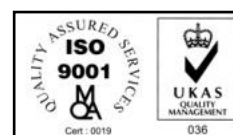


Community Life Survey

Investigating the viability of moving from a face-to-face to an online/postal mode: evidence from a series of methodological studies 2012-2015

August 2015

Becky Hamlyn, Alice Fitzpatrick and Joel Williams



Contents

Executive Summary	3
1. Introduction	6
2. Programme of methodological work: an overview	8
3. What fieldwork model should be employed?	13
4. What response rate can be achieved and how can this be maximised?	18
5. What is the level of compliance with the random adult selection and what steps can be taken to reduce non-compliance?	21
6. What is the profile of a web sample compared with a face-to-face sample?	24
7. What is the quality of the data using online designs?	26
8. Does the web/postal survey give similar estimates to the face-to-face survey both before and after demographic weighting?	30
9. If the online/interview results are still not aligned after weighting what is the main reason for this: residual sample bias or data collection mode?	33
10. Conclusions and recommendations	37
Appendix A: What weighting strategy is recommended for an online/postal Community Life Survey?	39

Executive Summary

The Community Life survey is a large scale face-to-face interview survey conducted using probability sampling methods. The survey, commissioned by the Cabinet Office since August 2012, tracks the latest trends and developments across areas that are key to encouraging social action and empowering communities.

The face-to-face survey provides robust and nationally representative Official Statistics but is an expensive and resource intensive method of undertaking research. As a result, TNS BMRB was commissioned to carry out development work to explore the feasibility of incorporating online/postal methods of data collection, which cost significantly less than face-to-face interviews, for future survey years. This work was carried out in parallel to the face-to-face survey.

The programme of methodological work was conducted between 2012 and 2015 and comprised four separate stages.

- Stage 1: Testing and refining an initial field model for online/postal survey delivery
- Stage 2: Larger scale online/postal pilot conducted alongside the face-to-face survey
- Stage 3: Testing the feasibility of sampling all adults in the household instead of one selected at random
- Stage 4: Study to investigate the relative contribution of sample effects and mode effects in explaining estimate differences between face-to-face and online/postal modes

Although these stages have been reported on separately (see Annex's) the purpose of this report is to bring together the findings from across the whole programme of methodological research in order to provide a more holistic set of findings and conclusions, and to help drive more informed decisions about the future of the survey. A summary of key findings is provided below.

What fieldwork model should be employed for an online/postal survey?

Based on a combination of experimental work, large-scale testing and knowledge of established best practice TNS BMRB recommends that a full online/postal survey model, if adopted, should be based on the following features:

- A random probability stratified sample of addresses drawn from the Postcode Address File (PAF)
- Survey invitations issued by letter; at each address all adults (up to a maximum of four) invited to take part
- Up to four mailings sent to each address to maximise response
- Postal survey available as an alternative to ensure inclusion where the household lacks internet access - postal survey to be made available on demand and also sent physically in the post to a targeted subset in second reminder packs

- Questionnaire designed to reflect the face-to-face survey construction as far as possible to maximise compatibility
- £10 incentive to each person completing the survey to maximise response

What response rate can be achieved and how can this be maximised?

- Although a £5 unconditional incentive is more effective in raising response rates, a conditional incentive of £10 represents the most cost effective strategy.
- Including an option to complete the survey on paper is effective in raising response rates, although the impact is more pronounced when the postal questionnaire is included in one of the mailings, rather than only being available on request.
- At Stage 2, based on a conditional £10 incentive to one adult in the household and postal survey available only on request, the web response rate was 25%, rising to 28% once the postal returns are added.
- When the Stage 2 design is amended to invite all adults (up to a maximum of four) to complete the survey the person-level response rate is estimated at 22%. However the lower response is compensated for by compliance with the sampling instructions (see Chapter 5 for more details).

What is the level of compliance with the random adult selection and what steps can be taken to reduce non-compliance?

- A web design does not allow for true random selection of an individual at each address as it cannot be assumed that the sampling procedures will be applied correctly in the absence of an interviewer.
- Therefore, a quasi-random approach was used, which involved the household selecting the adult with the "last birthday" (by Stage 2 this was tweaked so that half of households were asked to select the adult with the "last birthday" and half with the "next birthday").
- At Stage 2, one in four adults was identified as the "wrong" household respondent. Non-compliance may result from either the wrong adult completing the questionnaire to get the incentive or human error.
- To overcome this, a design that examined extending the invitation to all adults in the household was tested. This was found to reduce the level of non-compliance, although evidence suggested that occasionally one adult completed several questionnaires in order to obtain a larger incentive.
- Although the "all adults" approach is recommended, a range of further measures are recommended to validate responses.

What is the profile of an online/postal sample compared with a face-to-face sample?

- As the response rate to the web/postal survey is significantly lower than the response rate to the face-to-face survey, the risk of non-response bias is much greater.
- Compared with the face-to-face survey, the self-completion sample is more likely to comprise everyday internet users, high earners, owner-occupiers, native English speakers and the more highly educated.
- The paper questionnaire option helps bring in non-internet users, although take-up of these questionnaires on-request is low and there is minimal improvement in sample profile unless included with a targeted subset of reminder packs.

What is the quality of the data using an online/postal design?

- In general, there are no glaring concerns about the quality of the online data compared with face-to-face data, based on quality metrics such as time taken to complete and measures to detect survey "satisficing" (i.e. short-cutting tactics).
- However, there are indications of lower levels of engagement with the online survey, and evidence of a small risk that in some cases one adult completes several questionnaires.
- Different presentations of "don't know", "prefer not to say" and other spontaneous codes can affect their level of use in the online survey. While the presentation can be manipulated to more closely mirror the "face-to-face" survey there is no evidence that this produces results which are closer to the "true" opinions of respondents.

Does the online/postal survey give similar estimates to the face-to-face survey both before and after demographic weighting?

- Before weighting, around two-thirds of the web sample estimates are significantly different from the interview sample estimates; around 20% of these differences are substantial (at least five percentage points differences).
- After weighting to correct for demographic profile differences, the results are only slightly more aligned, which indicates that data collection mode and/or residual sample bias influence the results to the online survey. The relative contribution of these two effects was tested in Stage 4 (see Chapter 9).

If the results are still not aligned after weighting what is the main reason for this: residual sample bias or different data collection modes?

- A further set of experiments was devised to estimate the relative contribution of data collection mode on the one hand and sampling/fieldwork methods on the other to explaining the differences in results between the web/postal and interview designs.
- The evidence suggests that the difference in data collection mode is responsible for the bulk of the mismatch observed between the results.
- The wider evidence suggests that a self-completion mode is usually a better tool for questions which are sensitive or subject to social desirability bias. As these types of questions make up a large proportion of the Community Life survey, it is likely that the online survey will produce more accurate results than the interview survey for most (but not all) survey questions.

