Expert Review of Survey Methodology: paper to Steering Group

Document No:	245
Status:	Final
Title:	Expert Review of Survey Methodology: paper to Steering Group
Author:	3KQ and PSE sub-group
Notes:	Decision taken at 24th November 2011 Steering Group meeting

1 - Aim

This paper asks the Steering Group to take a decision on whether to change the methodology being used for the opinion survey at the end of PSE3, in the light of expert review results.

2 - Background

The Partnership has valued public opinion highly since the start of its work. This manifests itself in two ways. First, much effort has been put into public and stakeholder engagement (PSE). Second, the Partnership has stated that it will run a representative public opinion survey to gauge whether there are more people 'for' participating in the siting process than 'against'. This is called 'net support'.

The Partnership has agreed that *"it is right to use an opinion survey rather than a referendum at this stage in the process… and that a referendum might be an appropriate tool to use at some point in the future"* (Doc 176, para 6.3). However, there is sensitivity in the Partnership, as well as beyond, about the robustness of the actual methodology used to deliver the opinion survey. People want to know it can be relied upon within reasonable limits.

The Partnership, via the PSE sub-group, commissioned two expert reviews of the methodology being proposed. Both reviews are now complete and have been discussed by the PSE sub-group.

The Partnership agreed that *"the PSE Sub-Group will communicate the outcomes of its work to the Partnership for it to review, to ensure that members are convinced of the robustness of the survey and the methodology"* (Doc 176, para 6.3). As the PSE sub-group's work may lead to significant high profile changes, a Steering Group decision is required on what change to make to the methodology. This decision will in turn need communicating to the Partnership together with the full documentation.

3 - Current Methodology

In simple terms, the current methodology comprises of an independent contractor conducting 10-15 minute telephone interviews with 3000 people across Cumbria. This sample is split into 1000 in Allerdale, 1000 in Copeland, and 1000 in the rest of Cumbria. Quotas are set to ensure that interviews are achieved in the correct proportions (i.e. reflective of the whole population) in terms of age, gender, working status etc. This method is commonly used across the UK as a relatively reliable and comparatively cheap way of conducting opinion surveys.

4 - Expert Reviewers
A procurement process was undertaken to recruit up to three expert reviews of the methodology. 13 bids were submitted, and two chosen: Dr Sandy Ochojna (independent consultant) and Prof Patrick Sturgis (University of Southampton). Both appear to be very experienced and well respected in the field and the PSE subgroup had no reservations about appointing them. Their reviews are attached to this paper and should be read in full so Steering Group members are aware of the sensitive nature of their comments.

In summary, there are various aspects of the methodology that the reviewers either support, or make suggestions on things that will be easily factored in to the survey when the contractor is in place. There are however two keys things that the reviewers are questioning:

- Whether the sample is either truly random or quota based.
- Whether telephone or face-to-face interviewing is more appropriate.

These two points are covered below, together with the view of the PSE sub-group on each point.

5 – Whether the sample is either truly random or quota-based
The currently proposed method uses what is called a 'quota' sampling approach. In the first instance a list of telephone number is generated by a computer, this is done within area codes to ensure they cover the geographic areas of interest.

Interviewers then work through that list calling each number in turn. If nobody answers the phone that number is discounted and the interviewer moves to the next number.

The interviewer continues to make calls until they have completed the necessary number of interviews to fill the quotas (i.e. of women, men, different age groups, people in and out of work etc). This method is based on the assumption that people with similar characteristics will hold similar opinions, so if we achieve interviews in the right proportions our sample will be reflective of the views of the wider population.

This assumption is the key focus of criticism of this method as it introduces a potential source of bias into the survey findings.

A further problematic issue with quota surveys is something called 'non-response error'. This arises because the number of people picking up the phone and agreeing to be interviewed is low (typically around 5% of all the numbers generated), and so we are forced implicitly to make the assumption that the people picking up the phone and agreeing to be interviewed have exactly the same views, on balance, as those people not picking the phone up, or refusing to be interviewed. Although this system is very commonly used, the reviewers point out that this assumption provides another potential source of criticism.

The alternative approach is to use a 'truly random' approach. If a telephone methodology was used this would again involve a list of telephone numbers being

randomly generated in the same way as above. But this time the interviewers must try to ensure a response is received from every number on the list.

This means that if they do not receive a response the interviewer will call back to the same number up to 5 times at different times of the day and over a period of a few weeks, before 'dropping' it. The number is not called back if someone answers and refuses to participate.

