



Moving an established survey online - or not?

Tim Barber, Dave Chilvers and Sumran Kaul

International Journal of Market Research

Vol. 55, No. 2, 2013



Title: **Moving an established survey online - or not?**
Author(s): **Tim Barber, Dave Chilvers and Sumran Kaul**
Source: **International Journal of Market Research**
Issue: **Vol. 55, No. 2, 2013**

Moving an established survey online – or not?

[Tim Barber](#) and [Dave Chilvers](#)

BDRC Continental

[Sumran Kaul](#)

Ofcom

Introduction

Ofcom is the regulator for the UK communications industries, with responsibilities across television, radio, telecommunications, wireless communications and postal services. As part of its remit, Ofcom collects and publishes a wide range of data about developments in its markets, to ensure decisions are informed by a substantial evidence base.

Each year, Ofcom publishes its Communications Market Report, a reference document containing statistics and analysis of the UK communications sector for industry, stakeholders and consumers. The report also provides context to the work that Ofcom undertakes in furthering the interests of consumers and citizens in the markets it regulates. The report provides data and analysis on broadcast television and radio, the internet, fixed and mobile telephony and postal services. It also offers insights into how people are using mobile technology to access audio-visual and online content.

Many of the research sources used by Ofcom are peer reviewed and therefore it is imperative that the data used for the Communications Market Report is robust and capable of withstanding external scrutiny. The Media Tracker is one such data source, which commenced in its current form in 2007, but some of the questions monitored have remained the same since 1991 when the Independent Television Commission ran the survey before the formation of Ofcom. Fieldwork is undertaken using an in-home face-to-face methodology, employing a long (30–40-minute) questionnaire.

Ofcom is a public sector body, so must strive to achieve value for money. For this reason it is important to consider transitions from traditional data collection approaches to potentially more cost-effective methods.

Against this background, it was decided to test the possibility of migrating the Media Tracker from a face-to-face approach to an online methodology. If successful, this could help Ofcom continue to provide data for the Communications Market Report, while reducing the cost of the survey.

Previous work, however, has suggested that migrating surveys where the subject relates to technology is not straightforward

Downloaded from warc.com

and requires weighting in addition to simple demographic weighting. Williams (2012) found, when comparing results from parallel offline and online surveys, that even quite extensive demographic weighting will not eradicate differences between the online panel and probability sample estimates, particularly for questions relating to technology usage. His work showed that adding a number of subject-specific dimensions to the weighting process can reduce the difference between estimates for the two difference sources, although these do not appear to eradicate bias across the board. He observed that measuring change is more robust than measuring absolute levels, so if a mechanism can be found for correcting the absolute figures on one occasion, this is likely to be applicable to subsequent occasions. It was noted that, in some cases, the online approach may produce a more honest response (in other words, not all observed bias may be due to the new methodology).

Propensity weighting has been used in a variety of surveys since 2000 in an attempt to make data collected through online panels more representative of a target population – for instance, nationally representative or representative of all internet users. It is a form of weighting that attempts to compensate for the likelihood of respondents to participate in the research. The process of creating a propensity score adjustment uses specific attitudes or characteristics that are asked about in the online survey, as well as a probability-based reference survey of the target population. Propensity weighting then generates a propensity score adjustment and assigns each member of the online survey a weight. It generally does this through logistic regression on the combined reference survey and panel survey. In this way it reduces selection bias in an attempt to make the full range of survey measures representative.

In the AAPOR Report on Online Panels (March 2010), the authors highlight many of the issues of propensity weighting, while concluding 'Researchers should avoid nonprobability online panels when one of the research objectives is to accurately estimate population values', and specifically stating: 'The effectiveness of [propensity weighting] is yet to be demonstrated consistently and on a broad scale.' This reinforces why Ofcom must validate any move to an online methodology.

Duffy *et al.* (2005) undertook some early experiments and saw 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). They put forward some theories as to why data from online and face-to-face surveys might be different, and highlighted the particular issue of surveys about technology (not just the internet but other technology subjects).

They also raised a further note of caution when applying relatively heavy weighting to data sets, such as propensity score weights based on attitudinal questions. 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 comparatively low when a large number of online interviews is conducted, this may not be a problem.

Cooke *et al.* (2009) reported an extensive programme to migrate parts of the Financial Research Survey to an online methodology. Their experience reinforced the view that online surveys tend to be completed by heavier internet users, and that these exhibit different behaviours from lighter internet users (who in turn are different from non-users). In addition, it was found that, for some measures, data from online surveys correlate better with external sources than data from face-to-face surveys, and for some subjects online surveys can generate better-quality data.