Conclusions and recommendations

- With caveats, the different sampling/fieldwork methods do not appear to influence the distribution of the data so the lower response rate of the online/paper method compare to the interview method should not be a strong concern.
- However, there are very strong mode-conditioned measurement effects which make the two modes incompatible on many variables; if the online/paper mode was the only one commissioned, a break in the time series would need to be acknowledged and the nature and implications of this break would need to be clearly explained to users.
- TNS BMRB are still evolving the method but great strides have been made to improve the sample profile to a point where this is a better option than any random sample telephone survey. It achieves a sample profile that is biased towards the better educated and the better off but not much more so than the face-to-face interview survey.

1. Introduction

The Community Life survey, a large scale face-to-face interview survey conducted using probability sampling methods, was launched in August 2012. The survey was commissioned by the Cabinet Office to track the latest trends and developments across areas that are key to encouraging social action and empowering communities. The survey, which has been conducted by TNS BMRB since its launch, tracks measures including:

- Volunteering and charitable giving;
- Views about the local area;
- Community cohesion and belonging;
- Community empowerment and participation;
- Influencing local decisions and affairs; and,
- Subjective well-being

The survey also incorporates a number of key measures from the Citizenship Survey¹, which ran from 2001 to 2011.

In 2012-13, survey data were collected using face to face interviews with a representative sample of 6,600 individuals in England which, because of constraints in the commissioning timetable, ran from August 2012 through to April 2013, covering three-quarters of a year. In 2013-14, the face-to-face survey was reduced to just over 5,000 interviews across the year and then to 2,000 interviews in 2014-15.

The face-to-face survey provides robust and nationally representative Official Statistics but is an expensive and resource intensive method of undertaking research. As a result, TNS BMRB was commissioned to carry out development work to explore cost effective methods for future survey years. In particular, TNS BMRB explored options for incorporating online/postal methods of data collection, which cost significantly less than face-to-face interviews. This work was carried out in parallel to the face-to-face survey.

To provide evidence to support future decision-making, TNS BMRB carried out a series of methodological projects which investigated the viability of moving to an online/postal method to provide the majority of the Community Life users' data needs. The methodological work began with a large scale test of a probability sample online/postal survey in Autumn 2012. The design of this initial test was refined following consultation with a Technical Advisory Group of senior survey

¹ The Citizenship Survey began as the 'Home Office Citizenship Survey' (HOCS) before the responsibility moved to the Department of Communities and Local Government (DCLG) in May 2006.
<http://discover.ukdataservice.ac.uk/series/?sn=200007>

methodologists. Methodological development was also informed by a consultation among survey users².

This first test helped TNS BMRB and the Cabinet Office to draw initial conclusions on the viability of an online/postal survey. The results from this initial test were broadly positive. However, a number of questions remained about a) the optimal method of sampling individuals within households and b) the impact on the survey estimates of a change in (i) sampling and fieldwork method, and (ii) data collection mode.

As a result, TNS BMRB conducted a number of further methodological projects over the course of 2012-2015 in order to a) finalise the optimal design of an online/postal survey model; b) draw stronger conclusions about the relative impacts of sampling/fieldwork effects and mode effects and; c) develop recommendations for the longer-term future of the survey taking into account the desire for both robust official estimates and time series continuity.

The programme of methodological work was conducted between 2012 and 2015 and comprised four stages. Separate reports are available for these stages (see Annex's)

The purpose of this report is to bring together the findings from across the whole programme of methodological research in order to provide a more holistic set of findings and conclusions, and to help drive more informed decisions about the future of the survey.

1.1 Structure of this report

Chapter 2 provides an overview of the programme of methodological work. Chapters 3 to 9 then cover the main findings and conclusions under a series of headings which address the key questions the methodological work sought to address. These are as follows:

- What fieldwork model should be employed?
- What response rate can be achieved and how can this be maximised?
- What is the level of compliance with the random adult selection and what steps can be taken to reduce non-compliance?
- What is the profile of an online/postal sample compared with a face-to-face sample?
- What is the quality of the data using an online/postal design?
- Does the online/postal survey give similar estimates to the face-to-face survey both before and after demographic weighting?
- If the results are still not aligned after weighting what is the main reason for this: residual sample bias or different data collection modes?

Finally Chapter 10 sets out the conclusions from the programme of methodological work. This includes discussion and recommendations on the future implementation of the online survey as the core mode for generating Community Life statistics.

2. Programme of methodological work: an overview

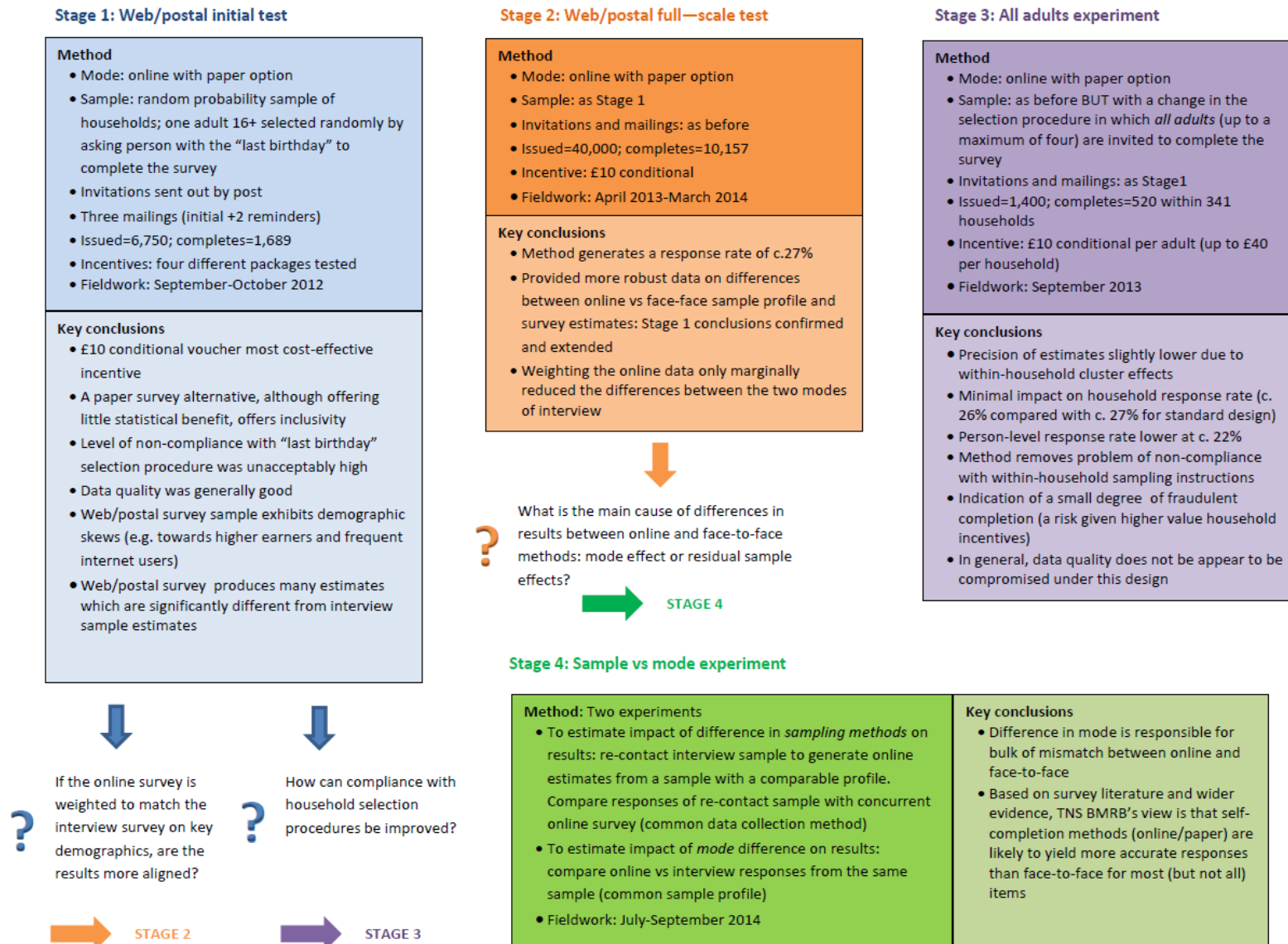
The methodological work was conducted over four stages:

