

Comparing data from online and face-to-face surveys

Bobby Duffy and Kate Smith

MORI Online

George Terhanian and John Bremer

Harris Interactive

This paper explores some of the issues surrounding the use of internet-based methodologies, in particular the extent to which data from an online survey can be matched to data from a face-to-face survey. Some hypotheses about what causes differences in data from online panel surveys and nationally representative face-to-face surveys are discussed. These include: interviewer effect and social desirability bias in face-to-face methodologies; the mode effects of online and face-to-face survey methodologies, including how response scales are used; and differences in the profile of online panellists – both demographic and attitudinal.

Parallel surveys were conducted using online panel and face-to-face (CAPI) methodologies, and data were compared before weighting, following demographic weighting and following ‘propensity score weighting’ – a technique developed by Harris Interactive to correct for attitudinal differences typically found in online respondents. This paper looks at the differences in data from online and face-to-face surveys and puts forward some theories about why these differences might exist. The varying degrees of success of the weighting are also examined.

Introduction

Growth of online research

During the past few years there has been considerable growth in internet use. According to the Office for National Statistics, only 9% of the UK adult population had access to the internet at home in 1998. However, this had increased to 52% by 2004 and, in February 2005, 59% of adults in Great Britain had used the internet prior to interview (National Statistics Omnibus Survey 2005).

In the USA, internet access is slightly ahead at 69% but growth in online research has been considerable – estimates suggest that it now accounts for anything between 7% and 20% of quantitative research (Comley 2003). While estimates of the current size of the market in the USA vary, it is very clear that it is likely to grow – to between 33% and 50% by 2005/6, depending on which projection you look at (Terhanian 2003). Although online research in the UK is much less developed (it accounted for maybe 1–2% of research revenue in the UK in 2002), as in the USA this is also growing considerably (Comley 2003).

However, it is important to note that the UK will not necessarily follow the USA down exactly the same path – it hasn't with other research methodologies and there is no reason to assume it will start now. First, the sheer size of the USA has made telephone and postal approaches more the norm than in the UK, where a large proportion of research is still done face to face (it is still the largest single research method in revenue terms by some distance).¹ It is perhaps easier to make the transition to online research from telephone and postal, with their more obvious problems with sample frames and self-selection among respondents than seen in a face-to-face survey with a high response rate using a fairly comprehensive sample frame of addresses (such as the Postcode Address File).

There are also a number of important circumstantial and cultural factors that could explain the adoption of approaches at different rates. These include the much greater investment funding that was available in the USA at the time of the first-wave internet research, the nature of the people who led the initial boom (higher profile and perhaps more charismatic in the USA) and the greater acceptance of risk within the US research industry (Terhanian 2003).

In any case, whatever the variations in trajectories, online research is only going to become more rather than less commonplace across all industrialised countries in the foreseeable future. For this reason it is important to identify the conditions that led to representative research being produced through an internet methodology.

In comparing two parallel surveys – one online, the other face to face – this study aims to establish the extent to which data from an online panel survey can be matched to data from a nationally representative face-to-face (CAPI) survey. Before discussing findings, we outline briefly some of the pros and cons of online panel research.

¹ According to BMRA figures, face-to-face interviewing accounted for 27% of BMRA members' revenue and 1.6 million interviews in 2002, whereas telephone interviewing accounted for 22% revenue and 2.4 million interviews.

Advantages and disadvantages of online research

A number of studies have outlined the relative advantages and disadvantages of online research. It should be noted, however, that many of these studies fail to differentiate between the various types of internet research. Just as there are better and more representative face-to-face and telephone surveys, some internet-based methodologies are better than others. For the purpose of this paper, we will therefore focus briefly on the advantages and disadvantages of online surveys that are conducted among a panel of people who have been contacted for market research purposes.

The key advantages nearly always quoted first are greater speed and lower cost. In a number of circumstances these are going to be significant – particularly for multinational research and research with specialist audiences.² However, even proponents downplay the general cost advantages, at least at this stage in development, with the cost of building and maintaining a panel being quite substantial in the beginning. The cost savings come in the low variable cost per interview and are borne out over a period of time.

Instead the focus is on speed of response. Here again the advantages can be considerable as it is possible to accumulate very large volumes of interviews in a short space of time. Having said this, a minimum fieldwork period is often recommended for online surveys to ensure good coverage, so speed of response becomes no more of an advantage for online than for face-to-face surveys, given a sufficiently large field force.

Other advantages highlighted relate to how the use of new technology in online surveys allows research that is more visual, flexible and interactive (Taylor 2000). Again there is clearly much truth in this, particularly where the comparison is with telephone surveys. This is less so when we compare with face-to-face surveys, as the significant growth of CAPI means that similar multimedia options are available, while CASI sections in surveys allow very similar direct interaction with visual tasks. Clearly, however, online surveys allow companies to avoid the capital costs, as the respondent effectively provides their own interviewing machinery.