Given this background, we might expect that questions relating to technology use in general (not just internet usage) will generate different responses from an online survey when compared to a traditional face-to-face approach. Given this, the main thrust of our approach is to see whether we can develop a mechanism to correct for these differences.

The experiment

In late 2011, it was decided to undertake a parallel run of the Media Tracker Survey, running the existing face-to-face methodology at the same time as running an equivalent online survey. The online survey comprised respondents from two sources: an established online panel and river sampling using around 100 diverse websites to invite participation.

River sampling was used for two reasons:

1. to try to mitigate the effect of the bias caused by the self-selection nature of online panels
2. to attempt to include lower-frequency internet users.

Several commentators have highlighted the problems caused by online generation of a non-probability sample, and the likelihood that this would produce bias. The AAPOR Report on Online Panels (2010) concluded that 'Researchers should avoid nonprobability online panels when one of the research objectives is to accurately estimate population values. There currently is no generally accepted theoretical basis from which to claim that survey results using samples from nonprobability online panels are projectable to the general population. Thus, claims of representativeness should be avoided when using these sample sources.' The report further noted that, 'The majority of studies comparing results from surveys using nonprobability online panels with those using probability-based methods (most often RDD telephone) often report significantly different results on a wide array of behaviours and attitudes.' Given these conclusions and the UK studies referenced earlier, we might expect to see major differences between the two data collection methodologies (and possibly between data from river sampling and that from the online panel).

As mentioned earlier, both online and face-to-face samples had quotas reflecting the profile of UK adults by gender, age, social class, working status, household size, presence of children in the household, region/nation (with smaller nations being over-sampled to provide adequate base sizes) and TV reception type. A total of 890 face-to-face interviews were undertaken and 728 online (with 232 from an online panel and 496 from river sampling, as described). At an overall level, these sample sizes were judged as adequate for testing for significant differences at the overall UK level

Data from each source were then rim weighted to these UK national profiles, using a range of demographics (gender, age, social class, region/nation).

Before looking at the weighted results, we provide some results on the unweighted profile of each sample source. As can be seen in Table 1, there are some variations by source, but nothing which indicates that the demographic weighting to be applied will lead to a substantial drop in weighting efficiency (quotas were set for face-to-face and online overall to ensure consistency of the sample source). What is clear is that, even with demographic quotas, the frequency of using the internet differs substantially for the online survey – even if non-users are excluded from the face-to-face survey.

Table 1 Sample profiles (unweighted)

	Face-to-face	Online overall	River	Panel
Sample size	890	728	496	232
<i>Gender (%)</i>				
Male	50	48	51	41
Female	50	52	49	59
<i>Age (%)</i>				
16–24	16	16	17	13
25–34	17	19	21	16
35–44	21	20	22	16
45–54	14	15	15	15
55–64	13	13	9	23
65+	19	16	17	17
<i>Frequency of using the internet (%)</i>				
Every day	54	91	92	88
Several times a week	18	6	4	10
At least once a month	4	2	2	2
Several times a year	2	1	1	0
Never	22	0	0	0

Comparison of key results

As expected, even when weighted by conventional demographics, substantial differences existed between the face-to-face and online approaches. Table 2 confirms previous conclusions that online respondents are skewed towards heavy internet use.

Table 2 Frequency of internet use (%)

Frequency of using the internet	Face-to-face		Online overall	River	Panel
	Face-to-face	excluding non-users			
Every day	53	68	92	93	89
Several times a week	18	24	5	4	9
At least once a month	4	5	3	3	2
Several times a year	2	3	1	1	0
Never	23	29	0	0	0

Q90: How often do you use the internet on any device (at home/elsewhere)?

Among the face-to-face sample, 68% of online users have a frequency of use of every day, compared to 92% for the online sample. Of some interest is the result that those recruited from the online panel are slightly less frequent users than those recruited via river sampling – the latter clearly is not a suitable source for recruiting light internet users.

Given this bias towards heavy internet use in the online sample, it is not surprising that the results from questions relating to technology differ between the two sample sources, as Table 3 shows.

Table 3 Frequency of watching recorded/catch-up TV (%)

Frequency of watching recorded/catch-up TV	Face-to-face				
	Face-to-face	excluding non-users	Online	River	Panel
Every day	14	15	29	30	29
Several times a week	25	29	37	41	29
At least once a month	10	12	15	15	16
Several times a year	4	4	6	4	11
Never	46	39	12	9	15

Q21: On average, how often, if at all, do you watch television that is not live scheduled TV?

