8% share of the vote (although ICM has not been lower than 12% share), the Wisdom Index suggests they hold a greater share. It would seem (particularly with the prompted Wisdom Index projection) that the electorate coalesce around a narrower vote share range than the standard polls suggest. Prompting with the previous Liberal Democrat election share has the effect of pushing up their Wisdom share of the vote prediction in both the general election pre-test and post-election tracker tests. There is little evidence to explain why this might be, but we might speculate that there is a lower level of general understanding about previous Liberal Democrat performance; unless told, people might be inclined to underestimate their vote share starting point. Certainly, it is the case over the years in ICM's telephone vote intention polls that recall of Liberal Democrat election scores is more volatile than that for the other main parties, and past vote weighting schemes have taken into account a greater preponderance to misremember voting for the UK's third party.

- Other parties also enjoy higher vote shares at the present time, up to four points higher than telephone vote shares. This might be expected mid-term, particularly with declining fortunes for the coalition government, and a new opposition leader attempting to find his feet.

It would seem that the Wisdom Index suite is not only capable of producing the most accurate projection of a general election, but also of mirroring vote intention trends picked up by other polls, albeit with some important variations, which might well be controversial where they deviate from the media narrative that conventional polls suggest, but which are certainly not beyond the realms of reason. There is, of course, no right or wrong answer for a mid-term poll, and linking a Wisdom Index finding to a simultaneously conducted telephone poll finding is somewhat disingenuous given vagaries that could as easily be applied to the poll findings (or indeed other polling companies' findings) as they could the Wisdom Index findings. The message, however, is clear: the Wisdom Index has real potential as a complement, or even a competitor, to classic vote intention polls.

Past vote weighting

ICM polls have been past vote weighted since the late 1990s – a technique that grounds each poll in political reality, and indeed in theory matches its national demographic representation with political representation. Since its introduction, its impact on ICM prediction polls has been nothing but advantageous to their accuracy.

For many poll watchers, weighting by recall of past voting has become a test of poll quality. If it is applied, the poll has the political balance and necessary internal structure to be accurate. If not, then it might suffer from imbalances that lead to inconsistency and potential inaccuracy. ICM's founder, Nick Sparrow, was the architect of modern opinion polls and the pioneer of past vote weighting, so it is with some trepidation that ICM now presents evidence to show that the Wisdom Index does not require it. The need to apply measures to counteract sample imbalances, political imbalances, turnout variations and spiral of silence issues has no place in the construction of the Wisdom Index.[8] In theory they are not needed; in practice they are irrelevant.

The Wisdom Index poll outcomes prove that past vote weighting has negligible or no impact on vote share predictions. Table 3 reveals what the telephone prediction poll looks like with past vote weighting applied, or not applied. It also shows the same for the first and the latest online Wisdom Index polls, although any of the four could have been included in the table such is the consistency of the comparison. There has been no movement beyond one point in any party share across all the Wisdom Index polls, implying that the maximum impact of past vote weighting is to move a party share no further than the impact of rounding up or rounding down a percentage figure. A Wisdom approach can avoid the controversy and complexity of past vote weighting, because it has little practical impact.

The value of past vote weighting, then – vital in classic ICM opinion polls – is largely inconsequential in the Wisdom approach. There should be no surprise in this: a classical vote intention poll has to avoid any form of internal bias for it to translate individual vote intentions into an aggregated, accurate prediction. On the Wisdom approach, the smartness of the crowd is not conditional upon its precise composition. Moreover, to guess what might happen at a future election, crowd members do not bring to the table the factors that underpin classically conducted poll inaccuracy:

[8] Spiral of silence issues imply that some party supporters might not tell pollsters their intentions because they are embarrassed to admit them.

