Community Life Survey

Disentangling sample and mode effects

Joel Williams

TNS BMRB

Introduction

Note on October 2017 revision

In October 2017, an error in one of the analysis weights was identified and corrected. **This revision does not change any of the conclusions drawn from this work**. The 'sample effects' and 'mode effects' quoted in the report are of the same direction and magnitude as in the original report. For example, the correlation between the 'mode effects' based on the erroneous weight and those based on the revised weight is .98 (see Chart 2). However, almost all the numbers in the report have had to be updated - even if each revision is tiny – and the appendix, which describes the construction of the now-revised weight, is substantially different.

In addition, Kantar Public (formerly TNS BMRB) recommended applying some sensible constraints on the set of survey variables for which 'mode effects' and 'sample effects' are estimated. The practical consequence is that only 216 estimates are in scope in the revised report, compared to 267 in the original report. This leads to an increase in the typical size of 'mode effects' and 'sample effects' (see Tables 2 and 3) but no meaningful increase in the proportions that are significantly different from zero.

Finally, Kantar Public has added two explanatory paragraphs, one to more fully explain the role of the analysis weight that has been revised (in the appendix and a shorter version in the section 'Study method'), and the other to quantify the relative scale of 'mode effects' and 'sample effects' (in the section 'Results', just below Table 3).

Significant changes to the text are highlighted in red.

The original published report is available on request from evidence@culture.gov.uk.

The *Community Life* Survey was launched in 2012-13, employing a face-to-face interview design. It provides Official Statistics on issues that are key to encouraging social action and empowering communities, including volunteering, giving, community engagement and well-being.

While the face-to-face interview survey provides robust, population representative statistics, it is a resource intensive approach. Consequently, **TNS BMRB was commissioned to explore the feasibility of introducing an online version of the survey**, a method that has the potential to significantly reduce costs without reducing the sample size.

In the first phase of its work, TNS BMRB developed an online survey method that employed probability sampling methods similar to those used for the face-to-face interview survey. However, it found that **the two designs produced significantly different results** even when the sample profiles were aligned using an appropriate statistical technique. The question that arose was this: was the difference in results due to (i) the different modes of data collection (online self-completion questionnaires vs. face-to-face interviews) or (ii) the different sampling and fieldwork methods (despite significant efforts to statistically align the two *post-hoc*)? To answer this question **it is vital to disentangle sample and mode effects** in order to determine which has the strongest influence on the results.

Sample effects are due to the different sampling and fieldwork methods used for each version of the survey. Mode effects are due to the different data collection methods used for each version of the survey.

Sample effects are more problematic than mode effects because the interview survey has a much higher response rate than the online/paper questionnaire (60% compared to 17%). If there *are* sample effects, then it is reasonable to assume that the interview-based estimate is the more accurate of the two.

However, if there are mode effects, we may make a judgment as to which is the more accurate estimate, but that judgment is reliant on generalising findings from the methodological literature. There is no objective indicator specific to this survey and to each of the variables within it. In the past, this had led some researchers to treat a survey with a higher response rate as better *in all respects* than a survey with a lower response rate. This is wrong. The interview data collection mode should not be assumed to be a more reliable measurement tool than the self-completion questionnaire. In fact, for many items in the Community Life survey, the face-to-face interview is probably a *less* reliable measurement tool than the self-completion questionnaire (although the concept of 'better/worse' is problematic for variables with no objective underpinning, such as those using opinion scales).

This report describes the results of a research project designed to separately quantify the sample and mode effects and thereby gain a better understanding of the results obtained from the two survey methods.

Headline findings and conclusions

The evidence presented here suggests that **the difference in data collection mode is responsible for the bulk of the mismatch observed between the results** produced from the face-to-face interview version of the *Community Life* survey and the results produced from the online/paper questionnaire version.

This conclusion depends on a few (tenable) assumptions. Principal among these is the assumption that TNS BMRB has been able to construct a counterfactual, i.e. that online questionnaire data has been generated, which has the sample properties of the interview survey. *If* this is the case (as seems likely) then the Cabinet Office can be reasonably confident that the difference in sample composition between its two parallel surveys is not primarily responsible for the difference in estimates they produce. The difference in data collection mode is much more impactful.

That being said, this study covers total population estimates only. **Findings might be different if sub-groups were assessed separately**. Unfortunately, sample size limitations prevent this. Consequently, there is a risk – even if small – that sample effects *are* meaningful for some parts of the population even if not in aggregate for the total population. This study's absence of evidence for sample effects does not imply that none exist for any sub-group.

These findings raise the obvious question: is the online/paper questionnaire survey reliable? If sample effects *are* minimal then this boils down to a judgment with regard to the different measurement properties of the two data collection modes.