Another advantage suggested is that online surveys do not require interviewers to be present and so interviewer effects are avoided. This again is likely to be a significant advantage for certain types of study,

² An article in the *New York Times* by Nancy Beth Jackson outlines how the cost to Consumer Reports (a research organisation that gathers views on consumer durables and other products) of an online survey is half that of the postal equivalent – with further savings to come now that development work has been done.

particularly where social desirability effects are likely to be large. One prominent example given is the higher admission of undesirable behaviour in online surveys than in interviewer-administered surveys (Comley 2003). Within political polls, the anonymity afforded by internet-based approaches is particularly highlighted as a way around the problem of the 'spiral of silence' that is seen by some to be responsible for under-recording Conservative voting in UK political polling, particularly in the 1990s (Kellner 2003a). In addition, the elimination of an interviewer is advantageous in many other areas of health care and public policy research – for example, disease prevalence rates are much closer to the known rates when using properly designed internet studies than when done either via the telephone or face to face. (Taylor *et al.* (2005) outline several topics that appear to produce more representative results solely through the elimination of an interviewer effect.)

A further category of advantages stems from the growing disillusionment with other methods, with online research seen as a possible way around these. In particular, these points focus on the increasing individualism and selectiveness of potential respondents, as well as their use of new technology such as voicemail and caller ID to avoid telephone surveys. The argument made here is that online surveys get around this by fitting in with a respondent's life; they can fill them in at their convenience and can partially complete and return whenever they like. It is argued that this may help explain the more 'socially liberal' attitudes seen in many online surveys, as respondents on average tend to lead less home-based lives and so are less cautious (Kellner 2003b). Indeed the more bullish examples of this argument suggest that this feature could actually produce achieved samples that are more representative than traditional approaches, as online interviewing reaches 'busy people – often educated and well-off – who systematically repel or ignore cold callers but are willing to answer questions posted on their computer screen' (Kellner 2004). However, others argue that it is the fact that online respondents are more 'viewpoint oriented' (i.e. more likely to have active opinions) that accounts for the different attitudes seen in online surveys.

The disadvantages cited for internet-based methodologies focus mainly on sampling issues, which are dealt with in the next section. However, other issues are raised around mode effects, where, for example, it is known that online respondents use scales differently from respondents in other modes. There is conflicting research on this, some showing that online respondents are more likely to choose midpoints in scales and 'don't know' options in general, and other research, in contrast, suggesting that

online respondents tend to choose extreme responses on these scales. Either way, these types of effect will be due to complex competing effects of response styles and do not necessarily make responses from online surveys less accurate, but they can cause problems when we attempt to switch to an online survey approach in tracking work. It is possible to correct for this to an extent through modelling, but this is likely to be viewed as less straightforward for those commissioning.

Online survey sampling: online panels

Unlike face-to-face surveys, which can be sampled from reasonably comprehensive databases, online surveys are most often conducted among respondents from a panel who have agreed to be contacted for market research. No simple database of everyone who is online exists, and it looks unlikely to exist for the foreseeable future. Furthermore, even if there were such a list, prohibitions against 'spamming' online users would prevent it from being used as a sampling frame.

There are therefore three main issues relating to coverage bias or selection error that are raised with the sampling approach to online panel surveys: first, of course, they can reach only those who are online; second, they can reach only those who agree to become part of the panel; and, third, not all those who are invited respond (Terhanian 2003).

What makes online surveys different from other survey approaches, such as telephone in the USA and face to face in the UK, is that such a large proportion of the population are excluded before the survey begins, and that these are known to be different from those who are included. Although internet access in the UK is around six in ten of the adult population and rising, the demographic profile of internet users is not representative of the UK adult population as a whole, tending towards younger age groups. Those who choose to sign up for online panels may also have a younger, more male profile (Terhanian 2005).

However, online surveys are not as different as some would like to make out, given the practical constraints on other survey approaches. Large sections of the public effectively rule themselves out of all surveys before they start, and these people also have a different profile from those who do take part.

Further, online surveys can match demographic profiles through disproportionate sampling, in the same way that traditional quota surveys do, using information on likely response rates in order to decide on the number of leads from particular demographic groups that are

issued.³ In addition, data can be weighted to adjust for any further differences in demographic profile.

However, it isn't only differences in the demographic make-up of online survey samples that need to be addressed. Even when data from online surveys have been weighted to the desired demographic profile, *attitudinal or behavioural* differences are still observed. Certain kinds of people are more or less inclined to complete web surveys – even after controlling for demographic characteristics. This is not a new finding; several comparative studies have been conducted in the USA and in the UK, and they have all shown that, even following demographic weighting, there are still differences in data. As noted above, it has been observed that online data tend to paint a more active picture of the population: online survey respondents tend to be more politically active, more likely to be earlier adopters of technology, and tend to travel and eat out more than face-to-face survey respondents (Baker *et al.* 2003).

Harris Interactive, along with a number of other research organisations, has been working for some time on a weighting system that aims to adjust for attitudinal and behavioural differences. However, Harris Interactive has been the first to identify the benefits that an approach called 'propensity score weighting' can bring. The technique behind propensity score weighting – propensity score matching (Rosenbaum & Rubin 1984) – has been used since the early 1980s, most commonly in evaluations of social policy, to ensure that experiment and control groups have similar characteristics (where random assignment is not possible).