The interviewer does not attempt to fill quotas. The assumption is that because the original list of number is random if interviews are achieved for each number then the resulting achieved sample will also be random, and therefore reflect the wider population. Additionally the approach is less prone to non-response bias because significant effort is made to achieve an interview with each number on the list, this means the response rate is higher (typically around 50-60%) and we can be more confident that the people that complete the interview are more likely to be reflective the wider population.

This truly random approach is considered the optimum in terms of statistical theory.

The implications of switching from a 'quota' to a 'truly-random' method are that it would cost about 10-15% more, and take around 8-10 weeks rather than 2 weeks to do the fieldwork.

PSE sub-group view. Given that both expert reviewers agree on the existence of non-response error and the assumptions inherent in quota surveys, and that addressing them can be accommodated fairly easily, **the PSE sub-group view is that the switch from 'quota' to 'truly-random' sampling should be made.** Appropriate budgeting for next financial year can accommodate the limited extra funds required (if any), and planning around timescales can minimise any delay to the programme. The extended fieldwork would occur in parallel with the reporting from PSE3, so little time would be lost.

6 – Whether telephone or face to face interviewing is more appropriate

The currently proposed method uses telephone interviews, rather than face-to-face interviews. The two reviewers appear to disagree on which is the most appropriate interview technique given the situation in West Cumbria.

Dr Sandy Ochojna says, *"telephone interviewing is by far the most appropriate fieldwork method for an area of widely dispersed populations*[1]*".* In contrast, Prof Sturgis takes a strong line saying that face-to-face interviewing is the *"gold standard" for* various reasons: to ensure that respondents can be supported in understanding the issues properly with prompt cards etc before they give their views, to further reduce non-response error, and also to reduce error arising from non-coverage i.e. those people who have an address but no landline number and

[1] Although he does recommend shortening and simplifying the questionnaire to minimise the likelihood of respondents becoming confused. This is being addressed by the PSE subgroup.

so would automatically be excluded from the sample (around 10% of the population by recent estimates).

The implications of switching from a telephone to a face-to-face method are firstly an extension of the fieldwork duration a month to 10-12 weeks, and secondly and an extension of costs from £75K to a minimum of £250K. The comparison is below.

	Cost	Fieldwork Duration
Telephone	£75,000	8-10 weeks
Face-to-face	£250,000 – 300,000	10-12 weeks

N.B. Approximate figures, assuming a truly-random approach to sampling in both.

PSE sub-group view. It is inherently impossible to define how much benefit will arise from switching to face-to-face interviewing, which leaves a difficult decision for the Partnership. Is it worth tripling the cost? Overall, the group notes:

- The strongly worded phrases in Prof Sturgis' review that will almost certainly be used against the Partnership in the media should it choose to employ either a quota or a telephone approach.
- Prof Sturgis' rough estimate that the original methodology proposed is a '4' on a scale of 1 to 10, a truly-random telephone methodology (as recommended above in section 5) would be a '6', and his recommended gold standard is a '10'.
- Dr Ochojna's view that the random telephone methodology is by far the most appropriate, although we do note that he is not willing to specify where on a scale of 1 to 10 the various approaches sit.
- A change in interview methodology could lead to changes in how people answer the questions. This could mean that the results of the previous three awareness tracking surveys may not be a good guide to the likely results from this survey[2]. This potential for a discrepancy in the findings between the survey waves could be used to challenge the results (from people in favour or opposed) .
- The cost of face-to-face interviewing is around that estimated for running a referendum, so may blur the rationale for not doing so[3].

Overall, the PSE sub-group feel that the majority of the deficit in our approach that Prof Sturgis focuses on can be addressed by switching from quota to fixed-random sampling (as per section 5). On balance, the PSE sub-group does not feel that it is worth tripling the cost in order to gain a certain, but essentially indefinable, increase in reliability. However, the potential for some players to undermine the Partnership's work is significant and does need to be weighed against the extra cost. The PSE sub-group ultimately acknowledges that this is a political decision, so leaves it to the Steering Group.

Note: the reviews from Prof Sturgis and Dr Ochojna are attached. It is recommended that they are read in full.

[2] For example, CCC usually conducts the Place survey via postal interview, and finds that around 30% of residents are content with the council's performance. When conducted once by face-to-face interview, this figure jumped to around 70%.
[3] Even though cost is not a stated reason for using opinion surveying instead of a referendum. The main reasons are to do with wanting a representative view (not the views of self-selected people), and that we don't know enough information yet to fairly hold a referendum.