Table 3 Wisdom Index unweighted vs past vote weighted comparisons

General election prediction poll	Conservative (%)	Labour (%)	Liberal Democrats (%)	Other (%)
3–4 May 2010				
Wisdom Index spontaneous Demographically weighted only	36	34	21	9
Wisdom Index spontaneous Past vote weighted	36	34	21	9
Wisdom prompted Demographically weighted only	35	31	24	10
Wisdom prompted Past vote weighted	35	31	24	10
21–23 January 2011				
Wisdom Index spontaneous Demographically weighted only	35	38	16	11
Wisdom Index spontaneous Past vote weighted	34	38	16	11
Wisdom prompted Demographically weighted only	33	36	18	13
Wisdom prompted Past vote weighted	33	37	18	13
22–23 June 2011				
Wisdom Index spontaneous Demographically weighted only	35	36	15	14
Wisdom Index spontaneous Past vote weighted	36	37	15	13
Wisdom prompted Demographically weighted only	34	35	17	15
Wisdom prompted Past vote weighted	34	36	16	14

the differential likelihood of party supporters to contribute to a poll, personal embarrassment to admit which party they support, differential likelihood to turnout to vote, etc. None of this matters in the construction of Wisdom predictions.

If such factors do not matter, and the impact of past vote weighting is insignificant, it does not imply that party supporters are incapable of introducing their own political prejudices into their thought processes and guesses. We should remember that the diversity of the crowd is not only a necessary condition of a smart crowd, it is a vital one. Sub-group analysis

Table 4 Prompted Wisdom prediction by party voted for in 2010 general election

22–23 June 2011	Voted May 2010			
	Overall prompted Wisdom (%)	Conservative (%)	Labour (%)	Liberal Democrats (%)	Others (%)
Conservative	34	40	30	32	30
Labour	35	31	43	35	31
Liberal Democrats	17	17	14	18	15
Other	15	13	13	14	24

of the Wisdom predictions confirms that an anticipated correlation between party support and predicted vote share does exist. In the latest Wisdom prediction (June 2011) the mean prediction by party supporter for their own party was, without fail, higher than the mean prediction for that party overall.

The theory of the crowd suggests that overestimated predictions of supporters' favoured party will be cancelled out overall, but if any one market research panel used to draw a smart crowd contains materially too many or too few supporters of any one party, the diversity principle may be damaged, to the detriment of the prediction itself.

Referendum failure: the inaccuracy of the Wisdom Index

Indeed, the Wisdom approach cannot guarantee accuracy; it is more than capable of producing seriously misleading predictions. Early 2011 presented two more chances for ICM to test the concept against a political reality, with a referendum held in Wales on 3 March to confirm additional powers for the Welsh Assembly Government, and a national referendum on switching the system for electing MPs to the House of Commons from first past the post to the alternative vote (AV). In both cases, ICM had been commissioned to produce prediction polls: the first, a representative sample of Welsh people on behalf of BBC Wales; the second, an AV prediction on behalf of the *Guardian*.

Both predictions were undertaken by telephone; the former slightly overstated support for additional powers but confirmed the prevailing media narrative, while the latter exactly predicted the AV referendum result to one decimal place, making it the most accurate polling prediction on record in the UK.

For practical and cost reasons, the Wisdom question was placed at the end of the BBC Wales referendum telephone prediction poll, while

we utilised our online omnibus to predict the AV referendum separately from our *Guardian* newspaper poll. An additional methodological factor also applied: given that these referenda had not occurred before, we were unable to ask the prompted Wisdom Index question.

Table 5 reveals how the Wisdom Index performed in comparison to the main published predictions. The errors on both were large and consistent, being nine points awry of the actual outcome – seriously inferior predictions to those produced by ICM using our standard telephone methodological approach.

We can only speculate as to why the predictions were so poor, but Surowiecki (2004) again offers clues. He suggests that there must be at least some information for the crowds to be smart, citing the likely inability of a group of children to buy and sell stocks in Thiokol in the way that traders successfully managed after the Challenger disaster. Here, we cannot assume that the crowd had sufficient information to evaluate the referendum outcome adequately. When the crowd predicted the weight of the ox, all members had the advantage of being able to stare at it and evaluate the size of the beast. When our general election prediction proved accurate, most people had the advantage of both a basic understanding of British politics at general election time, and a prompted understanding of how each party had fared at the previous election. In short, they had enough information to be smart.