Research on mode effects is often limited by confounding factors and the literature is riddled with inconsistent findings. However, a number of consistent messages do emerge. These include a broad consensus that a small number of fundamental factors (often in combination) lead to mode effects:

- the presence or otherwise of an interviewer;
- aural versus visual presentation of questions; and
- other differences in the way questions or responses are presented in different modes (for example the presentation of 'don't know' codes or the use of instructions/guidance).

The literature suggests that mode effects are often small. In particular, questions eliciting factual information (e.g. working status) tend to be largely unaffected by data collection mode, provided that sensible design practice is followed. However, **mode effects tend to be larger – and sometimes very large – for questions about values or behaviour where there is a clear societal norm or ideal**. This is especially true if the respondent is asked to use a scale when answering. For these types of question, interviewer-administered surveys typically elicit more positive or 'socially desirable' responses than self-completion formats. There are lots of questions of this type in the *Community Life* survey and - as predicted by the literature - the largest differences between the survey methods are found for these items. **Here, it seems reasonable to assume that the online/paper questionnaire is the better data collection tool**.

However, there are other variables where the interview might yield more accurate data. The literature suggests that some respondents take shortcuts (or 'satisfice') when answering questions, for example by focusing on response codes towards the top of the list, or routinely selecting midcategories when presented with a response scale. This tendency to 'satisfice' will affect both interviews and self-completion questionnaires but the effect is stronger for self-completion questionnaires. It is thought that the interviewer's presence encourages more conscientious response behaviours. In contrast, some self-completing respondents will speed through the questionnaire – especially if it is long and cognitively demanding - leading to lower data quality. Consequently, **we would expect the interview to yield better data if the question is complex or the response task demanding**.

The problem is that these are all *assumptions*. There are no benchmarks for the substantive questionnaire variables that would allow objective judgment of which data collection instrument is better. It is TNS BMRB's view that the *Community Life* self-completion questionnaire is likely to collect more accurate data than the interview for the majority of items in the *Community Life* survey. However, the lower response rate obtained for the online survey compared to the interview survey might add a small degree of bias to some estimates. On balance, **TNS BMRB expects the *net* error to be lower with the online/paper questionnaire survey than with the interview survey but this assertion is very much an informed opinion rather than a fact**.

Study method

The objective of this study was to better understand the difference observed between the results produced from the face-to-face interview version of the *Community Life* survey and the results produced from the parallel online/paper questionnaire version. In summary, the objective was to break this down into two components: *sample effects* and *mode effects*.

TNS BMRB planned two analysis projects to help identify the sample and mode effects that contribute to the difference in results produced by these two data collection systems.

Analysis project 1 was designed to quantify the difference in survey estimates that is due to the *different methods of sampling and fieldwork* used by the two data collection systems (system 1: up to six in-person interviewer visits; system 2: up to four letters requesting participation online (or on paper if preferred)).

Analysis project 2 was designed to quantify the difference in survey estimates that is due to the *different modes of data collection* used by the two data collection systems (system 1: in-person interview; system 2: online/paper questionnaire).

A summary of the design of each survey is given in table 1.

Table 1: The two data collection systems used for the *Community Life* survey

	System 1: Face-to-face interview survey	System 2: Online/paper questionnaire survey
Sample	Clustered & stratified random sample of addresses from the Postal Address File	Unclustered & stratified random sample of addresses from the Postal Address File
Within-household selection of respondent(s)	One adult selected at each sampled address (selection/verification by interviewer)	All adults selected at each sampled address, up to a maximum of four (no verification)
Data collection mode	In-person interviewing	Online (or paper if requested) self-completion questionnaires
Response rate	c60%	c17%

Both analysis projects were dependent upon an additional data collection stage: an online survey of respondents who had previously participated in the interview version of the *Community Life* Survey. Over the course of July-September 2014, TNS BMRB re-contacted as many as possible of those who took part in the face-to-face interview version of *Community Life* 2013-14 and had given their consent to be re-contacted. They were asked to complete the questionnaire online or on paper if requested. The data collection model matched the method used for the 'standard' 2014-15 online/paper questionnaire version of *Community Life* albeit a named individual was asked to complete the questionnaire rather than 'all adults' in the household.

In total, 5,105 interviews were carried out face-to-face in 2013-14 and 4,219 (83%) of these were invited to complete the 2014-15 questionnaire. In total, 1,576 did so. This represents a 37% response rate among invited cases, and a 31% response rate once the non-invited cases (i.e. those who did not consent to be re-contacted) are added to the denominator. The cumulative response rate after two surveys (the first collecting data face-to-face and the second collecting data online/on paper) was 19%.