Comments on Proposed Methodology for undertaking a survey of opinion toward siting of a nuclear waste storage facility in West Cumbria

Patrick Sturgis, October 2011

Prepared for the West Cumbria Managing Radioactive Waste Safely (MRWS) Partnership.

Background and proposed design

The proposal is to undertake a survey of public awareness of and opinion toward the proposal for West Cumbria to participate in the siting process for locating a geological nuclear waste storage facility. The target population for the survey is all adults aged 16 and above in Cumbria. Interviews will be undertaken by computer assisted telephone interview (CATI), using quota sampling on telephone numbers generated via random digit dialling (RDD). Quota controls on age, sex, and working status will be used to match the sample to population distributions from the ONS 2007 mid-year Census estimates.

There is to be an over-sample of households in the boroughs of Allerdale and Copeland such that 1000 households will be interviewed in each area and 1000 households in the rest of Cumbria (total achieved sample size = 3000). The supporting materials do not detail why these areas are over-sampled but I assume that it is because they are likely locations for the storage facility (if it goes ahead in Cumbria), so the residents here would be disproportionately affected by the potential risks and benefits.

The partnership will only recommend that West Cumbria proceed to the next stage of the siting process if 'net support' (total in favour minus total not in favour) is greater than zero.

Comments

My comments here are limited to the proposed methodology for undertaking the opinion survey and do not touch on the democratic legitimacy of the exercise, or the wisdom of using opinion polls to guide decision-making on contentious issues of this nature.

My overall assessment is that the proposed design is *capable of* providing a reasonably accurate estimate of public opinion toward participating in the siting process. However, the use of non-random sampling and telephone interviewing cannot be considered the 'gold standard' design for this purpose. Particularly as a result of the high rates of non-coverage and nonresponse that are bound to arise in the methodology proposed, there is clear potential for the survey to produce *biased* estimates of public sentiment. This will be the case if the probability of responding to the survey is correlated with attitudes toward the proposed policy.

Additionally, because the key questions in the survey relate to a proposal about a technology that many Cumbrians will be unfamiliar with, the questionnaire contains quite lengthy explanations about the nature of the process and how it

will work. Such a questionnaire format is not well suited to CATI administration, as many people will find it difficult to follow on a telephone.

Because there is no independent means of assessing the 'true' distribution of opinion on the outcome of interest (if there were, there would of course be no need for a survey), there is a high risk that those on either side of the debate will question the results of the survey on the grounds that a 'sub-standard' methodology has been employed, if the results do not support their position.

The methodological brief justifies the choice of quota CATI on the grounds of cost and speed of turnaround. While cost may be a reasonable justification for not employing a random probability design (assuming that the cost would actually exceed the partnership's budget, a case which is not in fact made in the briefing document), the speed of turnaround argument appears weak in the context of the overall time-frame for the siting process. Even the cost argument is liable to cut little ice with those who seek to question the validity of the findings; 'we couldn't afford something better' will not be very persuasive in the context of a decision of such local and national importance.

Elaboration
Given the high rates of non-contact and refusal that are standard for telephone surveys in the UK, there is clearly substantial scope for nonresponse bias arising out of the proposed design. No information is provided in the briefing document about likely non-contact and refusal rates but it is well known that response rates to (RDD) CATI surveys can easily fall into single figures. Given that the proposed design is to use quotas, the true response rate is almost certainly going to be considerably lower. The design is thus proposing to use a sample with a (notional) response rate of, at best, 5% (and in all likelihood, considerably lower) and assuming that (conditional on the quotas and weighting adjustments) there is no difference on the survey outcomes between the responding and nonresponding (and non-covered) units. This seems an heroic assumption. It would be useful to ask Ipsos MORI what the refusal rate was to the previous rounds of the survey, although it is likely that this information was not recorded.

I found the briefing document to be rather short on detail in the consideration it gives to the advantages and disadvantages of different survey modes. The primary reason given for preferring quota CATI to random CAPI and random CATI is that that the random designs will take longer, with around 3 months of fieldwork required to obtain acceptable response rates, possibly more.

However, given the high stakes involved in this exercise, with an accurate estimate of opinion being of the utmost importance, it seems strange to dismiss the idea of a more accurate mode on the grounds that data collection will take 3 months. After all, as the background information provided to me points out, the process of which this exercise forms a part has been going on for many years and is likely to continue for many more to come.