- Stage 1: Testing and refining an initial field model for online/postal survey delivery
- Stage 2: Larger scale online/postal pilot conducted alongside the face-to-face survey
- Stage 3: Testing the feasibility of sampling all adults in the household instead of one selected at random
- Stage 4: Study to investigate the relative contribution of sample effects and mode effects in explaining estimate differences between face-to-face and online/postal modes

Figure 1 displays a summary of the four stages of methodological work and how the stages linked together.

A description of the four stages then follows.

Figure 1: Summary of methodological work



2.1 Stage 1: Testing and refining an initial field model for online/postal survey delivery

The aim of the first stage was to scope and test a potential field model for online/postal survey delivery. The test was based on a relatively large-scale random probability survey with c. 6,700 issued addresses, with invitations to access the online survey sent out by post. At each address, one adult aged 16+ was invited to take part, with selection based on the adult in the household who had the "last birthday". Up to two postal reminders were sent to each address to maximise response.

Four different incentive packages were tested in the first letter: £5 conditional voucher; £10 conditional voucher; £5 unconditional voucher included; no incentive (the control group).

A random subset of non-responders received a postal questionnaire with their second reminder, which they could complete instead of the online version. The paper version of the questionnaire was an edited version of the online interview with a reduced number of questions, as the full survey was too lengthy to accommodate on paper.

Fieldwork took place in September-October 2012.

This stage set out to explore:

- What response rate can be achieved and how does this vary by different incentive packages?
- What value, in terms of response rate and survey quality, is added by the inclusion of a paper postal questionnaire for those who are unable or do not wish to complete the survey online?
- What is the profile of the achieved online/postal sample and how does this compare with the face-to-face interview sample profile?
- What is the quality of the data using an online/postal approach in terms of measures, such as time taken to complete, dropout rates and use of don't know/refusal codes?
- What is the level of compliance with the 'last birthday' rule for selection of a random adult?
- Does the online/postal survey produce similar estimates to the face-to-face survey for key measures?

2.2 Stage 2: Larger scale online/postal pilot conducted alongside a concurrent face-to-face survey

Based on recommendations for an optimal design arising from Stage 1, a larger scale online/postal test survey was conducted which ran concurrently with the standard face-to-face survey during the survey year April 2013-March 2014.

A larger annual sample size of c.10,000 achieved online/postal questionnaires over the survey year provided a more robust test of differences in a) sample composition and b) measurement between online/postal and face-to-face modes.

The larger sample size also allowed some initial exploration of the relative contribution of sampling/fieldwork methods and data collection mode in explaining differences between online/postal and face-to-face survey estimates (something explored more fully in Stage 4).

The Stage 1 study resulted in a set of recommendations for an optimal design which was then tested in this much larger-scale feasibility test. Stage 2 was based on the following features:

- Online survey methodology with option for paper version on-demand
- Incentive package of £10 voucher conditional on completion (administered through Perks, an online rewards platform and as a paper voucher for those completing by post)
- A change in the method for selecting one individual per household to a more balanced split sample design whereby half were asked to select the adult aged 16+ with the "last birthday" and half were asked to select the adult with the "next birthday".
- Changes in the presentation of "don't know" and "refused" codes in the online survey questionnaire to create greater compatibility with the face-to-face survey

2.3 Stage 3: Testing the feasibility of sampling *all* adults in the household instead of one selected at random

This stage was conducted as a concurrent extension to the Stage 2 large-scale online/postal survey experiment. This stage involved an additional issued sample of n=1,400 addresses in Quarter 3, which were issued with different instructions for selecting which adults should complete the survey; all other features of the survey were the same as the standard online/postal survey tested in Stage 2.

In this variant the invitation letter invited *all* adults in the household (up to a maximum of four) to complete the survey with an incentive of a £10 voucher offered to each responding adult. This was proposed as a solution to the problem of non-compliance with within-household sampling instructions identified at Stage 1.

The study allowed a comparison of the all adults vs single adult design on several measures including: completion behaviour associated with higher household conditional incentives; sample profile; and data quality.

Fieldwork for Stage 3 was conducted in September 2013.

2.4 Stage 4: Study to investigate the relative contribution of sample effects and mode effects in explaining differences between face-to-face and online/postal modes

Analysis of Stage 2 data revealed that the face-to-face and online/postal models produced significantly different results on key measures, even when the samples were aligned in terms of key demographics. This left a key unanswered question: were these residual differences in results due to i) the different mode of data collection (i.e. people giving different answers due to the different presentation of the questions and/or the presence/absence of an interviewer) or ii) residual systematic differences between the samples?

Stage 4 involved an additional experiment that had two parts as follows:

- *Estimating the relative contribution of sampling/fieldwork effects:* this was achieved by generating online survey estimates from a sample with the same properties as the sample recruited for the interview survey. To do this, TNS BMRB re-contacted respondents who had taken part in the face-to-face interview survey and asked them to complete the survey online. Comparing the responses of this sample with the concurrent "standard" online version of the survey allowed direct quantification of sampling/fieldwork effects after controlling for mode.

- *Estimating the relative contribution of mode effects:* this was achieved by comparing the responses of the re-contact sample (see above) with the responses of the same sample when they completed the original face-to-face interview while controlling for the effects of time by using the benchmark interview data from both time-points.

The sampling/fieldwork effects and mode effects were then plotted against the observed differences between the interview and online survey, which allowed conclusions to be drawn with regard to their relative impact.