TNS BMRB then constructed a model of response (see appendix 1) to the online/paper survey request, using variables from the original face-to-face interview as predictors. One of the outputs from this modelling procedure was an estimated probability of response for each case. This was converted into a weight that compensates for differential non-response to the online/paper questionnaire survey[1]. On the assumption that this weighting procedure was effective, this new data ought to have the sample properties of the interview survey but the measurement properties of the online/paper survey. This assumption is supported by the fact that, once this weight is applied, there is a very strong alignment between the interview data from the original 5,105 respondents and the interview data from the subset of 1,576 that later completed the online or paper questionnaire.

The weighted online/paper questionnaire data was used in two analysis projects.

Analysis project 1

For the first project, this new data was compared directly to the contemporary data (July-September 2014) collected for the 'standard' 2014-15 online/paper questionnaire version of *Community Life*. Any differences in results ought to reflect the difference in sample characteristics since the data collection tool is common to both samples.

Analysis project 2

The new data was also used to quantify mode effects. For this project, a 'difference-in-difference' method – frequently used for programme evaluation - was used. Two datasets were produced for this work.

Dataset 1 comprised:

(a) all face-to-face interview data from the period April 2013 - March 2014 (n=5,105), and

(b) all face-to-face interview data from the period July 2014 - September 2014 (n=666).

Given the common data collection system, the difference between (a) and (b) reflects the effect of *time*. On average, the time gap between (a) and (b) is ten and a half months but each dataset covers several months of data collection. The latest data from (a) was collected only three months before the earliest data from (b).

Dataset 2 comprised:

(a) a subset of face-to-face interview data from the period April 2013 - March 2014, namely the 1,576 who later completed a second online/paper questionnaire, and

---

[1] TNS BMRB selected 34 variables from the initial face-to-face interview to test as predictors of response to the online/paper questionnaire survey request. This total was reduced to 15 by discarding variables which were not additionally predictive of response. The final model has a fairly low level of variance explanation (Nagelkerke R^2 = 13%) but this is typical for survey response models. In this case, a low fit score suggests that re-contact survey response probability is not strongly associated with the first interview variables and that, consequently, a response rate of 31% is not a strong cause for concern. Where fewer – or less relevant – candidate variables are available, this assumption may be untenable.

(b) the same individuals' responses to the online/paper questionnaire (July 2014 - September 2014)[2].

The difference between (a) and (b) reflects the effect of both time *and* a change of data collection mode. The average time gap is the same as for dataset 1: ten and a half months, though varying from three to eighteen months in individual cases.

By subtracting the effect of time estimated from dataset 1, we should be left with just the effect of a change of data collection mode. This is the estimate of mode effect that is sought.

The key assumption here is that the effect of time is the same for datasets 1 and 2. The samples in dataset 1 are representative of the general adult population whereas the sample in dataset 2 is restricted to those who completed the second questionnaire and is therefore not necessarily representative of the general adult population. Consequently, there is a risk that the effect of time will have been different for this sample subset and that the mode effect might be mis-estimated as a result. To reduce this risk, the weight calculated for analysis project 1 has been applied to cases in dataset 2 to compensate for the difference in sample profile between datasets 1 and 2.

This report summarises the findings from this comparative analysis. Details about the weights used in each analysis are given at the end.

[2] The same data as used for analysis project 1.

Results

For this analysis, TNS BMRB has selected all questionnaire variables fulfilling two criteria: (i) there is a response from at least 80% of each data subset, and (ii) the proportion giving response x is between 5% and 95% for each data subset. These constraints strip out those questionnaire variables that are most strongly affected by the 'noise' produced by sampling variance.

There are 216 population estimates that fulfil these criteria (such as the proportion who feel they 'very strongly' belong to their immediate neighbourhood) spread between 76 variables, and 41 questions[3] (there are several variables for each multi-response question). The 216 separate population estimates are therefore clustered by variable and the variables clustered by question.

All the population estimates included in this study are *proportions*. A sample effect can be expressed simply as the difference between a proportion estimate derived from one sample and the same estimate derived from the other sample *should online/paper self-completion questionnaires be used with both samples*. In this case, the two samples compared are (i) respondents to the 'standard' online/paper questionnaire survey, and (ii) respondents to the interview survey.

Table 2 shows that the mean estimated sample effect was 1.9 percentage points. Overall, one third (31%) of sample effects are estimated to be smaller than one percentage point and three in five (61%) are estimated to be smaller than two percentage points. Only a handful (4%) are estimated to be greater than five percentage points.

Mode effects can be expressed in the same way, this time comparing a proportion estimate derived from an interview to the same estimate derived from an online/ data collection questionnaire *should the sample recruited for the interview survey complete both*.