In short, I did not find the 'longer time' justification to be convincing as a reason for preferring the quota CATI to the random face-to-face and random CATI

modes. It may be that, given where the process is in the timetable now, it is genuinely not feasible to undertake a random survey, whether in face-to-face or CATI modes. However, given the duration of the overall process, it seems to me that either would have been possible within the time available for reporting.

A further, and not inconsequential reason for preferring quota to random designs is cost, although no figures are presented, nor reference made in the briefing document to available budget. It is almost certainly the case that a quota survey will be cheaper than a random survey (holding sample size and mode constant) and, if the cost of the random alternatives exceed the available budget, then it is probably correct to conclude that the quota CATI approach is the best (or least worst) alternative, although there could well be a case for using random CATI, or even a random postal survey but these are not properly explored.

I have argued strongly in favour of the random option, now I expand on my reasons for taking this position, starting first with the random Computer Assisted Personal Interview (CAPI) gold standard.

The primary advantage of a CAPI design is that it enables the use of random sampling techniques, and generally yields high (relative to other modes) response rates. Additional measurement benefits accrue from the presence of an interviewer, who can use show-cards and the like to enable clearer administration of questions. Given the relatively high salience of the issue of nuclear waste siting in Cumbria, I would imagine that a well-conducted CAPI probability survey could achieve a response rate in the region of 60-80%. Although response rate is only an indicator as opposed to a direct measure of nonresponse bias, it is certainly the case that the potential for serious bias increases as response rates approach zero.

Random probability CAPI surveys (using standard approaches for random probability surveys in the UK) also provide better coverage of the target population than CATI surveys. The Post Offices Address File (PAF) is generally used to draw samples for CAPI surveys in the UK and has coverage rates approaching 100% for residential households (there is a small coverage issue with regard to institutional populations, although this is generally considered to be minor for most population parameters). For telephone surveys, however, there is a much bigger coverage problem, due to the substantial (and growing) minority of people who do not have a landline at all and who are different in important respects from the 'landline' population (younger, less affluent, less politically engaged, etc.). To the extent that such individuals have different opinions about the siting proposal relative to those who do have a landline phone, key survey estimates will be biased. The briefing document notes this problem but proposes to do nothing about it (beyond the use of quotas and weights) on the grounds of practicality and cost. While such arguments may hold some weight, they do little to address the bias that is likely to result from non-coverage.

For both nonresponse and non-coverage error the briefing document's proposed remedy is quota controls and post-survey weighting adjustments. Interviews

will be undertaken such that the achieved sample closely matches the 'known' population distribution on age, sex, and working status from the ONS 2007 mid-year census estimates. The raw data will then, presumably, be adjusted via post-stratification weights to reflect the population on a range of variables that are not specified in the briefing document (in fact, the briefing implies that the data will be weighted by the same variables used to produce the quotas but this would seem to be an odd strategy if the quotas have been achieved). For this approach to be successful in removing bias from the raw estimates, we must assume that a) the quota and weighting variables are the only variables that are predictive of both nonresponse and survey outcomes and b) that the distribution of public opinion is *the same* for both respondents and nonrespondents within each of these variables (e.g. women who do not have a landline or who decline to take part in the survey are no different in their opinions toward the siting process than women who take part in the survey and provide an opinion). In my view, these assumptions are very fragile and there is a high chance of nonresponse bias if they are violated.

We might speculate on the likely direction of nonresponse bias if the proposed (or, indeed, another design) were implemented. Would it produce an estimate of public opinion which is more supportive or more opposed to participating in the siting process? This is, of course, a difficult question to address and any answer can only be regarded as speculative and highly imprecise. However, it is well known that one of the strongest predictors of nonresponse is interest and engagement with the topic of the survey. This is why surveys about topics such as crime and health generally achieve higher response rates than surveys about expenditure or about science and technology. In the current instance, interest and engagement is likely to be strongest amongst those with more polarised attitudes, on either side of the debate. In the most fortunate case, we might imagine a scenario in which nonresponse bias is self-cancelling, with supporters and opponents outweighing those with a more ambivalent position in more or less equal numbers and the survey estimate converging on the true value as a result. However, a more likely scenario is that interest and engagement will be strongest amongst those who are opposed to the proposal on the grounds that, in most policy contexts, opposition is a stronger mobilising force for action than support. If this were to be the case here, the direction of the nonresponse bias would be to over-estimate opposition to the proposal. However, as I noted above, this is entirely speculative and bias in the opposite direction, or no bias at all could easily pertain. My intention is not so much to diagnose the likely bias bu to give an indication of how bias might arise and how we might think about its probable direction and magnitude.