The propensity score matching process is as follows.

- Parallel online and telephone or face-to-face surveys are conducted where the same questions are asked at the same time using different modes (an online survey and a telephone or face-to-face survey).
- Logistic regression is then employed to develop a statistical model that estimates the probability that each respondent, conditional on his or her characteristics, participated in the telephone or face-to-face study rather than the online one. The probability, or 'estimated propensity score', is based on answers to several socio-demographic, behavioural, opinion and attitudinal questions.
- Next, in the 'propensity score adjustment' step, respondents are grouped by propensity score within the survey group (telephone/

³ As an aside, it is worth noting that those less likely to be online and to join a panel (i.e. the older and women) are more likely to respond when they are asked. This clearly has no influence on sample error, but it does mean that panels need to contain proportionally fewer of these difficult to recruit groups, which will keep costs down.

face-to-face or online) they represent. Statistical theory (Rosenbaum & Rubin 1984) shows us that when the propensity score groupings are developed methodically, the distribution of characteristics within each internet grouping will be asymptotically the same as the distribution of characteristics within each corresponding telephone or face-to-face grouping. Therefore, by weighting the internet sample's propensity group proportions to be the same as the telephone or face-to-face sample's propensity group proportions, the distribution of characteristics will be asymptotically the same across all propensity groupings within both samples, assuming there is sufficient overlap between the two distributions.

This procedure produces a result similar to randomisation: the estimated probability of belonging to one group rather than the other will be the same given the variables in the model. Because the model includes behavioural, attitudinal and socio-demographic information, we can be far more confident that we have linked the right people together than would have been the case had we relied on only basic socio-demographic variables. More importantly, there should be no differences in the survey responses of interest, as long as the propensity score model includes the right variables and the survey has been designed in a way that minimises design effects (e.g. scale presentation and wording differences).

Careful consideration and repeated testing is vital in the generation of all propensity score models; key characteristics of the sample must be accounted for in order to correct disparities across modes. The model used in the experiment reported in this paper includes measures of attitudes about privacy, security and risk, and measures of physical activities and product boycotting behaviour, among other things. These measures have been chosen specifically because they cover some broad areas and issues that differentiate online users and potential panellists from those who would not join an online survey panel.

This study

Objectives

As part of continuing development of research approaches at MORI and Harris Interactive, parallel surveys were run comparing an online panel survey (Harris Interactive) with a face-to-face CAPI omnibus survey (MORI). This is one of the first major UK studies comparing online and

face-to-face data as opposed to online and telephone research. The objective of the study is to establish whether data from an online panel survey can be successfully matched to data from a nationally representative face-to-face survey. Specifically, the study aims to make comparisons at a number of levels. First of all, we are looking at differences between raw online and face-to-face data, and whether these are the same type of differences seen to exist between raw online and telephone data. Second, we are looking at the relative impact of demographic and propensity score weighting on attitudinal/behavioural variables in the online survey, and whether they close any gaps between online and face-to-face data. Throughout this we will be trying to identify the relative impact of sample and mode effects.

It should be noted at this stage that we are not making any a priori judgment about which methodology produces responses closest to the true figure: this, and the competing effects that will cause the differences, are discussed throughout the paper.

Methodology

The MORI omnibus is an in-home survey, conducted using CAPI. The sample is stratified by region, with constituencies used as the primary sampling units, and quotas set on a number of demographic factors – sex, age, work status and tenure. The comparison would therefore seem to be a very relevant one, as this is the type of low-cost survey vehicle that online polling will mainly be competing against, particularly for straightforward advocacy and PR work. The fact that the MORI omnibus uses a quota-based sample rather than a random sample in some sense also helps level the playing field between the two approaches.

The Harris Interactive online panel was utilised for the internet-based component of this experiment. The sample pulled from the panel was stratified by age, gender and region.

Several questions were placed on both surveys, with the target questions covering voting intention, socio-political activism, knowledge of/attitudes towards cholesterol, views of immigration and access to technology. These questions were selected to provide a relatively stern test of how close an online survey can get to a face-to-face survey, given that there are likely to be significant mode effects (particularly interviewer effects) and a noticeable impact from any attitudinal bias in the online sample.

In addition, five 'propensity score' questions were asked on each survey. These cover issues such as online purchasing behaviour, views on the amount of information respondents receive and personal attitudes towards risk, social pressure and rules.

Question wordings on both surveys were kept as similar as possible, but some adaptations were required to reflect the different interviewing methods. Show cards were used in the face-to-face survey for all questions except for those with a simple 'yes/no' or numerical response, and the order of response scales and statements was rotated in both surveys.

The fieldwork periods for both surveys were kept as close as possible: the MORI omnibus ran from 31 July to 5 August 2003, and the Harris Interactive online survey ran from 31 July to 11 August 2003. The longer survey period for the Harris Interactive study was unrelated to the sample used in this analysis: an additional sample of youth was contacted by Harris Interactive for a separate series of questions within the survey and the survey was kept open to all respondents until the required completes for the youth sample had been achieved.