The mean estimated mode effect was twice as large as the mean sample effect (3.8 percentage points compared to 1.9 percentage points). Overall, one in five (20%) of mode effects are estimated to be smaller than one percentage point and one third (33%) are estimated to be smaller than two percentage points. However, one in four (25%) are estimated to be greater than five percentage points with a small number (5% of the total) greater than ten percentage points.

Both the estimated sample effects and the estimated mode effects will be subject to some sampling variance. This means that even if the different sampling/fieldwork methods or the different data collection modes had **no** impact on the estimates (i.e. the 'null' hypothesis), some random differences between the results would nevertheless be observed.

The distribution of estimated sample effects broadly follows what would be expected if sample source had *no* systematic impact on the estimates (i.e. the 'null' hypothesis) albeit with slightly more large differences than the null hypothesis would predict. *If* there were no systematic differences attributable to sample source then approximately 5% of the estimated sample effects

[3] Although some of the questions use scale response options, TNS BMRB has treated all questions as categorical, leading to *k* variables per question, where *k* is the number of categories. The alternative of treating the response list as *metric* (i.e. with equal intervals between scale points) and analysing mean scores would hide any sample effects that are specific to some parts of the scale.

would have a t-score greater than 1.96[4]. TNS BMRB found that 11% of the estimated sample effects had t-scores above 1.96, only a little above the 'null' expectation of 5%. This is low enough for it to be reasonable to conclude that sample effects are small in the *Community Life* survey (especially as one would not expect the sample source to have *zero* impact (the premise of the null expectation)).

In contrast, TNS BMRB found that 42% of the estimated mode effects had t-scores above 1.96. This is substantially above the 'null' expectation of 5% and is strong evidence that mode effects *do* exist.

Table 2: Aggregated analysis of estimated sample and mode effects (online/paper self-completion questionnaire vs in-person interview)

	Estimated sample effect	**Estimated mode effect**
Mean absolute difference	1.9pp	3.8pp
Median absolute difference	1.6pp	3.0pp
% of differences <1pp	31%	20%
% of differences <2pp	61%	33%
% of differences <3pp	80%	51%
% of differences <4pp	92%	67%
% of differences <5pp	96%	75%
% of differences 5pp+	4%	25%
% of differences 10pp+	0%	5%
% of differences that are statistically significant (null expectation = c5%)	11%	42%

One straightforward way to assess the respective contribution of sample and mode effects to the overall 'data collection system effect' is to separately plot the estimated sample effects and the estimated mode effects against the observed difference in results produced from the two data collection systems. This can be seen in charts 1 and 2.

Chart 1 shows a positive but weak correlation between the estimated sample effects and the observed difference in population estimates produced from the two data collection systems (R=.36)[5]. This is consistent with a hypothesis that sample effects are mostly small.

---

[4] The t-scores are standardised versions of the observed differences, taking into account the different measurement properties of the variables under consideration as well as any differences in sample design or size.

[5] R is the correlation coefficient, ranging from -1 (the highest possible level of negative correlation) to +1 (the highest possible level of positive correlation). As a rule of thumb, R values between -0.3 and +0.3 are generally considered to be weak correlations. The square of R is the proportion of the variance of one set of estimates that can be 'predicted' by the other set. This is a frequently used statistic to summarise regression models.

Chart 1: Estimated sample effects plotted against observed differences between data collection systems (July-September 2014)