A consequence of using weighting to adjust for under-representation of population sub-groups in the sample is an increase in the variance of estimators. This means that, for a given sample size, confidence intervals will be wider if weights are used than if they are not. The increase in variance due to weighting (the design effect) is proportional to the variance of the weights. Thus, using weights to reduce bias comes at a cost of loss of precision, it is in effect a trade-off between the two sources of error. This weighing design effect is discussed in the report for wave 3 of the survey, although no mention is made of the likely

effect on estimates in the methodological briefing document. The likely impact of design weights on the precision of estimates should be considered more explicitly in the methodological briefing document. Where a design effect is particularly large, it is standard practice to truncate the weights at some point. It would be useful to discover whether truncation was undertaken for the existing surveys undertaken by Ipsos MORI.

In addition to the nonresponse and non-coverage advantage, CAPI surveys are better suited to asking longer and more technically complex questions because questions of this nature are more difficult to follow on the telephone. Additionally CAPI surveys can use show-cards to aid respondents in seeing and remembering the full range of response alternatives.

An alternative to CAPI, but which retains random sampling procedures would be to use a true RDD CATI design. The briefing document rejects this approach on the grounds that fieldwork would take too long (and costs would be higher as a result). As I have already noted, I do not consider this to be a persuasive justification. However, it is also the case that the benefits of moving to a true RDD design would be considerably less than moving to random CAPI because the nonresponse and non-coverage problems would remain substantial in a random CATI approach. Nonetheless, a random sample design would be more defensible against potential critics of the process because it adheres to the principles of scientific sampling, which quota sampling does not. This is not an esoteric point; the principles that enable us to draw inferences about populations from samples are based on the properties of sampling distributions and these are only relevant in the context of samples which are drawn at random. The concept of a confidence interval is hard to justify when a sample is not selected at random.

The methodological briefing note states that the decision on recommendation to proceed will be based only on the results from the Allerdale and Copleand samples. However, it does not state whether this will be a combined estimate, or whether each area will be considered independently. This has implications for the sample design because, if a combined estimate is to be produced then Alledrale would need to be down-weighted in order to reflect the relative proportions living in each area. This weight would have an effect on the precision of the estimate, as discussed above. An alternative strategy would be to draw samples which are proportional to the size of the population in each area (i.e. 1120 in Allerdale and 880 in Copeland). Allocating the sample in this way would yield a more precise (combined) estimate for the same total sample size. Of course, if separate estimates are required, this strategy would result in confidence intervals that are too wide for Copeland.

To summarise, although the proposed sample design can be considered to be more or less 'state-of-the-art' with regard to current practice for quota CATI surveys in the UK, I have reservations about the high potential for non-coverage and nonresponse bias in the survey estimates using this approach.

Questionnaire
No questionnaire is currently available, although the version used in wave 3 of the Awareness Tracking survey (conducted by Ipsos MORI) can serve as a guide to the likely content of the proposed survey. I have no particular comments about the questions relating to demographic variables, or about other issues of concern in the area.

For the key questions relating to awareness and approval of the siting of nuclear waste, I have the following observations:

Many Cumbrians are not aware, or are only very dimly aware of the fact that the government is looking for a site at which to store higher activity nuclear waste and that West Cumbria is investigating the possibility of becoming that site. This poses a difficulty for asking questions about approval; how can someone approve or disapprove of something that they are unaware of? There are two primary approaches that one can take when asking questions about such issues. The first is to ask people at the outset if they are aware of the issue(s) in question. If they are not, then no further questions are asked about their attitude for or against. This has some disadvantages, notably that many people do not express an opinion either way, so the distribution of opinion is derived only on those who are already aware of the issues. This would not be considered a very democratic way of making a decision that affects everyone, not least because those who are unaware tend to be disproportionately located within lower socio-economic groups. Additionally, it is possible that some people will select this option as a way of avoiding answering further questions, when they could in fact express an opinion ('satisficing' in Jon Krosnkick's terminology).

An alternative approach, and one that is adopted in the wave three questionnaire, is to provide a 'preamble' in which the issues are set out as succinctly and accurately as possible. Respondents can then give their opinion based on the information provided, even if they were completely unaware of the technology or the politics of the matter before the interview commenced. While this approach has the merit of increasing the rate of opinion holding, it places a very large burden of neutrality on the text employed in the preamble. Depending on the content of the preamble, the distribution of opinion might change markedly. In addition to enabling those who are unaware to provide an opinion, the provision of a preamble might also *change* the opinions of those who had heard about the issues prior to taking part in the survey. So its content becomes absolutely key.