3. What fieldwork model should be employed for an online/postal survey?

Summary

Based on a combination of experimental work, large-scale testing and knowledge of established best practice TNS BMRB recommends that a full online/postal survey model, if adopted, should be based on the following features:

- A random probability stratified sample of addresses drawn from the Postcode Address File (PAF)
- Survey invitations issued by letter; at each address all adults (up to a maximum of four) invited to take part
- Up to four³ mailings sent to each address to maximise response
- Postal survey available as an alternative to ensure inclusion where the household lacks internet access - postal survey to be made available on demand and also sent physically in the post to a targeted subset in second reminder packs
- Questionnaire designed to reflect the face-to-face survey construction as far as possible to maximise compatibility
- £10 incentive to each person completing the survey to maximise response

In this chapter we document some of the findings from Stages 1 and 2, which between them tested a number of fieldwork features. This has helped to refine a best-practice model for a random probability online/postal survey. This chapter covers:

- Sample design
- Initial invitation to take part
- Reminder mailings
- Offering a postal alternative
- Layout the questionnaire
- Incentives (covered in more detail in Chapter 4)
- Random selection of individuals within households (covered in more detail in Chapter 5)

3.1 Sample design

At Stage 2 (2013-14) the Community Life online/postal survey delivered a nationally representative sample of c. 10,000 adults aged 16 years and over in England. The fieldwork took place over four quarters between June 2013 and March 2014. One individual was invited to take part at over 40,000 addresses, with approximately 10,000 interviews completed over the four quarters of fieldwork (c. 2,500 a quarter).

³ In this survey three mailings have been used, but response may be increased further with the use of up to four mailings

The survey retained the random probability approach used in the face-to-face-survey, and used the postcode address file held by the Post Office as its sample source. A systematic random sample design⁴ was employed, with two steps as follows.

Step 1:

Firstly, a random sample of 40,620 English addresses was drawn from the Postcode Address File (PAF) and allocated to survey quarter (10,155 addresses in each) in such a way that each sample quarter was representative of the full PAF.

Before selection, the PAF was filtered to exclude addresses for businesses or other organisations or communal establishments. The addresses were then sorted by (i) local authority, (ii) Lower-layer Super Output Area (LSOA) and (iii) postcode to ensure that a systematic sample drawn from the PAF would have maximal geographic dispersion; a design feature that usually enhances the precision of survey estimates compared to more clustered designs.

Addresses were listed alphanumerically within each postcode before the sample was drawn.

Step 2

As explained further below, a letter was sent to each sampled address inviting one adult aged 16+ to complete the survey online (or to request a postal version of the questionnaire) based on (quasi) random selection.

All residential addresses in England had an equal probability of selection, but the total sampling probability of each individual varied due to differing within-address sampling fractions⁵ (for example individuals living in a one person household had a higher chance of selection than individuals living in larger households). These variations in sampling probability are compensated for when the data are weighted.

3.2 Inviting the household to take part

The sample for the online/postal survey was based on postal addresses. Therefore the first contact with the household was by post in the form of an invitation letter (which is equivalent to the advance letter sent out to all households selected as part of the face-to-face survey).

The letter explained the purpose and nature of the interview and contained a simple link to a web survey, plus a unique user name and instructions for selecting which individual to take part. The letter explained that the survey could be completed in any location with internet access including on a desktop computer, laptop or tablet. However, survey completion on a smartphone was discouraged for quality control reasons, as the questionnaire was not specifically designed (or tested) for use on devices with very small screen sizes.

Instructions were provided explaining how to request a paper version of the questionnaire for those who preferred not to - or who were unable to - complete the survey online. The letter also mentioned the incentive that was due on completion of the survey. On the reverse of the letter more detailed information about the survey was included in the form of an FAQ.

⁴ Systematic random sampling is a type of probability sampling technique. With a systematic random sample, the sample frame is before drawing 1 in n addresses from a random start-point.

⁵ The within-address sampling fraction is the proportion of individuals living at the address that is asked to take part in the survey.

3.3 Reminder mailings

In line with best practice for self-completion surveys the online/postal survey involved three mailings. Following the initial letter up to two reminder letters were sent to target those who had not responded to the previous round. Mailings were scheduled to provide a gap of around 14 days between mailings and were timed so that the letters arrived for the weekend. The reminder letter contained the same information as the initial letter and provided a prompt about how to access the survey online, as well as how to access the paper version of the survey.

Reminder mailings were found to have a significant impact on response rates, especially between the initial mailing and the first reminder. At Stage 1 the conversion rate⁶ was found to double or almost double between these two mailings within each of the different experimental cells. Between the first and second reminders the impact was reduced, although it was still effective in raising response rates beyond the level achieved after the second stage.

3.4 Offering a postal alternative

Offering a postal alternative increased population representation by allowing those without internet access to take part in the survey.

However, in order to keep the paper questionnaire to an acceptable length, the postal questionnaire was designed as a slightly truncated version of the online questionnaire, focussing on the most important survey measures. As it does not offer full coverage of the questionnaire, the expanded sample coverage is only applicable to questions common to both forms of the questionnaire.

At Stage 1, for a random subset of non-responders, a postal questionnaire was added at the second reminder stage and was included in the mailing with a postage-paid return envelope. However, at Stage 2 the postal questionnaire was made available on demand from the start of fieldwork and was requested via a telephone orderline. This approach ensured that the full-content web version was the default mode of completion for those with online access.

At Stage 2 the inclusion of a postal alternative increased the response rate significantly (see Chapter 4). However, despite this, the *uniform* addition of a postal questionnaire to the second reminder pack had relatively little impact on the quality of the sample profile: the profile by housing tenure and working status improved slightly, while there was no improvement in the age profile and its value is more limited because of the omission of some of the questionnaire content (see Chapter 6). With that in mind it was originally recommended that the postal questionnaire should be available only on demand as per the model adopted in Stage 2. However (currently unpublished) evidence from another test of this method - albeit for a different survey sponsor - suggested that the sample profile *could* be improved by placing postal questionnaires in approximately half the second reminder packs and targeting these packs at more deprived areas of the country. This evidence was strong enough for this approach to be adopted for the Community Life survey from 2015-16.

⁶ The conversion rate is the number of completed questionnaires expressed as a proportion of the sampled addresses. This is not the same as the response rate, which is usually higher and is estimated by taking account of the expected c.9% of PAF addresses that are classified as non-residential in face-to-face interviews.

3.5 Questionnaire content and presentation

The online questionnaire and postal surveys were designed to mimic the face-to-face version as far as possible, although mode constraints meant that this was not always possible. For example questions in the face-to-face survey, which were designed to be read out by an interviewer, or where responses were presented on a show card, were re-phrased in the online/postal version to suit a self-completion context.

Following the initial testing at Stage 1, a number of enhancements to the questionnaire were made to improve user experience and face-to-face/online comparability. These were as follows:

Presentation of don't know/refused and other spontaneous categories

One of the most significant issues which affected compatibility between the two modes was the presentation of 'Don't know', 'Prefer not to say'⁷ and other spontaneous answer options. In the face-to-face survey script the option to select these answers is available at every question, although they are only used when the respondent specifically volunteers them as an answer. This presented a challenge in the design of the online and paper questionnaires as this approach could not be replicated exactly.