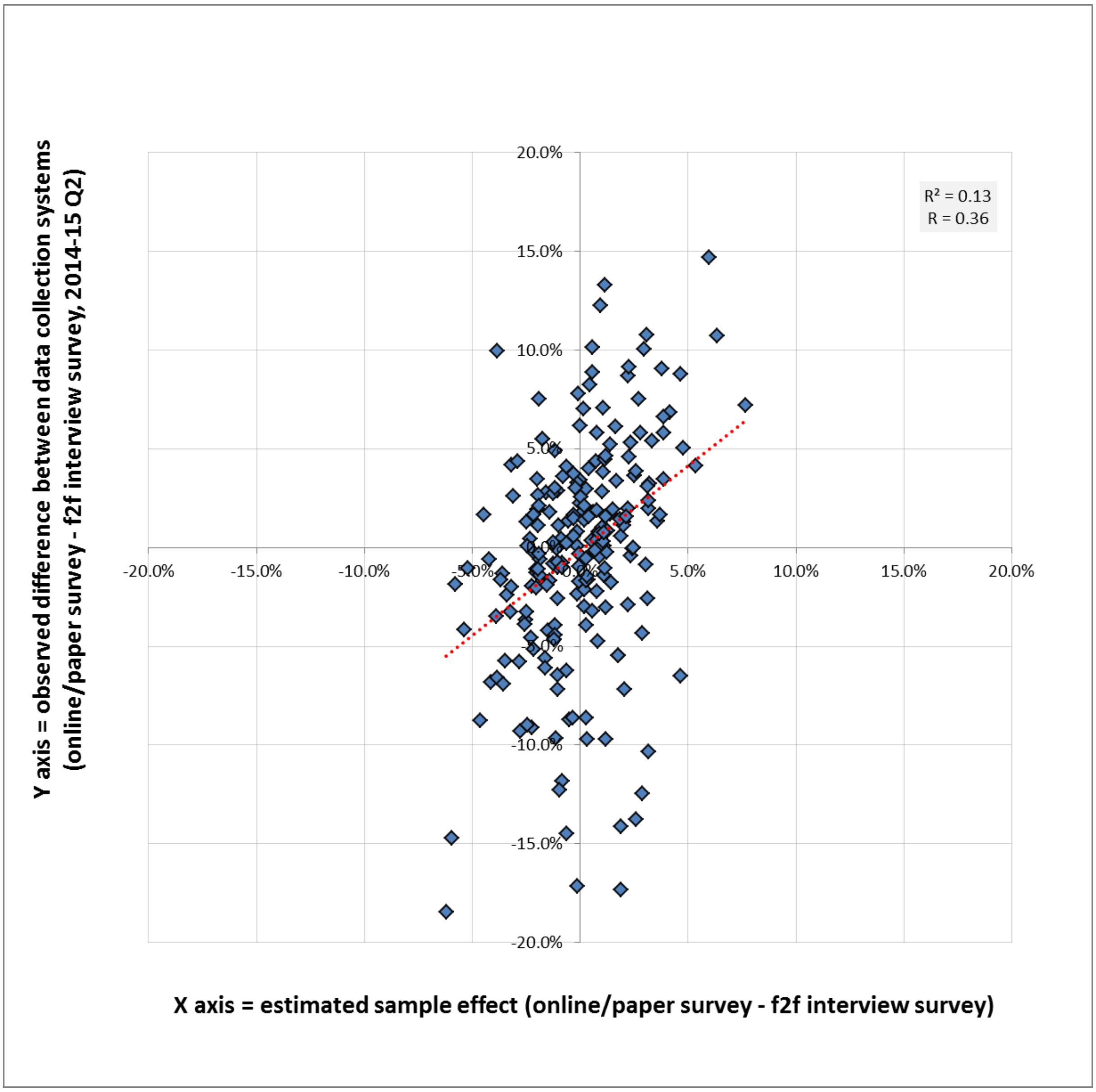
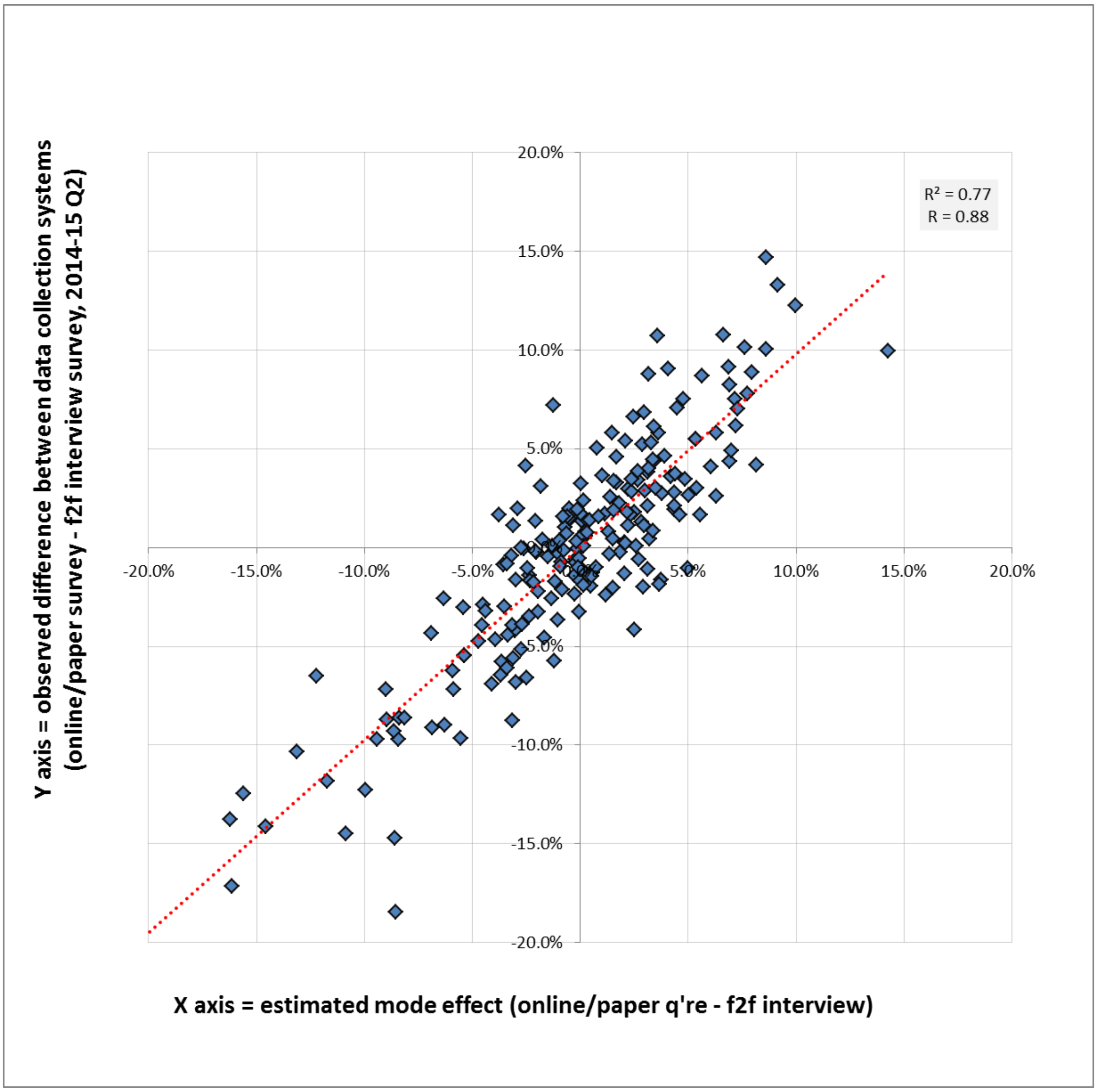