Clearly, demographic weighting on its own is insufficient to remove bias caused by the different methods of data collection, and there is a need to identify additional weighting variables if this bias is to be reduced or removed. For this question, those recruited via river sampling have greater frequency of watching recorded/catch-up TV (more viewing 'several times per week' and less 'several times per year/never'). However, both sources have substantially fewer claiming to never watch recorded/catch-up TV than the face-to-face sample.

A mechanism to identify covariates to be used in additional weighting

Previous work in this area has attempted to use other variables to reduce bias – referred to as 'hooks' or, more technically, covariates. Williams (2012), for example, predetermined these covariates for two different public-sector surveys. For the British Crime Survey, the selected covariate was an attitudinal statement: 'Taking everything into account, I have confidence in the police in my area.' For the Scottish Household Survey, the selected covariate was 'satisfaction with public transport'.

In previous work we have undertaken, we have used an attitudinal statement – 'I am interested in new technology' – as an attempt to correct the bias introduced by online samples featuring a disproportionately high number of heavy internet users.

All these examples identify subject-specific covariates, which intuitively seem to be more likely to adjust any biases relating to the specific subject. However, all these attempts suffer from the fact that the choice of covariate has been subjective.

We have therefore looked at using a non-subjective method to identify the most appropriate covariates for this particular survey. To work well, a covariate needs to be a question that has been answered by the whole sample (so that, for example, covariates based on questions just asked of internet users in the face-to-face survey cannot easily be extrapolated to the non-internet users). A covariate also needs to show different results from the two surveys (a question that gives the same results will not provide any substantial correction when used as an additional weight).

To facilitate this, we have used a discriminant analysis, where the respondents from the face-to-face survey are placed in one group and those from the online survey in another group. The discriminant analysis then looks at all questions covered and identifies those that discriminate the most between the two groups – the face-to-face sample and the online sample. It is possible to use discriminant analysis in a stepwise mode, so that increasing numbers of questions can be included to improve the discrimination.

The more variables that are used, the better the discrimination – and so the more the results between the two surveys are likely to converge if all these variables are used as additional weights. However, the greater the number of variables, the worse the impact on weighting efficiency – the more variables that are used as additional weights to reduce bias, the worse the weighting efficiency will become.

There is thus a trade-off between making the results from the online survey match those from the face-to-face survey compared to reducing the weighting efficiency and thus reducing the robustness of the reweighted results. There is no 'right' answer to this trade-off; it requires a judgement to be made as to whether the inclusion of an additional variable is worth the reduction in weighting efficiency the inclusion generates. There also needs to be consideration of how many variables in total should be used, as a very large number will suggest that the data have been manipulated to achieve a required result. Each project needs to be looked at in its own right to determine the recommended number of additional weighting variables to be used.

The results: first iteration

Initially, all questions in the survey were included in the discriminant analysis, including frequency of internet use; this variable can only be included by combining the non-users in the face-to-face survey with the least frequent users. As 92% of the online panel are daily users, the only way of redefining the frequency question is between daily users and less frequent/non-users. Table 4 shows these data.

Table 4 Frequency of internet use (grouped)

Frequency of using the internet	Face-to-face (%)	Online (%)	Weighting factor
Every day	53	92	0.576
Less frequently/never	47	8	5.875

Q90: How often do you use the internet on any device (at home/elsewhere)?

Less frequent users in the online sample are given a weight ten times the size of frequent users; this is bound to affect the weighting efficiency of the online sample reweighted under this regime. In fact, the weighting efficiency applying this weighting as well as the original demographic weights is reduced to 10%. This is clearly indefensible as a methodology for a publicly available dataset, stretching the data so much as to reduce the effective sample size from 728 to around 70.

The results: second iteration

A second discriminant analysis was therefore undertaken excluding the 'frequency of internet usage' question. The objective was to find a number of questions that effectively act as proxies for this initial question, but without having such a dramatic effect on weighting efficiency. Table 5 shows the results of this second iteration.