In order to replicate the face-to-face approach as far as possible, where these codes were not available in the main response list, these responses were initially hidden in the online survey and appeared on a 'second screen'. At each question respondents were shown the normal response list without the 'Don't know', 'Prefer not to say' or other spontaneous options. If they clicked forward to move to the next question without selecting an answer, these codes then appeared on the screen. This ensured the online survey was as compatible to the face-to-face survey as possible in this respect.

The process of how to display these answer options was explained to the respondent at the beginning of the survey with additional prompts at selected questions where relevant. In addition to this, help links were placed on screen at questions covering particularly sensitive issues or other questions where 'Don't know' and 'Prefer not to say' codes were expected to be used more, by at least 3% of respondents.

With the paper questionnaire, hiding certain responses was not an option due to the questionnaire being completed on paper. In order to ensure the responses were as compatible as possible between the different modes, the 'Don't know', 'Prefer not to say' and spontaneous codes were generally not included in the paper questionnaire except where we expected the use of these codes to be at least 5% (based on the face-to-face survey approach – these were usually the more sensitive questions). Respondents were informed in the instructions on the front that they should leave the questions blank if they did not want to or were not able to provide an answer.

Progress bars/survey banners

To provide an indication of how far through the online questionnaire a respondent had reached a progress bar was added to the start of each section and within some of the longer sections.

HM Government and survey logos were also added to the banner at the top of the survey screens to provide a more professional look.

⁷ In the web survey, the 'Refused' option was re-named as 'Prefer not to say'

3.6 Incentives

As discussed further in Chapter 4, a number of different incentive packages were tested at Stage 1. A £10 conditional incentive was taken forward into Stage 2 as this was found to represent the best compromise between maximising response and cost efficiency. For the 'all adults' design tested at Stage 3 (see section 5) the incentive package changes from a £10 incentive for one randomly selected adult to a £10 incentive for up to four adults (a maximum total of £40 per household).

3.7 Random selection of individuals within households

Stage 1 and Stage 2 were based on the model that one adult aged 16+ is selected at random and is asked to complete the questionnaire. In a face-to-face survey random selection can be automated and controlled. However, in an online/postal survey we are reliant on the household applying correct procedures. At Stage 2 it was found that around a quarter of respondents (26%) were not the "correct" respondent i.e. the sampling instructions were ignored as the survey was not completed by the person with the right birthday. This clearly cast doubt on the representativeness of the sample. An "all adults" design, which allowed all adults (up to a maximum of four) the opportunity to complete the questionnaire removes the problematic within-household sampling stage. Although there are still some issues associated with this model (see Chapter 5 for more details), it is considered the best model in terms of ensuring a design which most closely adheres to random sampling ideals.

4. What response rate can be achieved and how can this be maximised?

Summary

- Although a £5 unconditional incentive is more effective in raising response rates, a conditional incentive of £10 represents the most cost effective strategy.
- Including an option to complete the survey on paper is effective in raising response rates, although the impact is more pronounced when the postal questionnaire is included in one of the mailings, rather than only being available on request.
- At Stage 2, based on a conditional £10 incentive to one adult in the household and postal survey available only on request, the web response rate was 25%, rising to 28% once the postal returns are added.
- When the Stage 2 design is amended to invite all adults (up to a maximum of four) to complete the survey the person-level response rate is estimated at 22%. However the lower response is compensated for by compliance with the sampling instructions (see Chapter 5 for more details).

In this chapter we cover the variation in response rates according to the results of different experiments, which sought to test the impact of different survey features on response rate.

Drawing on findings from Stages 1 to 3, in this chapter we consider the impact on response rate by:

- whether an incentive is included (either unconditional or conditional)
- whether a postal survey alternative is offered
- the number of people in the household who complete the questionnaire

In this chapter we refer to two measures of response:

- *Conversion rate*: This is the number of completed questionnaires expressed as a proportion of the sampled addresses. Under the 'all adults' design, the numerator is the number of addresses yielding at least one completed questionnaire.
- *Response rate*: The response rate is estimated and is calculated by dividing the conversion rate by 0.91. This is based on the assumption that 9% of PAF addresses are non-residential which is the standard proportion identified as non-residential in face-to-face interview surveys.

4.1 Impact of incentive value and postal survey

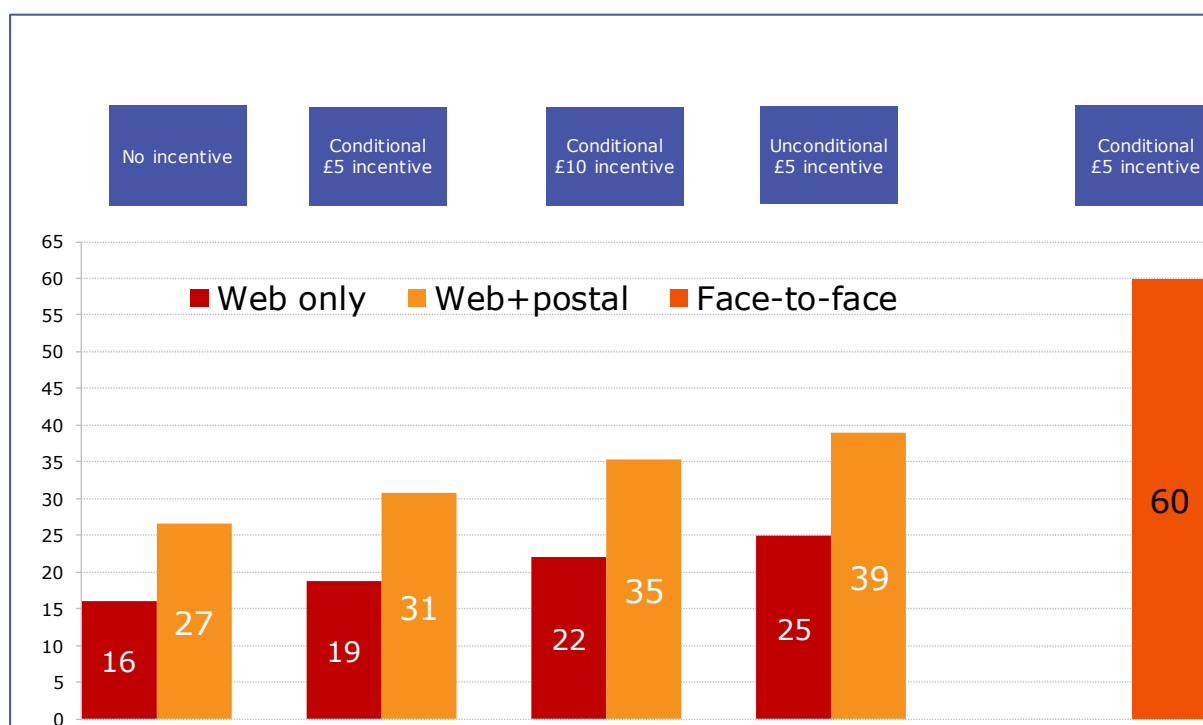
The face-to-face survey generally achieved a response rate of 60% during the 2012-13 survey which ran concurrent to the Stage 1 web survey. At Stage 1 a number of different experimental web designs were tested. Four different incentive packages were tested:

- £5 conditional voucher
- £10 conditional voucher
- £5 unconditional voucher included in the first letter
- No incentive (the control group)

In addition, the effect of including a paper questionnaire in the first reminder mailing was tested across each of these four incentive cells.