Once the surveys had been completed, both sets of data were weighted to the correct demographic profile (UK adults aged 15+). In the case of the omnibus survey this involved applying simple rim weights on region, social class, car ownership, and age and work status within gender. For the online survey the demographic weights that were applied were age within gender, ITV region, education level, income level and internet usage (ranging from high to low, measured in number of hours per week). In addition, the propensity score distribution was incorporated to ensure the same distribution of characteristics on attitudinal and behavioural characteristics across surveys.

When comparing unweighted and weighted data from face-to-face and online surveys, significance testing has been applied. Although, strictly speaking, this technique can be used only with probability samples, its use with non-probability samples, such as the quota samples used in these face-to-face and online surveys, is generally accepted practice in the market research industry. Significance testing of all weighted data is based on effective sample sizes (i.e. accounting for the 'design effect' of weighting). This is important, as the design effect of the online sample weighting is substantial – as will be seen in the following section.

Main findings

Voting intention: propensity score weighting works on voting

The first question area we looked at was voting intention, which has been a particular focus for many online survey experiments, due to the success of some companies in predicting election results using online methods.

Comparison of unweighted face-to-face and online data shows us what previous studies of online research methodologies have suggested: online respondents are more likely to say they would vote Liberal Democrat or Conservative than their face-to-face counterparts (Baker *et al.* 2003; Kellner 2004). This is likely to be because of two competing effects that we see throughout the study.

First, it has been hypothesised and shown to some degree that online panels tend to achieve samples that are more educated and active – and we know Lib Dem supporters in particular tend to be more politically active.⁴ The finding that fewer online respondents say they ‘would not vote’ also supports this argument. On the other hand, the fact that online respondents are more likely than face-to-face respondents to say they would vote Conservative is likely to be at least partly because there is no interviewer present, as discussed in the Introduction.

The application of demographic weighting to both sets of data does serve to close the gap between online and face-to-face results (reducing the proportion of votes for the Conservatives and Liberal Democrats in particular), but it is the *propensity score weights* that bring online data closer into line – to the extent that there is now no statistically significant difference between the two samples. While the face-to-face weighting has had very little effect on data (increasing Conservative and Liberal Democrat support by just one percentage point), the propensity score weighting has had a significant impact on online data (for example, increasing Labour support by eight percentage points) (see Table 1).

It should also be noted that the design effect of propensity score weighting has nearly halved the effective sample size of the fully weighted online data, whereas demographic weighting has very little effect on the face-to-face effective sample size. However, as the original online sample was very large, comparisons are still relatively robust (a difference greater than +/- three percentage points would be significant).

4. This is seen in aggregates of MORI data from the first half of 2003, where 14% of Liberal Democrat voters are classed as ‘socio-political activists’, compared with 12% of Conservatives and 9% of Labour voters. This will of course partly be a function of the relative popularity of the parties (i.e. the leading party is likely to have a higher proportion of those less interested in political issues).

Table 1

Q How would you vote if there were a general election tomorrow?

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
Conservative	19	20	25	23	21
Labour	27	27	18	18	26
Liberal Democrats	12	13	19	17	15
Scottish/Welsh Nationalists	3	3	3	2	2
Green Party	1	1	–	–	–
Democratic Party	0	0	–	–	–
UK Independence Party	*	*	–	–	–
Other	2	2	6	6	5
Would not vote	14	14	9	10	9
Undecided	21	20	21	24	23
Base	2057	2057	4131	4098	3956
<i>Effective sample size</i>	2057	1706	4131	2590	2100

Source:

MORI

Harris Interactive

* Less than 0.5% but greater than zero.

As outlined earlier, these data are based on interviews conducted in July/August 2003. Both MORI and Harris Interactive produced voting estimates in the run-up to the 2005 general election, and their final polls came out with very similar figures, after weighting: MORI data had 33% Conservative, 38% Labour, 23% Liberal Democrat and 6% other, Harris Interactive had 33% Conservative, 38% Labour, 22% Liberal Democrat and 7% other, compared with the final result of 33% Conservative, 36% Labour, 23% Liberal Democrat and 8% other.

However, these simple voting intention figures are not those typically quoted in published polls. These tend to be followed up with a question on likely party support for those who are undecided, and the combined figure is re-based on only those who say they will vote/support a party.

As shown in Table 2, the relationships remain similar – the fully weighted online figures come close to matching the face-to-face figures, and it is the propensity score weighting that is most responsible for this.

At MORI we also ask a ‘certainty to vote’ question, on a scale of 1 to 10, and increasingly count only those who say they are certain to vote in our estimate of likely voting behaviour (Table 3). The findings for this question are actually pretty similar between the two studies and different versions of the weighting. The weighting of the online results makes the more politically active online sample somewhat less likely to vote, as we would expect.

Table 2

Q How would you vote if there were a general election tomorrow (excluding undecided, would not vote, refused)?

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
Conservative	30	31	35	34	31
Labour	42	41	25	27	38
Liberal Democrats	19	19	27	26	22
Other	9	9	12	12	10
Base	1329	1329	2900	2867	2876
<i>Effective sample size</i>	<i>1329</i>	<i>1092</i>	<i>2900</i>	<i>1822</i>	<i>1450</i>
Source:	MORI		Harris Interactive		

Table 3

Q How likely would you be to vote in an immediate general election?