Chart 2 shows a positive and strong correlation between the estimated mode effects and the observed difference in results produced from the two data collection systems (R=.88). This is consistent with a hypothesis that mode effects have an influence on a large number of variables and, in many cases, this influence is very strong.

Chart 2: Estimated mode effects plotted against observed differences between data collection systems (July-September 2014)



These findings can also be shown in tabular form. Table 3 shows that the estimated mode effects are only very slightly smaller than the observed difference in results between the two data collection systems. This reinforces the view that mode effects are largely responsible.

Table 3: Aggregated analysis of estimated mode effects (online/paper self-completion vs in-person interview) against observed differences in results (July-September 2014)

	Observed difference between data collection systems (July-Sept 2014)	Estimated mode effect
Mean absolute difference	4.2pp	3.8pp
Median absolute difference	3.0pp	3.0pp
% of differences <1pp	18%	20%
% of differences <2pp	40%	33%
% of differences <3pp	50%	51%
% of differences <4pp	60%	67%
% of differences <5pp	68%	75%
% of differences 5pp+	32%	25%
% of differences 10pp+	8%	5%
% of differences that are statistically significant (null expectation = 5%)	39%	42%

One way of summarising the relative impact of mode effects and sample effects is to regress the system effects on these two effects simultaneously in the same 'model'. The R^2 - a popular measure of explanatory power – of this combined model is .96, compared to .77 and .13 respectively for models that contain just the mode effect or just the sample effect. From this we may estimate that the mode effect has four to five times the explanatory power of the sample effect.

Finally, table 4 shows the outliers, the ten questions with the largest estimated mode effects. Generally speaking, they are questions where some social desirability bias would be expected in an interview situation. In each case, the mode effect is in the direction expected (i.e. a less positive response given online/on paper than given in an interview).

Table 4: Largest estimated mode effects

	Estimated mode effect (online/paper mode minus interview mode)
WellB3 On a scale from 0 to 10, how anxious did you feel yesterday (= '0')	-20.2%pts
WellB4 On a scale from 0 to 10, to what extent do you feel that the things you do in your life are worthwhile (= '10')	-16.2%pts

SBeNeigh How strongly do you feel you belong to your immediate neighbourhood (= ‘very strongly’)	-16.1%pts
PInfl How important is it for you personally to feel that you can influence decisions in your local area? (= ’very important’)	-15.6%pts
WellB1 On a scale from 0 to 10, how satisfied are you with your life as a whole nowadays? (= ’10’)	-14.6%pts
WellB2 On a scale from 0 to 10, how happy did you feel yesterday? (= ’10’)	-13.1%pts
NComfort1 How comfortable would you be asking a neighbour to keep a set of keys to your home for emergencies (= ‘very comfortable’)	-12.2%pts
STogeth To what extent do you agree or disagree that this local area is a place where people from different backgrounds get on well together? (= ‘definitely agree’)	-11.7%pts
FrndSat2 To what extent do you agree or disagree that if I wanted company or to socialise, there are people I can call on (= ‘definitely agree’)	-10.9%pts
GGroup2l Bought goods from charity shop or catalogue (=’yes’)	+10.0%pts

Weights used in the analysis

Analysis project 1:

(a) all online/postal questionnaire data from those sampled for the ‘standard’ online/paper questionnaire survey, July 2014 - September 2014 (n=834).