Figure 2 shows the achieved response rates for the web-only designs and the web/postal designs compared to the standard face-to-face interview design. Response rates across these eight experimental cells ranged from 16% (web only – no incentive) to 39% (web with postal survey option, unconditional £5 incentive).

Figure 2: Impact of incentive and postal reminder on response rates (Stage 1)



There are four clear conclusions:

- Response rate increases with value of the incentive (when comparing £5 to £10 offered on condition of survey completion)
- Unconditional incentives are associated with a higher response rate than conditional incentives even if the face value of the incentive is lower
- Including a paper questionnaire in the second reminder helps increase the response rate
- Although the response rate is highest when a £5 unconditional voucher is used, in cost terms this is equivalent to a £20 conditional incentive, which means that a £10 conditional voucher is a more cost effective option even if the response rate is slightly lower.

Whilst adding a paper questionnaire with the second reminder increased the response rate, the profile of the achieved sample was not substantially improved. Adding the paper questionnaire brought in more middle-aged and older people, but not young people. The age profile actually got slightly worse despite the increase in response rate. In some other respects (e.g. housing tenure and working status) the addition of a postal questionnaire option improved the sample profile, but not as much as might be expected given the significant increase in response rate.

As a result, at Stage 2, the method of distributing postal questionnaires changed. Instead of sending with the second reminder, the postal questionnaire was made available only on demand. At Stage 2, based on a conditional £10 incentive to one adult per household, the response rate was 25% web-only rising to 28% once the postal survey returns are included.

However, in 2015-16 this approach was revised further by including paper questionnaires in a targeted 50% of second reminder packs to improve both the response rate and the sample profile (based on successful implementation of this approach on another survey). By targeting paper questionnaire reminders to households sampled in deprived areas it is possible to even up the response rates between types of area through selective allocation.

4.2 Response rates achieved using an all-adults design (Stage 3)

Based on an issued sample of 1,400 addresses, and the invitation extended to all adults in the household (up to a practical maximum of four), at least one questionnaire was completed in 341 households. The estimated household response rate for the all-adults design was 27%. This compares with a similar household response rate of 28% for the standard "one adult" design.

In total, 520 questionnaires were completed either online or by post, an average of 1.5 per cooperating household. Based on information provided by cooperating households on the total number of resident adults, the average was 2.1 adults per household, or 701 resident adults across all cooperating households. The limit of four questionnaires per household reduces the maximum possible number of completed questionnaires within cooperating households to 697 and means an effective within-household response rate of 75%.

The person level response rate (incorporating both the household and within-household response rates) is less straightforward to compute. Our approach is to use the comparable face-to-face interview data to estimate the number of adults in sampled addresses and divide the number of completed questionnaires by this number. This leads to a person level response rate of 22%.

In summary, it is estimated that the response rate for the 'all adults' design is slightly lower than that of the standard design, although this is compensated for by full compliance with the sampling instructions (see Chapter 5).

5. What is the level of compliance with the random adult selection and what steps can be taken to reduce non-compliance?

Summary

- A web design does not allow for true random selection of an individual at each address, as it cannot be assumed that the sampling procedures will be applied correctly in the absence of an interviewer.
- Therefore, a quasi-random approach was used, which involved the household selecting the adult with the “last birthday” (by Stage 2 this was tweaked so that half of households were asked to select the adult with the “last birthday” and half with the “next birthday”).
- At Stage 2, one in four adults were identified as the “wrong” household respondent. Non-compliance may result from either the wrong adult completing the questionnaire to get the incentive, or human error.
- To overcome this, a design that examined extending the invitation to all adults in the household was tested. This was found to reduce the level of non-compliance, although evidence suggested that occasionally one adult completed several questionnaires in order to obtain a larger incentive.
- Although the “all adults” approach is recommended, a range of further measures are recommended to validate responses⁸.

In the face-to-face version of the Community Life Survey, at each selected address one adult aged 16+ is selected at random using a computer-generated random sampling procedure administered via the interviewer. However, true random sampling procedures are considered too complex to administer in a self-completion survey and different quasi-random alternatives have been tested.

These include:

- a) At Stage 1 one adult was selected per household by asking for the individual with the most recent birthday to complete the survey
- b) At Stage 2 this design was tweaked to make the sample more balanced, whereby half the sampled households were asked to select the person with the most recent birthday, the other half with the “next birthday”
- c) At Stage 3 an alternative design was tested whereby all adults, up to a maximum of four, were asked to complete the survey

⁸ An approach to validate responses has been developed and implemented on the 2015-16 survey – this includes a process for flagging and back checking questionable cases in the data.

The incentive for each of these designs was a £10 voucher per survey completion – so £10 per completion in versions a and b, and up to £40 for up to four individuals in version c. Incentivisation increases the risk of an individual either selecting the “wrong” individual to gain an incentive (versions a or b) or claiming a falsely high number of adults in the household to increase the household incentive (version c). Non-compliance can also result from human error.

Evidence on each of the above is discussed in this section.

5.1 Level of compliance with random selection procedures in the single adult design

With the self-completion design, the selection mechanism is outlined in the advance letter. To test compliance, the birth month was collected about each resident adult as part of the questionnaire⁹. From this it was identified that a significant number of respondents completing the survey were not the person with the right birthday, and that the sampling instructions were ignored, either deliberately or due to human error.

Overall, 74% of respondents had the right birthday, leaving 26% as ‘wrong respondents’. The non-compliance rate was higher if the household main language was not English and in households with a higher number of resident adults. This indicates that level of compliance decreases with the complexity of the task. This is a sub-optimal outcome for a random sample survey because the credibility of its data is derived from its sampling method. A degree of non-compliance with the sampling instructions does not automatically lead to biased estimates but the risk of bias is certainly greater.

It was the view of TNS-BMRB that a one in four error rate presented a risk to the validity of the survey estimates. As a result an alternative option was tested, which allowed all resident adults (up to a maximum of four) to complete the questionnaire, discussed below.

5.2 Level of compliance with random selection procedures in the all adult design

In this experiment (Stage 3) the invitation letter was changed to provide instructions that all adults aged 16+ (up to a maximum of four) should complete the online survey in order to prevent potentially incorrect within-household person selection. A postal version was made available as in the standard variant of the survey. Incentives were provided per completed questionnaire.

In summary, this method worked well. As discussed in Chapter 4, the response rate was estimated at 22%, lower than the standard design response rate of 28%, but without the problem of non-compliance with the sampling instructions.

Given the potential to earn up to £40 per household, there is a risk that one adult will complete multiple questionnaires rather than just one for additional incentives. Evidence is currently limited, but assessment of the stage 3 data suggested that this happens, but at a low level, as evidenced by an implausibly high response rate in four-adult households¹⁰ and more rapid questionnaire completion. TNS BMRB believes that the high level of non-compliance under the standard design is a more significant problem than a small number of cases where one adult completes multiple questionnaires under the ‘all adults’ design.