For the second of these procedures to be an acceptable way of gauging public opinion it is, in my view, essential for the wording to be produced in an open and transparent manner and, if at all possible, to be acceptable and agreed in advance by all interested parties and stakeholders. Without this consensus, it is likely that the results of the survey will be contested on the grounds of the impartiality of the preamble used.

An additional problem with the use of an explanatory preamble is that (particularly for those not engaged with the issues) it can be off-putting and

difficult to follow. This can result in break-offs from the survey and people basing their responses on information that they have not properly understood. I note that the preamble in the current version of the questionnaire comprises 8 sentences, many with sub-clauses containing rather a lot of technical and scientific terms. There must be a question about how well respondents are able to follow this information and whether they have sufficiently understood and digested it by the time they are asked to provide their opinions. As I have already noted, these difficulties of comprehension are likely to be compounded when the interview is administered on the telephone (relative to face-to-face).

A second question relates to the administration of the questions. Respondents are asked 'could you please indicate how strongly you agree or disagree with the following statement….'. The answer categories run from strongly agree, through tend to agree, neither agree nor disagree, tend to disagree, strongly disagree. If this question (and the ones that follow it, which use the same approach) were administered via CAPI, the interviewer would hold up a show card containing all the response alternatives and the respondent would select the one that most closely corresponds to his or her position. However, as this is a telephone interview, it is not clear to me how the respondent is intended to interpret 'how strongly you agree or disagree' *qua* the available alternatives (which they will not have heard or be able to see at this point). A respondent might well say something like, 'I'm totally against it'. At which point the interviewer would either have to interpret this in his/her own way (probably as 'strongly disagree') or ask the respondent to clarify 'so would that be strongly disagree?'. The point here is that there is a lack of correspondence between what the respondent is asked and the list of available response alternatives, which leaves room for error – the interviewer might wrongly interpret the response as being 'tend to disagree' when it should have been 'strongly disagree'.

It is common in telephone surveys to use a 'branching' or 'unfolding' approach for this type of unipolar attitude item. That is, respondents are first asked 'do you agree or disagree' (perhaps with an additional, or are you undecided?'). Depending on their answer to this question, they are then followed up with 'and do you tend to agree or strongly agree'? I am surprised that this approach does not appear to have been used here and suggest that Ipsos MORI be asked if there was a reason for not doing so.

My final point relates to the choice of measure of public opinion to determine whether to propose proceeding to the next stage of the consultation process. Although no detail regarding the derivation of this measure is provided in the briefing document, I note that the measure of 'net support' appears to be the simple difference between those who are in support (presumably defined as those selecting strongly or tend to agree) and those who oppose (those selecting tend to disagree or strongly disagree). If this is the case, then those selecting don't know (and presumably 'neither/nor') are effectively dropped from the data. There is a good case for arguing that a better measure of a majority in favour would be the simple proportion of those selecting 'tend to' or 'strongly agree'. The partnership should be prepared to defend its choice of measure

against those who are likely to argue, reasonably in my assessment, that the current choice advantages the case for proceeding.

Postscript

I have been asked to provide an additional response to the question *"To what extent can the Partnership and public rely on the results from an opinion survey carried out via the approach outlined (RDD CATI)? For example, where on a scale of 1 to 10 of reliability does our old methodology lie, and where does our new methodology lie (if 10 is the ideal-no-expense-spared-face-to-face approach, and 1 is an approach that cannot be relied on at all)."*

First, I should point out that my discussion in the text above gives the best account of my assessment of the relative advantages and disadvantages of the available alternatives. I do not feel entirely comfortable boiling this rather complex set of factors down to a 'reliability number' on a continuous scale. Nonetheless, I will attempt to do so as a (very rough and with clear 'health warning' attached) further guide to my assessment of their relative merits.