However, this may not have been the case in the referenda; both were characterised by the electorate's limited understanding. ICM's own BBC Wales referendum prediction poll revealed that almost as many people (48%) did not have enough information to make an informed decision on how to vote as did (50%).[9] As for the AV referendum, qualitative research

Table 5 Spontaneous Wisdom Index in the 2011 Welsh Assembly Government election and the 2011 UK alternative vote (AV) referendum

	Wales Assembly Government referendum 1–2 March 2011			AV referendum 4–6 May 2011		
	Result (%)	ICM telephone prediction (%)	Wisdom prediction (telephone) (%)	Result (%)	ICM telephone prediction (%)	Wisdom prediction (online) (%)
Yes	63.5	69	54	32	32	41
No	36.5	31	46	68	68	59
Error	**–**	**5.5**	**9.5**	**–**	**0**	**9**

[9] BBC Wales/ICM, March 2011.

conducted by ICM post-referendum found that swathes of the electorate were completely dumbfounded by the AV proposition, and some reported stories of friends and colleagues not even being aware that a referendum was occurring.[10] Under such circumstances, we speculate that the crowds' assumptions were superficial, and their ability to predict was limited. It is perhaps enlightening that, in both predictions, there was coalescence around the median point of 50%. Just as coin tosses even themselves out over time, so might the guesses where knowledge was limited or absent.

Caution, then, is strongly advised when used in scenarios where prompting cannot be employed. The evidence so far suggests that the prompted Wisdom question is a superior predictor to its purer spontaneous counterpart. Prompting respondents may perhaps confound a central principle of the Wisdom of Crowds, but in doing so it seems to improve accuracy where knowledge cannot be assumed or depended upon.

The practical application of the Wisdom Index in market research

The political polling environment in the UK has been reshaped over the past 15 years, first by the introduction of telephone polling, associated weighting and adjustment schemes, and more recently by the emergence of multiple online agencies that offer large-scale samples quickly and cost effectively, but sometimes inaccurately.

Within such an environment, the addition of the Wisdom Index could add to the feeding frenzy, either where small online agencies seek to build their own brand through high-profile polling work, or even where individual websites poll their own members to produce vote share projections. There is certainly a time and place for the Wisdom Index, but whether that place is as a complement to classically conducted opinion polls or as a competitor is open to debate.

The failure of the referendum tests raises questions about the information people need to contribute to a smart crowd. For sure they must have some information or else the diversity condition is compromised – but where do they get that information from? Of course people will dig in to their own historical pool of knowledge to seek answers, but those answers will also be updated by current political narratives that are themselves premised in part on contemporary voting intention polls. In short, Wisdom predictions may be only a 'competitor' to standard voting intention polls so long as the latter are being conducted. In this sense, the Wisdom Index may best

[10] ICM/Electoral Reform Society, 2011.

be viewed as an interesting set of additional questions on a vote intention poll, rather than the central focus of a poll itself.

There is a further quandary to contend with. We have seen that methods used by pollsters to force a poll sample to be representative have little impact on Wisdom outcomes. However, we also point out that a crowd mostly or entirely comprising a specific mind-set, such as right-of-centre Conservative Party voting, is unlikely to be smart. The evidence suggests that such a crowd would most likely overstate the vote share that the Conservatives end up winning. Implicit in this is something of a contradiction: on one hand we argue that a crowd's composition is of no material concern, and yet on the other we show that if the diversity principle is compromised, the crowd will not be smart. We cannot have it both ways, and some future work needs to be done on the precise nature and composition of the crowd that can be counted on to be smart, or at least on the conditions that must be avoided for it to have a chance of being smart. One of those conditions is, of course, the size of the crowd. On this we do have some interesting evidence.

Surowiecki (2004) explained his work with examples of quite small crowds being smart. To date, ICM's poll 'crowds' have comprised 1,000- and 2,000-strong samples because it has been convenient and easy for us to do so using our omnibus services. However, market researchers will no doubt be interested to learn that much smaller sample sizes are capable of being equally or even more accurate.