⁹ A simpler method was used in the edited paper variant by asking if the respondent’s birthday was the next one in the household. Combining the answer to this with information about number of eligible adults in the household, compliance can be inferred.

¹⁰ 33% person-level response compared with 19-22% across one to three person households

5.3 Recommended approach going forward

TNS BMRB recommends adopting an 'all adults' design if a self-completion survey model is used as this design effectively overcomes the problem of non-random within-household sampling without bringing substantial additional drawbacks.

However, it should be combined with stronger deterrents than were used for the Stage 3 test. Robust prevention measures were developed and put in place in 2015-16. The current approach includes:

- A declaration screen at the end of the survey that asks respondents to confirm that they have answered all questions honestly and completed the survey only about themselves.
- A process for flagging questionable cases in the data. This is based on an algorithm which takes into account factors, such as the stated number of adults in the household versus the number of complete interviews received; the length of time taken to complete the interview; and comparisons between the responses given to questions relating to the household by each individual
- Asking respondents to provide a phone number to enable back-checks
- Removing cases, identified as invalid through this process, from the data (likely to be 5-10% of the total)

6. What is the profile of a web sample compared with a face-to-face sample?

Summary

- As the response rate to the web/postal survey is significantly lower than the response rate to the face-to-face survey, the risk of non-response bias is much greater.
- Compared with the face-to-face survey, the self-completion sample is more likely to comprise everyday internet users, high earners, owner-occupiers, native English speakers and the more highly educated.
- The paper questionnaire option helps bring in non-internet users, although take-up of these questionnaires on-request is low and there is minimal improvement in sample profile, unless included with a targeted subset of reminder packs.

The interview survey achieves a response rate of around 60% while the web/postal survey achieves a response rate of around 28%. Therefore, there is a substantially greater risk of non-response bias in the web survey when compared with the interview survey.

This chapter provides an overview of the unweighted sample profile for the web survey in comparison to the face-to-face survey. This highlights the groups which are over-represented and under-represented in comparison to the interview survey.

6.1 Composition of the self-completion sample

The Stage 1 and Stage 2 tests generally reached the same conclusions about the composition of the web sample. In comparison with the interview survey, the online/postal sample contains:

- a higher concentration of those who use the internet everyday
- a lower concentration of younger people aged 16-24 and a higher concentration of middle aged/older people aged 45-74
- a lower concentration of social renters and a higher concentration of home owners
- a lower concentration of those who have lived at their address for less than a year
- a lower concentration of those whose native language is not English
- a higher concentration of those who earn more
- a higher concentration of people who are degree educated

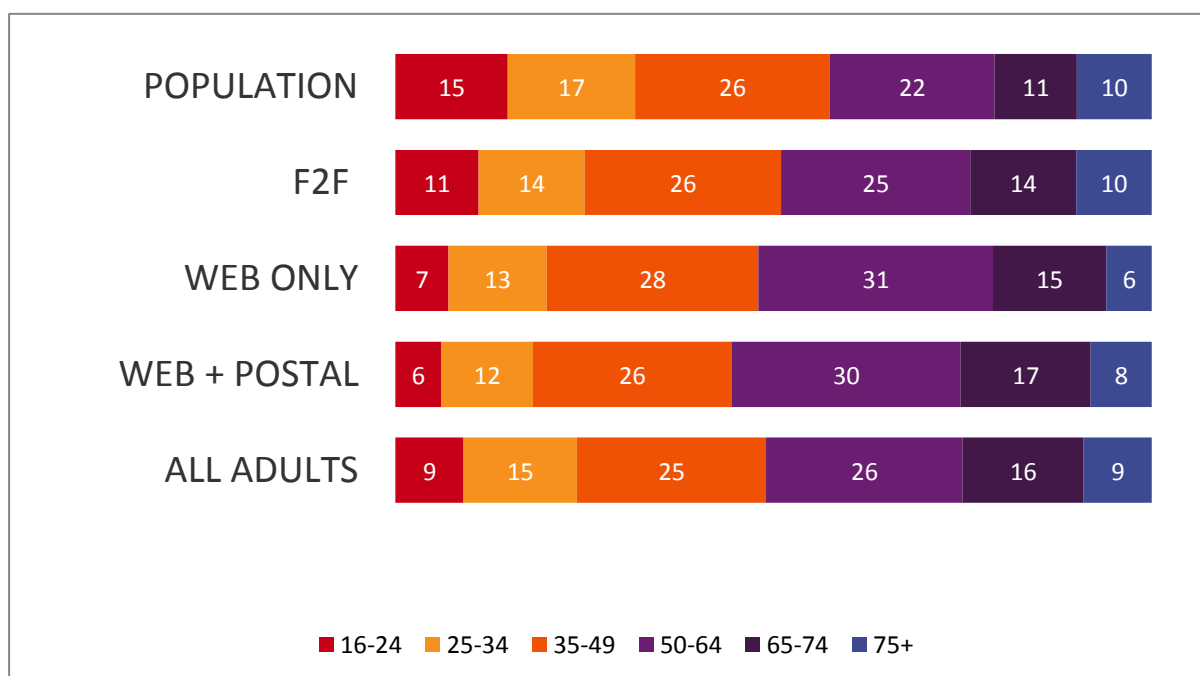
All of these findings are in line with expectations. A skew towards heavy internet users is expected given that the primary mode is online; income indicators could be an indicator of lower levels of online access; a longer length of residence is associated with strength of connection to the community (the headline topic for this survey); and non-native English may act a barrier to accessing the online survey.

Of course the face-to-face survey, even with a response rate of 60%, does not perfectly reflect the population. As an example, Figure 1.3 shows how the age profile of the sample differed by test design in the 2013-14 surveys¹¹.

This shows that across all survey designs there is a bias towards the 50-74 age groups, a bias which is further accentuated across the different online designs. The age profile for the 2013-14 survey (Stage 2) is shown for the online-only sample and the combined online/postal sample. This shows that adding the postal survey makes very little difference in terms of improving the age profile.

Figure 3 also shows the age profile for the all-adults design (Stage 3) based on data from the 2014-15 online/postal survey, which adopted the 'all adult design'. This indicates an improved representation of younger people compared with the standard web survey design.

Figure 3: Age profile of the different survey designs



Base sizes: F2F (n=5,105); Web only (n=9,332); Web/postal (n=10,157); all adults 2014-15 (n=2323)

A clear recommendation is to weight the online/postal sample to compensate for any measurable sample bias. However, sample bias can remain even after weighting because sample bias is more than just demographic bias; bias may also be associated with behaviour and attitudes. This issue is addressed in Chapters 8 to 10.