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
1 – Absolutely certain not to vote	12	12	7	8	7
2	2	2	2	3	3
3	3	3	4	5	4
4	2	2	3	4	4
5	8	8	8	9	10
6	4	4	4	4	4
7	5	6	7	7	7
8	7	7	9	9	9
9	6	6	10	8	9
10 – Absolutely certain to vote	48	47	47	43	43
Base	2000	2000	4131	4098	4098
<i>Effective sample size</i>	<i>2000</i>	<i>1656</i>	<i>4131</i>	<i>2590</i>	<i>2100</i>
Source:	MORI		Harris Interactive		

Table 4

Q How would you vote if there were a general election tomorrow (excluding undecided, would not vote, refused – and filtered by ‘certain to vote’)?

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
Conservative	35	35	39	38	34
Labour	38	38	26	26	35
Liberal Democrats	19	20	24	25	22
Other	8	8	12	11	9
Base	817	817	1848	1714	1714
<i>Effective sample size</i>	<i>817</i>	<i>682</i>	<i>1848</i>	<i>1078</i>	<i>866</i>
Source:	MORI		Harris Interactive		

The voting intention figures for just those certain to vote are shown in Table 4. This actually has a similar type of impact on each of the surveys, increasing the Conservative vote and slightly reducing the Labour share. However, the final fully weighted online figures are somewhat out of line with other published polls around that time.⁵

Socio-political activism: online respondents are more active

Comparing face-to-face and unweighted online data for questions on socio-political activism confirms our expectations (Table 5); online respondents are more likely to have presented their views to a local councillor or MP, or to have urged someone else to do so, than face-to-face respondents. However, whereas they are more likely to have urged someone else to vote, online respondents are as likely as face-to-face respondents to say they have voted in the last general election themselves, or to have made a speech before an organised group. They are also *less* likely to say they have helped on a fundraising drive.

Again it is possible to explain these differences by the fact that the online results come from a more politically active sample facing lower social

⁵ An ICM poll for the *Guardian* with fieldwork from 14–17 August 2003 put Labour on 37%, Conservatives on 32% and the Liberal Democrats on 22%.

Table 5

Q Which, if any, of the things on this list have you done in the last two or three years?

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
Presented my views to a local councillor or MP	16	15	23	20	20
Written a letter to an editor	7	7	15	12	11
Urged someone outside my family to vote	15	15	27	24	24
Urged someone to get in touch with a local councillor or MP	14	14	27	25	25
Made a speech before an organised group	14	15	13	11	10
Been an officer of an organisation or club	9	9	15	12	12
Stood for public office	1	1	1	1	1
Taken an active part in a political campaign	3	2	5	4	4
Helped on fundraising drives	22	23	13	13	12
Voted in the last general election*	61	65	63	59	60
None of these	27	27	23	27	26
Base	2057	2057	4131	4098	3956
<i>Effective sample size</i>	<i>2057</i>	<i>1706</i>	<i>4131</i>	<i>2590</i>	<i>2100</i>
Source:	MORI		Harris Interactive		

* Filtered to include only those aged 20+ (i.e. the group that would be eligible to vote at the time of the last election).

desirability pressure, because there is no interviewer present. For example, we know that 61% of the UK population actually voted in the 2005 general election, which is very similar to the 60% seen in the weighted online data, but somewhat lower than the 65% of the weighted face-to-face sample. This does suggest some over-claiming among face-to-face respondents with respect to some activities, particularly given that we know they are generally less politically active than the online sample.

Significant differences between the two samples remain after the weighting has been applied – suggesting that this is one of those factors where online and face-to-face surveys can be less successfully matched. It is also interesting that it is the demographic weighting that has brought the online data closer to the face-to-face data – the propensity score weighting has in fact had a minimal additional effect, in contrast to voting intention. This result is consistent with the work of Terhanian and Bremer, who show

that demographic biases comprise the largest component of the total bias of online studies when it comes to factual responses. Intent, attitudinal and forward-looking behavioural metrics are those that are most affected by propensity score matching techniques (Terhanian & Bremer 2001).

Attitudes towards immigration: more or less considered views online?

Attitudes towards immigration have been surveyed by MORI a number of times, and findings have varied greatly by education, social class and general world-view. Further, these questions cover sensitive issues and are likely to be susceptible to eliciting socially desirable responses, particularly when an interviewer is present. These questions were therefore interesting to repeat in the online vs face-to-face experiment, as large differences could be expected.

Weighting does not have much effect on either online or face-to-face survey data, and the key finding from these questions is that online survey respondents seem much more inclined to select the neutral point ('neither agree nor disagree') than face-to-face respondents. This is consistent with a number of other studies on scale usage across modes, and a number of different possible explanations exist.

First, the questions were placed towards the end of both surveys, and there are good reasons to think that respondent fatigue has more noticeable consequences for online surveys; there is no interviewer to encourage respondents to answer questions, and respondents may be more likely to click down or give answers less thought. (This is consistent with theories, such as 'satisficing', where people put in the minimum of required effort; see Krosnick 2000.) Having said this, previous studies by Harris Interactive of respondent fatigue in online surveys have shown that the first drop-off in response levels tends to take place after 18 minutes, depending on the survey population and subject – and the questions were placed well before that point.