Our 2010 general election prediction poll was based on a total sample of 2,022 respondents, producing a spontaneous Wisdom prediction of 2.2%. We stripped out interviews on this poll and looked at the predictive capability of the remaining randomly selected and demographically weighted sub-samples of 500, 250 and 100 respondents. Table 6 shows that the impact on the spontaneous prediction was minimal, worsening

Table 6 Spontaneous 2010 predictions based on varying samples sizes

	Election result	Wisdom spontaneous prediction *N* = 2,022	Wisdom spontaneous prediction *N* = 500	Wisdom spontaneous prediction *N* = 250	Wisdom spontaneous prediction *N* = 100
Labour (%)	29.6	34	34	34	36
Conservative (%)	36.8	36	37	37	35
Liberal Democrats (%)	23.8	21	21	21	23
Others (%)	9.8	9	8	8	6
Average error	–	**2.2**	**2.3**	**2.3**	**3.2**

the average error by only 0.1 point to 2.3% for 500 and 250 interviews, before the error widens materially at a sample size of 100 interviews.

The outcome on the more accurate prompted version is even more dramatic. Table 7 reveals that, the smaller the sub-sample, the greater the accuracy of the prediction up to the 100 sample size option. The average error on 500 interviews fell to 0.6%, and to 0.5% for 250 interviews, but widened to 1.5% for 100 interviews. On this basis, it almost defies belief to report that a crowd comprising only 250 people is perfectly capable of producing smart and accurate predictions of future election results. A sample size that is lower than this should not be considered.

On this basis (but with innate reservations that do not correspond with the facts), we conclude that the Wisdom Index could easily be constructed as a competitor to conventional opinion polling – anything is possible with an approach so grounded in simplicity, and where almost any diverse crowd can be sourced. Certainly, any market research company could do so with ease. But the concept has further value outside of election predictions: bookmakers would, no doubt, be interested to hear that the Wisdom Index could identify winners and losers in various high-profile gaming situations, while big consumer brands might be interested to understand how their brand is predicted to perform in the future, rather than what public perception of it looks like now. No doubt researchers have and will find profitable avenues to explore.

However, it is in the realm of electoral predictions that the Wisdom Index must prove its worth. It could be, in one major way, a revolution in electoral opinion research, turning the traditional relationship between representative sampling and aggregating the individual's own behaviour on its head. But, given the variable performances of the traditional opinion polls in the UK and elsewhere over the past 20 years, the conversion from 'I will vote for party X' to 'We think this will be the outcome' may stand

Table 7 Prompted 2010 predictions based on varying samples sizes

	Election result	Wisdom prompted prediction *N* = 2,022	Wisdom prompted prediction *N* = 500	Wisdom prompted prediction *N* = 250	Wisdom prompted prediction *N* = 100
Labour (%)	29.6	31	31	30	32
Conservative (%)	36.8	35	36	36	35
Liberal Democrats (%)	23.8	24	24	24	25
Others (%)	9.8	10	9	9	9
Average error	**–**	**0.9**	**0.6**	**0.5**	**1.5**

as much chance of success in accurately predicting election outcomes as traditional polling methods.

Acknowledgements

With thanks to Matthew Burn and Professor John Curtice for inspiration and advice.

References

Earls, M. & Kearon, J. (2009) ME-TO-WE RESEARCH. From asking unreliable witnesses about themselves to asking people what they notice, believe and predict about others. ESOMAR Congress research paper.

Curtice, J. & Sparrow, N. (2010) Too clever by half? The impact of weighting and adjustments on the accuracy of the polls. *International Journal of Market Research*, **52**, 5, pp. 678–686.

Sparrow, N. (2007) Developing reliable online polls. *International Journal of Market Research*, **48**, 6.

Surowiecki, J. (2004) *The Wisdom of Crowds. Why the Many are Smarter than the Few.* USA: Doubleday.

About the author

Martin Boon is a Director at ICM Research, having worked at the company for 18 years. He heads both the Social & Government research team and the Political Polling unit at ICM, taking full responsibility for the latter in 2009. He led the company's successful 2010 general election polling and subsequent success in the 2011 AV referendum, in which ICM predicted the exact result to one decimal place.

ICM Research is a major opinion and market research company. It offers a full range of quantitative and qualitative techniques to political and business clients.

Address correspondence to: Martin Boon, ICM Research, Berkshire House, 168–173 High Holborn, London WC1V 7AA.

Email: martin.boon@icmresearch.com