Weight = design weight

(b) all online/postal questionnaire data from re-contacted interview respondents (initially sampled April 2013 – March 2014; second questionnaire completed July 2014 - September 2014) (n=1,576).

Weight = design weight divided by the estimated probability of completing the second questionnaire

Analysis project 2:

(1a) all face-to-face interview data from the period April 2013 - March 2014 (n=5,105)

Weight = design weight

(1b) all face-to-face interview data from the period July 2014 - September 2014 (n=666)

Weight = design weight

(2a) a subset of face-to-face interview data from the period April 2013 - March 2014, namely those who later completed a second online/paper questionnaire (n=1,576)

Weight = design weight divided by the estimated probability of completing the second questionnaire

(2b) the same individuals’ responses to the online/paper questionnaire (July 2014 - September 2014) (n=1,576)

Weight = design weight divided by the estimated probability of completing the second questionnaire

Appendix 1

Modelling response to the online/postal follow up survey

Respondents to the face-to-face *Community Life* survey (CLS-F2F) in 2013-14 were invited to respond to an online or postal follow up survey and their probabilities of providing data were estimated by means of a logistic regression technique. This technique predicts the outcome of either obtaining or not obtaining data given respondent characteristics recorded at the CLS-F2F[6].

Following a thorough examination of variables providing information on CLS-F2F respondents' characteristics and attitudes, a set of candidate predictor variables was selected for the model. Variables with substantial proportions of missing values were excluded from the set of candidate predictors to avoid suppressing the statistical power of the model. Depending on their frequency distributions, some categorical variables were re-coded in order to combine low frequency categories together.

The following CLS-F2F variables were tested as possible predictors:

1. The number of adults in the household;
2. The age of the respondent;
3. The gender of the respondent;
4. Whether the respondent was recorded as being single (i.e. never married and never registered in a same-sex civil partnership);
5. The respondent's number of children;
6. Whether the respondent reported having a paid job;
7. Whether the respondent reported using emails or the internet;
8. The respondent's frequency of internet usage;
9. The respondent's frequency of meeting up with family members or friends;
10. How strongly the respondent feels he/she belongs to their immediate neighbourhood;
11. The respondent's satisfaction with their local area as a place to live;
12. Whether the respondent voted in the last local election;
13. Whether the respondent had formally volunteered in the 12 months before the CLS-F2F;

[6] Note that the model denominator is *all* respondents to the interview survey, not just those who were willing to be re-contacted. The model therefore accounts for both stated willingness to be re-contacted and the probability of response if invited to take part in the survey.

14. The extent to which the respondent was satisfied with his/her life as a whole;
15. The extent to which the respondent was happy the day before the CLS-F2F;
16. The extent to which the respondent was anxious the day before the CLS-F2F;
17. The extent to which the respondent felt that the things he/she does in their life are worthwhile;
18. The frequency by which the respondent feels lonely;
19. The respondent's housing tenure;
20. Whether the respondent comes from a white ethnic background;
21. Whether the respondent's main language is English;
22. The respondent's religion;
23. The respondent's health status;
24. The respondent's level of civic participation in the 12 months before the CLS-F2F;
25. Whether the respondent has formal qualifications;
26. Whether the primary sampling unit the respondent was sampled from was located in an inner city area;
27. The primary sampling unit's Census 2011 proportion of the population self-defining as from an ethnic minority;
28. The urbanisation level of the area where the respondent lives;
29. The respondent's ACORN category classification (© CACI);
30. The respondent's gross income;
31. The region where the respondent lives;
32. Whether the respondent had a caring responsibility;

A 'stepwise' logistic regression process that eliminates uninformative candidate predictor variables one by one was employed to construct the final model.