So a second possible explanation is that these are complex questions with no straightforward answers, and a 'neither' response is logical. This explanation is given more weight by the fact that the effect is greatest in the question that is probably most susceptible to socially desirable responses (the statement that 'it is a good thing that Britain is a multicultural society'). It could therefore be argued that the face-to-face results artificially emphasise opinions, when actually there are few strongly held views on these sensitive, complex issues (see Tables 6–9).

Table 6

Q It is a good thing that Britain is a multicultural society?

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
Strongly agree	19	20	20	18	18
Tend to agree	44	44	30	30	29
Neither agree/disagree	16	15	26	28	28
Tend to disagree	13	13	16	16	16
Strongly disagree	7	7	9	9	9
Agree	63	64	50	48	47
Disagree	20	20	25	25	25
Net agree	43	44	25	23	22
Base	2035	2035	4131	4098	3956
<i>Effective sample size</i>	<i>2035</i>	<i>1688</i>	<i>4131</i>	<i>2590</i>	<i>2100</i>
Source:	MORI		Harris Interactive		

Table 7

Q The government has immigration under control.

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
Strongly agree	1	1	2	2	2
Tend to agree	6	7	4	4	4
Neither agree/disagree	7	8	13	12	12
Tend to disagree	26	27	28	28	29
Strongly disagree	60	58	54	54	53
Agree	7	8	6	6	6
Disagree	86	85	81	82	82
Net agree	-79	-77	-75	-76	-76
Base	1998	1998	4131	4098	3956
<i>Effective sample size</i>	<i>1998</i>	<i>1658</i>	<i>4131</i>	<i>2590</i>	<i>2100</i>
Source:	MORI		Harris Interactive		

Table 8

Q I am concerned that Britain is losing its own culture.

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
Strongly agree	32	30	26	27	27
Tend to agree	29	29	27	26	26
Neither agree/disagree	12	13	18	18	18
Tend to disagree	19	20	16	15	16
Strongly disagree	8	8	14	14	14
Agree	61	59	54	54	53
Disagree	27	28	30	29	30
Net agree	34	31	23	25	23
Base	2017	2017	4131	4098	3956
<i>Effective sample size</i>	<i>2017</i>	<i>1672</i>	<i>4131</i>	<i>2590</i>	<i>2100</i>
Source:	MORI		Harris Interactive		

Table 9

Q I am unwilling to say what I really think about immigration in case I am seen as racist.

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
Strongly agree	10	9	10	10	10
Tend to agree	20	20	18	18	18
Neither agree/disagree	12	12	18	19	19
Tend to disagree	29	29	23	23	23
Strongly disagree	28	30	31	30	29
Agree	30	29	28	28	28
Disagree	57	59	54	53	52
Net agree	-27	-30	-26	-25	-24
Base	2033	2033	4131	4098	3956
<i>Effective sample size</i>	<i>2033</i>	<i>1690</i>	<i>4131</i>	<i>2590</i>	<i>2100</i>
Source:	MORI		Harris Interactive		

Cholesterol: online respondents are better informed

The results on understanding of issues surrounding cholesterol (Tables 10–15) appear to confirm that online respondents are generally better informed than face-to-face samples, with a significantly higher number of online respondents correctly saying that cholesterol is ‘a type of fat that circulates in the bloodstream’. The rating of the seriousness of cholesterol as a health risk clearly illustrates the pattern seen in other studies, where online respondents are less likely to choose extreme options.

In the true/false questions (where the correct answer in both cases is false) we see that the online respondents are less likely to give an incorrect answer, at least to one statement. For the other statement (‘Men are more likely than women to have a high level of cholesterol’), online respondents are more likely to give a ‘Don’t know’ response and less likely to answer incorrectly. This again is likely to be a combination of effects, with the presence of an interviewer discouraging ‘Don’t know’ responses in the face-to-face survey and/or a satisficing effect, alongside the greater knowledge of online respondents resulting in fewer incorrect answers.

This raises interesting questions about asking knowledge-based questions in online studies. It is possible for online respondents to search for answers to factual questions on the internet – while this option is clearly not available to those who are interviewed face to face. Indeed this has been noted by an American journalist who signed up to a number of online panels in order to write an article on online research: ‘Occasionally I felt so ill informed about a topic that I cribbed an answer, pausing to do a Google search, for example, to help me on a couple of sports surveys even though nothing but my ego was riding on an informed response’ (*New York Times* 2004). While online surveys may do away with interviewer bias, they may bring other forms of inaccuracy that we need to be aware of when deciding which surveys are appropriate.

To try to give some indication of whether online respondents had cribbed their answers from the internet in this instance, data from the face-to-face survey were examined more closely. Answers given by face-to-face respondents with internet access (either at home or at work) were compared with responses given by respondents without internet access. Interestingly, face-to-face respondents with internet access were more likely to answer most questions on cholesterol correctly (or less likely to give an incorrect answer) than those without. This seems to suggest that online respondents are more likely to answer knowledge-based questions correctly because, as internet users, they tend to be more knowledgeable – and not because they have looked up the answers.