One of the reasons why this is a difficult question to answer is that surveys are not just topic and population specific but *question specific*. That is, one variable in a survey may yield a very biased estimate while another variable *in the same survey* may be completely unbiased. Thus, it is very difficult to state categorically that a particular mode will definitely yield more accurate estimates than another in any specific context. Having said that, however, there are good reasons for preferring a random over a non-random sample design; we can use the principles of scientific sampling to draw inferences within explicit confidence intervals; the rates of response (and nonresponse) to the survey are clear and explicit; and the degree of nonresponse bias is likely to be lower. So, let us assume that we allocate the number 1 to a survey design which had, say, a small number of interviewers intercepting shoppers in a mall with a clip-board and the number 10 to a random CAPI survey conducted by a reputable agency according to standard procedures in the UK and with a response rate of 80% or above[1]. I would give the previous survey design (quota CATI) 4 and the proposed new methodology (of which I have yet to see the details) (RDD CATI) 6.

[1] Although, note that this design could still yield a biased estimate due to nonresponse and/or measurement error.

West Cumbria MRWSP

A review of your proposed survey method

by the survey doctor

The survey doctor

1. Introduction

This note considers your Draft Briefing Note re Public Opinion Surveys. It assumes that the questionnaire which will be used in this final phase of opinion surveys will not differ significantly in nature or substance from those used in earlier waves (although at the appropriate stage I will be making several suggestions about how these earlier questionnaires might be changed).

This note also considers how your final deliberations might be best incorporated into the ITT.

2. Your proposal

2.1 The structure

The survey you are contemplating will generate much controversy, as indeed have the earlier phases. It is essential then that the debate it brings about revolves around its findings rather than its mechanics. Hence my comments in this note address three of the four basic building blocks of any successful survey, namely

Survey universe	who do we ask?
Sample size	how many do we ask?
Questionnaire	what do we ask?
Fieldwork method	how do we ask it?

All these aspects are equally important. My comments are based on a careful reading of the documents on your website, particularly those relating to your objectives, and to the results of the earlier Ipsos MORI public opinion surveys and their related 'discussions'.

The issues raised have no right or wrong answers. Rather they must be addressed so that you have clear, concise, defensible and readily understandable reasons for doing whatever it is you finally decide to do, whether it be on the grounds of practicality, time, or cost. The key issue is one of **proportionality**. Whatever is the final outcome there will be those who say that the survey was too expensive, and others who will say it was not accurate nor big enough.

2.2 Survey Objectives

It has to be acknowledged from the outset that while the objective is to gauge net public support to proceed to the next stage of discussion with Government on possible sites, without any presumption of agreement, many respondents, especially those with a view for or against having a site in West Cumbria, will see this survey as being one about whether or not there should be a storage site in the area. Those against will be seeking to stop the process at this early stage by generating opposition to further consultation, while those for it will support the process as the next step on the way.

The wave 3 survey results show a strong correlation between the two stages; amongst those in favour of a facility in the area, net support for discussions with Government is +77: amongst those against a local facility, the net support is -83.

2.3 Survey Universe...and proportionality

The universe is defined as all adult residents aged 16 and over: a local one of the Districts of Allerdale and Copeland, and a rest–of-Cumbria one. This seems a fair distinction to maintain. However the definition of these universes might merit some reconsideration.

For one, the definition may not be solely one of people or location, rather it could be one of time. The effects of siting in the area will be very long term, but of even more significance, the construction and implementation phase itself is still many years away...so, who has the right to say 'yes' or 'no' today? Such an intergenerational factor was touched upon on wave 1, but in practical terms of course we can only ask those who are here now. Perhaps however, in assessing the responses specific attention should be given to younger respondents, those who intend to be living in the area in twenty years time, and those with young families. The same short term/long term timeframe consideration needs to be applied to the motivations of those working in the nuclear industry compared to those who do not. But the universe for this survey – all adults in Cumbria -

should remain; it is the possible demographic questions that need to be fully considered for analytic and presentational purposes.

For another, it is strongly recommended that you reconsider and convince yourselves that having two sample areas is still 'valid'...in the public arena at least. While the two Districts could merit special interest because they have by far the highest incidence of nuclear industry employees and skills, (mis)perceptions of the size of the Fukushima exclusion zones could mean that all of Cumbria could be seen as meriting similar attention. Clearly this has repercussions on Doc 74 and the meaning of 'local area net support'.

2.4 **Sample Size**... and proportionality

2.4.1 Quotas

The quota sampling method has been well defended in an earlier Ipsos MORI rebuttal. If properly controlled then such a sample of 1000 respondents offers a 95% confidence interval of +/-3% on a 50% statistic. [This is not strictly the case here since these confidence limits assume that the sample is a purely random one: a quota sample is not.]