Table 5 Most important discriminators

Question	F value
Q21: Frequency of viewing television that is not live scheduled TV	199
Q88: Awareness of catch-up TV services on PC?	60
Q27: Activities undertaken whilst watching TV	52
Q82: As far as you are aware is the internet regulated?	37
Q4: Working status	10

The *F* value is a measure of the discriminatory power of the question. Typically a value exceeding 4 is significant at the 95% level. There are further questions that are significant discriminators but reweighting the online data using these five questions in addition to the original demographics brings most key questions into line. Table 6 shows the comparison for the main method of TV reception in the household.

Table 6 Main method of TV reception (%)

Main method of TV reception	Online original	Face-to-face	Online with five rims
Freeview	23	33	30
Satellite	45	41	42
Cable	20	13	15
Other	17	13	13

Q11/12: Which if any of these types of television does your household receive?
If more than one, which would you regard as your main type of television?

The original online data, when weighted just by demographics, shows 23% of households with Freeview as the main method of TV reception, compared to 33% for the face-to-face survey. Compensating this, the original online survey shows 20% of households with cable TV, compared to only 13% for the face-to-face survey. Weighting additionally by the five questions shown in Table 5 reduces the difference so that online data weighted this way now compare well to the original face-to-face survey.

Previous work confirmed the existence of modal bias but raised the question of which methodology was biased: the original face-to-face, the online or indeed both. Cooke *et al.* (2009) concluded that, for FRS, the online approach gave results more consistent with market data for heavy internet users and there was reason to believe that the original face-to-face survey generated more bias than its online equivalent. This helped them make the decision to migrate the FRS to online for heavy internet users aged 16–64 (while retaining the face-to-face approach for light/non-users and older respondents).

The question thus arises for the Media Tracker as to which data collection method generates bias (as there clearly is bias somewhere given the large differences in results from the two methods). Luckily, a number of the questions in the Media Tracker concern measuring market statistics, for which universe data are available. For example, Table 7 shows the market data compared to the estimates shown earlier.

Table 7 Main method of TV reception, with market data for satellite and cable (%)

Main method of TV reception	Online original	Face-to-face	Online with five rims	Market data
Freeview	23	33	30	
Satellite	45	41	42	38
Cable	20	13	15	14
Other	17	13	13	

Q11/12: Which if any of these types of television does your household receive?
If more than one, which would you regard as your main type of television?

Satellite data: http://corporate.sky.com/about_sky/key_facts_and_figures (accessed 22 January 2013), 10.268 million homes, end March 2012.

Cable data: <http://www.guardian.co.uk/media/2012/feb/08/virgin-media-posts-first-ever-profit> (accessed 22 January 2013), 3.625 million homes, end March 2012.

Total households in UK: <http://www.jicpops.co.uk/about.html> (accessed 2 June 2012), 26.835 million homes, end March 2012.

The Sky figure of 38% will understate the satellite figure, as some homes receive free services via satellite (Freesat) and a smaller number receive pay services other than Sky (these are primarily foreign language services). The likely level of these other satellite services is 2–3% of households, bringing the total figure up to just over 40%.

From this comparison, it can be seen that the face-to-face data (and indeed the reweighted online data) are in line with the market data available. What is clear is that the original online data, weighted just by demographics, generates large biases for the number of cable TV homes (and as a result for the number of homes where Freeview is the main reception method). The

additional weighting therefore appears to reduce or remove significant differences that were apparent from the original online data weighted just by demographics. But, as expected, there has been a cost in terms of reduced weighting efficiency, as Table 8 shows.

Table 8 Impact of weighting on weighting efficiency*

	Cumulative weighting efficiency (%)
Demographics	78
Q21: Frequency of viewing television that is not live scheduled TV	36
Q88: Awareness of catch-up TV services on PC?	30
Q27: Activities undertaken whilst watching TV	23
Q82: As far as you are aware is the internet regulated?	22
Q4: Working status	20

* Weighting efficiency is a measure of how much the weighting has reduced the effective sample size. A weighting efficiency of 50% is equivalent to saying that an initial sample size of 1,000 has been reduced to 500 for the purpose of calculating confidence intervals.

This table shows how the weighting efficiency reduces as additional weighting variables are added. Demographics alone result in a 78% weighting efficiency – this is primarily due to (deliberate) differential sampling by nation (over-sampling Northern Ireland and Wales). When frequency of viewing television that is not live scheduled TV is added, the weighting efficiency reduces to 36%. When awareness of catch-up services on PC is additionally added, the cumulative weighting efficiency reduces to 30%. Adding in each of the extra variables continues to reduce the weighting efficiency, but at a decreasing rate.