Table 10

Q Which one of these descriptions, if any, most closely defines what you think cholesterol is?

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
A part of the skin	1	1	*	*	*
A medicine	1	1	*	*	*
A type of fat that circulates in the bloodstream	81	82	92	90	90
A stomach disorder	1	1	*	*	*
Heart disease	6	6	3	4	4
Being overweight	6	6	2	2	3
A lung disease	*	*	0	0	0
Other	*	*	*	1	*
None of these	1	1	1	1	1
Not sure	4	4	1	2	2
Base	2057	2057	4131	4103	3956
<i>Effective sample size</i>	<i>2057</i>	<i>1706</i>	<i>4131</i>	<i>2590</i>	<i>2100</i>
Source:	MORI		Harris Interactive		

*Less than 0.5% but greater than zero.

Table 11

Q How serious a health risk, if at all, do you think it is to have a high level of cholesterol?

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
Not at all serious	*	1	1	1	1
Not very serious	3	3	4	4	4
Fairly serious	43	44	53	51	52
Extremely serious	54	52	43	45	44
Base	1980	1980	4131	4098	4098
<i>Effective sample size</i>	<i>1980</i>	<i>1641</i>	<i>4131</i>	<i>2590</i>	<i>2100</i>
Source:	MORI		Harris Interactive		

*Less than 0.5% but greater than zero.

Table 12

Q Please indicate whether you think the following statement is true or false:

If you are not overweight, it is unlikely that you will ever have to worry about cholesterol.

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
True	13	14	7	7	8
False	80	80	84	83	83
Don't know	6	6	9	10	10
Base	2057	2057	4131	4098	3956
<i>Effective sample size</i>	<i>2057</i>	<i>1706</i>	<i>4131</i>	<i>2590</i>	<i>2100</i>
Source:	MORI		Harris Interactive		

Table 13

Q Please indicate whether you think the following statement is true or false:

Men are more likely than women to have a high level of cholesterol.

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
True	30	32	29	29	28
False	52	50	40	40	40
Don't know	18	18	32	31	32
Base	2057	2057	4131	4098	3956
<i>Effective sample size</i>	<i>2057</i>	<i>1706</i>	<i>4131</i>	<i>2590</i>	<i>2100</i>
Source:	MORI		Harris Interactive		

Table 14

Q Please indicate whether you think the following statement is true or false:

If you are not overweight, it is unlikely that you will ever have to worry about cholesterol.

	F2F survey		Online	
	Internet access at home or work		No internet access	
	%		%	
True	9		19	
False	89		72	
Don't know	2		10	
Base	1012		1045	

Source: MORI

Table 15

Q Please indicate whether you think the following statement is true or false:

Men are more likely than women to have a high level of cholesterol.

	F2F survey		Online	
	Internet access		No internet access	
	%		%	
True	34		31	
False	52		48	
Don't know	14		21	
Base	1012		1045	

Source: MORI

In this section of the interview we also asked about actions taken to maintain a healthy heart (Table 16) – and there are some similarities and differences in response, with no clear pattern or explanation of why this should be the case. Face-to-face respondents are more likely to say they exercise often, but less likely to say they have given up smoking, and slightly less likely to say they eat low-fat foods.

Overall, as seen with previous questions, weighting, whether it be demographic weighting or propensity score weighting, does very little (if anything) to reduce differences between online and face-to-face data.

Table 16

Q Which of these, if any, are you currently doing to try to maintain a healthy heart?

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
Taking tablets/medication	15	13	16	15	15
Exercising often	50	51	46	45	44
Eating low-fat foods	46	45	49	49	49
Giving up/have given up smoking	13	13	23	21	21
Eating a healthy diet	65	65	65	61	60
Eating a cholesterol-lowering spread such as Benecol or Flora Pro-active	16	16	17	17	17
Other	1	1	5	5	4
Nothing	13	13	13	14	14
Base	2057	2057	4131	4098	3956
<i>Effective sample size</i>	<i>2057</i>	<i>1706</i>	<i>4131</i>	<i>2590</i>	<i>2100</i>
<i>Source:</i>	<i>MORI</i>		<i>Harris Interactive</i>		

Technology usage: do not ask online

As recognised in just about all methodological literature on internet research, there are certain survey questions that it will never be appropriate to ask online when you are trying to represent the population as a whole – and technology use is certainly one of those. The entire sample source by the very nature of the survey approach has internet access of some sort, and internet access correlates very highly with many types of technology usage, particularly computer ownership.

As the results in Table 17 show, the differences are indeed significant, particularly on computer/internet-related technologies, but also less obvious ones such as mobile phones and DVD players. Further, weighting again has very little effect. It is important to note that while incidence studies of technological adoption in the general population are ill advised online, studies of a particular population that owns a particular technology – for example, internet users, digital camera users or mobile phone users – can be done very successfully using an online panel.