Given the universes set out above then samples of 1000 in each of the two local Districts, and 1000 across the rest of the County, are easy to defend. However, to minimise design effect and maximise the effective sample (while perhaps at the same time ignoring local sensitivities) the two Districts could be treated as one local area sample so avoiding post-survey weighting. That is, the sample could be 1120 in Allerdale and 880 in Copeland.

Closely allied to the issue of sample size is that of cost: detractors could well claim that 3000 interviews is simply an extravagance. A 2000 interview survey, with 1000 in the combined local areas of Allerdale and Copeland,(eg 560 and 440 respectively) and a 1000 across the rest of the county is an option. However, while this would certainly reduce costs it might restrict significantly the capacity of the data to sustain statistically meaningful analysis with respect to sub-groups of interest (ie the young, those intending to stay in the area).

Since it is accepted that the minimum size for a sub-cell of interest is 80-100 interviews, the sample composition of earlier waves may help is assessing the sample size for this exercise. In simple terms, let us consider two sample sizes, A, the proposed 3000, and B, a possible 2000,

and let us ignore all weighting and other influences on effective sample size.

The two 'target groups' for which there is information are set out below:

Target group	**Incidence**	**Local 1000**	**Sample A**	**Sample B**
Under 24s	12%	120	360	240
Families with children under 5	11%	110	330	220

That is, these sample sizes can accommodate sub-cell analysis; but they are not sustainable at the individual District level if a combined 1000 (Sample B) is used for both Copeland and Allerdale; that is, 11% of 440 interviews is only 48: with Sample A it could be 11% of 880, ie 97. The incidence of respondents expecting to be living in the area in say ten years time is not known.

Furthermore, with respect to the key questions on support, earlier waves' results show a sizeable 20% of respondents stating 'neither support not oppose': this could be grasped upon by detractors, rightly or wrongly, as a significant weakness in the use of a 'net support' indicator. You should examine the wave 3 data tables in detail to ascertain just who these neither/not respondents are in terms of demographics and area to ensure that their key discriminators are represented in the main sample in sufficient numbers to sustain a meaningful investigation into why they are undecided, apart from the 'obvious' one of not knowing enough about the issue - which in itself could be presented as a criticism of your activities to date.

2.4.2 The Fukushima effect

Reporting on the 'local' and 'rest' samples separately will minimise the design effect of trying to produce some overall County-wide response. However, in response to the possible concerns raised by Fukushima, you might see merit in considering (then discarding? but at least with reasons) the following sampling design which seeks to acknowledge the Fukushima effect clearly and simply.

There may well be a case for setting the main sample as Cumbria-wide, in which case Allerdale would constitute 18% of the sample, and Copeland 14%. To achieve 1000 interviews in the smaller area of Copeland would demand an overall Cumbria-wide sample of c7100: however, if the two

target areas are classed as one, with 32% of the survey universe, then the target 1000 is achieved with 3125 interviews. In such a sample there would be 560 respondents in Allerdale and 440 in Copeland, hence a local booster in these two areas to make them up to 1000 in each sees the overall sample rise to 4125.

While such an approach does not change in any way the statistical rigour of your proposed method, it does show that you are acknowledging public concern...if indeed it does exist.

2.5 **Fieldwork**...and proportionality

Random Digit Dialling telephone interviewing is by far the most appropriate fieldwork method for an area of widely dispersed populations. It could be argued, though, that the questionnaires used in earlier waves were too wordy and introduced perhaps too complex sets of information to be assimilated adequately over a 'phone call. Hence, if telephone is the chosen approach, then the questionnaire should be simplified.

If this exercise were being carried out with no consideration of cost then my recommended approach would be a pre-selected fixed random face-to-face in-home survey: this being so you might consider asking bidders to submit a fee estimate for doing the survey this way so that the significant cost advantages of carrying out the work by telephone are made apparent.

3. The Invitation to Tender

Whatever the desired specification turns out to be, you should consider the following:

a) should your ITT be prescriptive, or should it set out the problem and invite bidders to present the sampling and fieldwork options which they recommend within set budget limits?

b) the ITT should ask for rough price estimates for alternative fieldwork methods and sample sizes to allow you to defend the method finally adopted on the grounds of proportionality.

4. Recommended actions

The proposed method is good, but it might be prudent for you to –

- revisit and rehearse your arguments for having differential sampling fractions across Cumbria
- define your target groups of respondents, by demographics as well as area, to help assess the final sample size
- decide on how prescriptive the ITT should be, and whether it seeks rough price estimates for various methodological options.