A weighting efficiency of 20% is quite severe, although it can be counteracted by multiplying the initial sample size by five to compensate and generate a sample with the same effective sample size as the original unweighted data. Unit costs of online surveys mean this is cost-effective compared to the original face-to-face survey. In addition, a hybrid of the type adopted by FRS would also probably be cost-effective.

The business outcome

In the event, Ofcom decided not to migrate the Media Tracker at this point in time, as it was felt that the level of weighting that needed to be applied could jeopardise the robustness of the data, even if the sample size were increased five-fold as suggested, and also the resulting cost savings were reduced.

Nevertheless, the exercise has been of commercial value as it provides a rationale for continuing to use a face-to-face methodology for this particular survey at present (although this will be re-evaluated every two to three years). It also confirms previous work (some of which we have referenced above) that migration for less technologically orientated subjects may work with the application of a small number of subject-specific additional weights (this will help Ofcom identify subjects for which an online approach may be appropriate). The project confirms that a non-judgemental approach to identifying the questions to use for these additional weights does work in practice.

Conclusions

This experiment has provided a useful adjunct to the body of evidence already in existence regarding the effects of modal switch. The key insights are as follows.

-
- To use online panel sources to generate estimates similar to those we would obtain by using probability-based methods, we need to identify and understand the key covariates of the particular survey topic, to enable an effective weighting regime to be developed.
 - Even when we can do that effectively, we still require a high-quality probability-based survey as the basis for the weights.
 - Respondents recruited by river sampling show similar characteristics to those recruited from online panels and thus do not solve the bias problem identified in the Media Tracker Survey.

From a technical perspective, applying discriminant analysis to data from both data collection methods has been seen to be a useful way to identify the most appropriate covariates, and results in a non-judgemental approach to this task. The covariates can then be added to the more conventional demographic weights to reduce, and possibly eliminate, any bias generated by use of a new mode of data collection.

To follow this process, it is clearly necessary to collect data from both traditional and online samples, to enable the most effective covariates to be identified. The methodology developed is thus most suitable for continuous or tracking surveys, rather than ad hoc or one-off surveys.

References

AAPOR (2010) Report on Online Panels. *Public Opinion Quarterly*, **74**, 4, pp. 711–781.

Cooke, M., Watkins, N. & Moy, C. (2009) A hybrid online and offline approach to market measurement studies. *International Journal of Market Research*, **52**, 1, pp. 29–48.

Duffy, B., Smith, K., Terhanian, G. & Brewer, J. (2005) Comparing data from online and face-to-face surveys. *International Journal of Market Research*, **47**, 6, pp. 615–639.

Williams, J. (2012) Forum: Survey design in an age of austerity: driving value in survey design. *International Journal of Market Research*, **54**, 1, p. 13.

About the authors

Tim Barber is a research director in the Media and Technology team at BDRG Continental. He first started using online interviewing in 1995 while at university, to provide primary research for his dissertation on Premiership football sponsorship. He has worked with Ofcom and its predecessors for more than ten years and has run the Media Tracker since 2008. He has spoken at a number of conferences about the findings of this research and experiment, including for the Market Research Society and Association of Survey Computing.

David Chilvers is a Director at BDRG Continental. He has more than 30 years' experience in market research, and specialises in the application of analytic tools to research and other data. He won the MRS Innovation Award in 2001 for a segmentation project for Electronic Arts, which transformed the company's marketing strategy. He was a segmentation adviser to COI until its demise and is a statistical mentor to Ofcom.

Sumran Kaul was Senior Audience Analyst at Ofcom, where he had responsibility for television, radio and internet audience intelligence. He also led ad hoc projects, including, among others, Ofcom's Media Tracker survey, which explores the public's attitudes and opinions on a range of media activities and issues. He joined Ofcom in 2008 from ITV, where he was a senior

market analyst in the commercial division, working closely with the sales and marketing teams.

Address correspondence to: Tim Barber, BDRC Continental, 229–231 High Holborn, London WC1V 7DA.

Email: Tim.Barber@bdrc-continental.com

© Copyright Warc 2013

Warc Ltd.
85 Newman Street, London, United Kingdom, W1T 3EU
Tel: +44 (0)20 7467 8100, Fax: +(0)20 7467 8101

www.warc.com

All rights reserved including database rights. This electronic file is for the personal use of authorised users based at the subscribing company's office location. It may not be reproduced, posted on intranets, extranets or the internet, e-mailed, archived or shared electronically either within the purchaser's organisation or externally without express written permission from Warc.

Downloaded from warc.com