The fifteen predictors in the final model are: (a) the number of adults in the household; (b) the age of the respondent; (c) the gender of the respondent; (d) whether or not the respondent reported using emails or the internet; (e) the frequency of internet usage; (f) whether the respondent reported having a paid job; (g) the respondent's frequency of meeting up with family members or friends; (h) the respondent's satisfaction with their local area as a place to live; (i) whether or not the respondent voted in the last local election; (j) whether the respondent had formally volunteered in the 12 months before the CLS-F2F; (k) the respondent's housing tenure; (l) whether or not the respondent comes from a white ethnic background; (m) the respondent's health status; (n) the respondent's civic participation in the 12 months before the CLS-F2F; and (o) the respondent's ACORN category classification.[7]

[7] The impact of multi-collinearity between predictors in the final model on the precision of the estimated coefficients was assessed by means of the Variance Inflation Factor (VIF). The highest level of multi-collinearity

Table A1 presents some key parameters of the model. The odds ratios show the ratio between the named sub-group's odds of responding to the online/postal follow up survey[8] and the odds of a reference group (described in square brackets). The lower and upper bounds of the 95% confidence intervals indicate the uncertainty around the quoted odds ratio. Finally, p-values for the coefficients that are under 0.05 indicate that the named sub group's odds of responding to the follow up survey are significantly different from the reference group's odds of responding, *controlling for other factors in the model*. Put another way, inclusion of the variable as a predictor makes a statistically significant difference to the model's fit with the data.

Table A1 – Predictive model of response probability to follow up survey: model parameters

Predictor	Category [vs. reference category], if predictor is categorical	Odds Ratio	Lower bound of odds ratio (95% C.I.)	Upper bound of odds ratio (95% C.I.)	p-value for coefficient
Number of adults in the household	2 adults [vs. 1 adult]	1.27	1.09	1.47	.00
	3 adults [vs. 1 adult]	1.19	0.93	1.52	.16
	4 or more adults [vs. 1 adult]	1.09	0.79	1.51	.60
Age of the respondent	26 to 34 [vs. 16 to 25]	1.22	0.90	1.66	.20
	35 to 49 [vs. 16 to 25]	1.23	0.92	1.64	.16
	50 to 64 [vs. 16 to 25]	1.83	1.37	2.46	.00
	65 to 74 [vs. 16 to 25]	1.87	1.34	2.61	.00
	75 plus [vs. 16 to 25]	1.70	1.17	2.47	.01

was detected for 'frequency of internet usage' (VIF=3.1) and 'whether the respondent uses emails or the Internet' (VIF=2.9), while VIF for the other variables was lower than 2.2.

[8] The odds of responding to the follow up survey represent the ratio of the probability of responding to the follow up survey to the probability of *not* responding. For example, if the probability of responding to the follow up survey for an individual in the reference group is estimated at 25%, then the odds of response are estimated at 0.33 (0.25/(1-0.25)). If the odds ratio for a non-reference group is 1.5 then the odds become 0.5 (0.33*1.5), and the probability of response becomes 33% (0.5/(0.5+1)).

Frequency of internet usage	Once a day [vs. more than once a day]	0.90	0.75	1.08	.27
	2 to 3 times a week [vs. more than once a day]	0.70	0.53	0.91	.01
	Less often than 2-3 times a week / not at all / refused to day [vs. more than once a day]	0.57	0.42	0.78	.00
Gender of the respondent	Male [vs. female]	0.78	0.69	0.89	.00
Working status of the respondent	Paid work [vs. no paid work]	0.75	0.63	0.88	.00
Respondent's frequency of meeting up with family members or friends	Once a fortnight or less often [vs. more frequently]	1.17	0.99	1.38	.06
Respondent's satisfaction with their local area as a place to live	Fairly or very satisfied [vs. not]	1.28	1.06	1.54	.01
Whether the respondent uses emails or the internet	Yes [vs. no]	1.71	1.23	2.40	.00
Whether respondent voted in the last local election	Yes [vs. no]	1.48	1.29	1.70	.00
Whether the respondent had formally	Yes [vs. no]	1.27	1.11	1.45	.00

volunteered in the 12 months before the CLS-F2F					
Respondent's housing tenure	Part or whole ownership [vs. no ownership]	1.47	1.26	1.71	.00
Whether respondent comes from a white ethnic background	Yes [vs. no]	1.70	1.35	2.15	.00
Respondent's health status	Bad or very bad [vs. better]	0.76	0.58	1.00	.05
Whether respondent reported civic participation	Yes [vs. no]	1.34	1.17	1.54	.00
Respondent's ACORN category classification	Comfortable communities [vs. others]	1.20	1.05	1.38	.01
Constant		0.06			.00

With respect to the fifteen predictor variables in the model, there is alignment between the respondents to the CLS-F2F in 2013-14 and the *weighted* subset that responded to the online/paper survey some months later. The weight is equal to one divided by the estimated probability of response. This follows as a consequence of model construction. However, this alignment is also observed across the set of 216 population estimates that are considered in the main body of this report: 16 of the 216 are significantly different at the 5% level, compared to a null hypothesis expectation of 11.