Table 17*Q Which of these, if any, do you personally use?*

	F2F survey		Online		
	Unweighted	Demographic weighted	Unweighted	Demographic weighted only	Demographic and propensity weighted
	%	%	%	%	%
Mobile phone	74	76	87	87	86
Text messaging (SMS) on a mobile phone	47	50	69	69	69
Picture messaging (MMS) on a mobile phone	5	7	9	10	10
PC – at home	50	53	92	91	91
PC – at work, place of study or elsewhere	27	30	48	43	43
Internet at home	38	41	90	88	89
Internet at work, place of study or elsewhere	25	28	44	41	40
Internet at home via broadband connection	10	11	40	37	38
Digital TV	40	42	63	62	62
Interactive services on digital TV	13	15	25	24	25
DVD player	44	47	70	68	69
Digital radio	13	13	21	19	19
None of these	13	13	*	*	*
Don't know	14	13	*	*	*
Base	2057	2057	4131	4098	3956
<i>Effective sample size</i>	<i>2057</i>	<i>1706</i>	<i>4131</i>	<i>2590</i>	<i>2100</i>

Source:

MORI

Harris Interactive

*Less than 0.5% but greater than zero.

Conclusions

We have seen a range of outcomes when comparing online and face-to-face questions: some are close without weighting (e.g. attitudes towards immigration), some are very close after weighting (e.g. voting intention), some very different and not helped by weighting (e.g. political activism and knowledge-based cholesterol questions). We have put forward some theories as to why data from online and face-to-face surveys might be different; however, we need to understand more about why weighting by both demographics and attitudes has varying degrees of success.

There seem to be two main competing effects at play when comparing online and face-to-face methodologies. Online research using panel

approaches appears to attract a more knowledgeable, viewpoint-orientated sample than face-to-face surveys. This could be because this is a prior characteristic of those with access to the internet or those who join online panels, or it could be a learned behaviour from taking part in a number of surveys. However, face-to-face respondents are more susceptible to social desirability bias due to the presence of an interviewer. Sometimes these effects appear to balance, bringing the outcomes from the two methodologies together, but sometimes they don't.

Voting intention is an example of a question area that has been successfully matched online, suggesting that, for some areas of study, well-designed internet-based surveys with appropriate weighting strategies can produce similar results to well-designed face-to-face surveys. However, a number of other question areas are not so encouraging, particularly where the issues are sensitive. The higher number of 'Don't know' and 'Neither/not sure' responses in online surveys needs to be explored further to assess the extent to which they are satisficing behaviour or a true reflection of views when there are no interviewer effects. It has also been observed that questions that test knowledge may pose particular issues for online research, possibly because online respondents tend to be more informed generally rather than them using the internet as a source of information during interviews.

A further note of caution when applying relatively heavy weighting to data sets, such as the propensity score weights used in this study, relates to design effect and impact on effective sample size. If such weights are to be used, the sample must be large enough to ensure that the resulting effective sample size will stand up to significance testing. Of course, as the cost per interview is low when a large number of online interviews are conducted, this may not be a problem.

Despite these limitations it seems likely that online surveys will grow substantially over the next few years. This is partly because there are some doubts over either the capacity for or methodological advantages of traditional methods. First, face-to-face interviewing resources are limited and increasingly expensive. Landline telephone penetration is dropping, with currently 7% of households having no phone or mobile only. This is likely to grow fairly significantly and, more importantly, there is significant bias involved, with young households in particular much more likely to have mobiles alone. It is therefore important to continue to think about in what circumstances and how internet-based methodologies can be used for data collection, and to develop approaches that will be as robust and representative as possible.

References

- Baker, K., Curtice, J. & Sparrow, N. (2003) *Internet Poll Trial: Research Report*.
- Comley, P. (2003) Innovation in online research – who needs online panels?, MRS Research Conference Paper 36.
- Jackson, N.B. (2003) ‘Opinions to spare? Click here’. *New York Times*, 3 July.
- Kellner, P. (2003a) First among equals. *Research*, September.
- Kellner, P. (2003b) Personal communication to Nick Sparrow.
- Kellner, P. (2004) Can online polls produce accurate findings? *International Journal of Market Research*, **46**, 1.
- Krosnick, J. (2000) The threat of satisficing in surveys: the shortcuts respondents take in answering questions. *Survey Methods Newsletter*, **20**, 1.
- National Statistics Omnibus Survey (2005) Individuals accessing the internet, National Statistics Omnibus Survey, April.
- Rosenbaum, P.R. & Rubin, D.B. (1984) Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, **79**, 387, pp. 516–524.
- Taylor, H. (2000) Does internet research work? *International Journal of Market Research*, **42**, 1, pp. 51–63.
- Taylor, H., Krane, D. & Thomas, R.K. (2005) How does social desirability affect responses? Differences in telephone and online. *Public Opinion Pros*, February.
- Terhanian, G. (2003) The unfulfilled promise of internet research, MRS Conference Paper 37.
- Terhanian, G. (2005) Personal communication, 20 June.
- Terhanian, G. & Bremer, J. (2001) Reducing error associated with non-probability sampling with propensity scores, Paper presented at the 2001 Joint Statistical Meetings.