¹¹ All data in the chart has been weighted to correct for within-household sampling probabilities only

7. What is the quality of the data using online designs?

Summary

- In general, there are no glaring concerns about the quality of the online data compared with face-to-face data, based on quality metrics such as time taken to complete and measures to detect survey “satisficing” (i.e. short-cutting tactics)
- However, there are indications of lower levels of engagement with the online survey, and evidence of a small risk that in some cases one adult completes several questionnaires.
- Different presentations of “don’t know”, “prefer not to say” and other spontaneous codes can affect their level of use in the online survey. While the presentation can be manipulated to more closely mirror the “face-to-face” survey there is no evidence that this produces results which are closer to the “true” opinions of respondents.

Beyond sample quality, data quality is an important consideration when choosing between survey designs. Face-to-face interviews are regarded as a high quality interview mode in this respect, as the interviewer can help ensure engagement throughout the interview and can pace the survey so that the respondent does not complete it too quickly, giving their full attention to each question. There is a concern that online/postal surveys on the other hand are more susceptible to survey “satisficing” which is defined as a tendency by survey respondents to take short cuts rather than giving a fully considered response.

As discussed in Chapter 6, there are a number of noticeable differences in findings when the online/postal and the face-to-face survey are compared, even after the sample is aligned by key demographics. However, it is not always clear which is closer to the “true” answer. This is because interviewer-based surveys can be more prone to social desirability bias – that is the tendency of survey respondents to reply in a manner that they believe will be viewed favourably by others. For example this might include a tendency to over-state more socially desirable behaviours (volunteering, charitable giving etc.) and to under-state less socially desirable behaviours (e.g. feeling lonely).

Data quality is difficult to measure in an empirical way. However, we used a number of metrics to provide an indication of data quality. These are summarised Table 4.

This shows that:

- when comparing the online/postal survey with the face-to-face survey, in general there is no evidence to suggest any level of concern about reduced data quality/increased levels of survey satisficing.

- However, there are indications of lower levels of engagement with the online/postal survey compared with the face-to-face survey, as evidenced by higher levels of drop-outs (individuals who exit the survey before coming to the end of the questionnaire) and reduced willingness to be re-contacted by TNS BMRB for further research. Engagement and completion rate might be raised by considering the questionnaire order (for example by placing easier to complete and more "interesting" questions nearer the start and more difficult and/or sensitive questions nearer the end).
- If the presentation of missing responses (such as don't know, prefer not to say) is made more similar to the face-to-face survey, the levels of missing responses are more aligned. However, it is more difficult to state which mode/presentation offers answers which are closer to the "truth" - for example it could be argued that making "don't know" responses more difficult to access forces people who genuinely do not have an opinion into an answer option which doesn't fully reflect their view.

Table 4: A summary of data quality metrics for the different online survey models

Measure	Possible indication of...	Standard web/postal survey (Stage 2)	All adults design (Stage 3)
Length of time taken to complete interview	Level of effort (a faster completion might indicate short-cutting or fraudulent completion)	Same length as interview survey	Evidence that mean interview length is reduced in households self-classified as four person households. Possible evidence of a small degree of fraudulent completion.
Number of items selected at multi-coded questions	Level of effort put into considering all options at multi-response questions	Similar to face-to-face survey	Evidence that slightly lower number of items selected when four questionnaires were returned in a household – possible evidence of a small degree of fraudulent completion
Differentiation of questions that use the same response scale (e.g. the tendency to say “Agree strongly” across all items in a battery which use the same 5 point agree/disagree scale).	Non-differentiation across batteries (sometimes referred to as “flat-lining”) can be an indicator of reduced cognitive effort.	Similar to face-to-face survey	Not covered
Bias towards top answers when a long list of responses is presented.	Also known as “primacy effects” this can be an indication of reduced cognitive effort (selecting those from the top and not reading to the bottom of the whole list)	Similar to face-to-face survey	Not covered
Level of drop-out (i.e. started the survey but did not finish)	Can be indication of dissatisfaction/boredom with the survey; plus leads to a higher proportion of survey returns which are partial and incomplete	Higher than face-to-face survey (11% vs 1%). In addition to missing data arising from web drop-outs the postal version of the questionnaire is abridged and therefore does not offer full questionnaire coverage.	Similar to Standard Stage 2 design (9% dropout rate)

UNCLASSIFIED

<p>Proportion of “don’t know” and “prefer not to say” responses</p>	<p>This can be attributed to two factors: it is sometimes thought that a higher level of non-responses, such as these, are evidence of survey satisficing; however, differing levels may also reflect different presentation of such responses and how prominently they are displayed to respondents.</p>	<p>Compared with f2f survey, higher at Stage 1 when these responses were displayed on screen. However, at Stage 2 - when a change was made to make these responses less prominent in the web survey – rates were similar to F2F survey.</p>	<p>Not covered</p>
<p>Willingness to be re-contacted by TNS BMRB for further research</p>	<p>Lower consent rates can be an indication of dissatisfaction, low engagement or perceived burden; it could also be an indicator of fraudulent completion (if false information has been provided it could be argued that willingness to be re-contacted might be lower)</p>	<p>Compared with an 83% consent rate in the f2f survey, this is much lower in self-completion designs: at Stage 2 42% of online respondents and 52% of postal respondents were willing to be re-contacted (43% overall).</p>	<p>Overall consent rate (41%) was similar to the standard design. However, the consent rate fell with number of questionnaires returned from 55% among households where one questionnaire was returned to 32% among households where four questionnaires were returned. Consistent with other findings this may indicate a level of fraudulent completion.</p>

8. Does the web/postal survey give similar estimates to the face-to-face survey both before and after demographic weighting?

Summary

- Before weighting, around two-thirds of the web sample estimates are significantly different from the interview sample estimates; around 20% of these differences are substantial (at least five percentage points differences).
- After weighting to correct for demographic profile differences, the results are only slightly more aligned, which indicates that data collection mode and/or residual sample bias influence the results to the online survey. The relative contribution of these two effects was tested in Stage 4 (see Chapter 9).

8.1 How aligned are the result before demographic weighting?

To answer this question, the online/postal data at Stage 2 together with the interview data from the comparable time period (April 2013-March 2014) were weighted to compensate for differences in sampling probability but not for sample profile differences.

As the sample sizes are large (5,105 interviews and 10,205 web/postal questionnaires) this means that relatively small differences (for example 1.5 to 2.5%) will be labelled as statistically significant.

Across 157 categorical variables, 68% of the differences were statistically significant at the 5% significance level. However, most differences were small:

- 46% were less than two percentage points
- 33% were greater than two but less than five percentage points
- 17% were greater than five but less than ten percentage points
- 4% were at least ten percentage points

8.2 Are the results more aligned if the web survey is weighted to match the interview survey on key demographic measures?

To answer this, the self-completion sample was weighted to match the interview sample on twelve different demographic characteristics: (i) gender, (ii) age group, (iii) region, (iv) working status, (v) presence of children, (vi) number of adults in the household, (vii) ethnic group, (viii) religion, (ix) highest level of qualification, (x) housing tenure, (xi) length of residence in the neighbourhood, and (xii) level of internet use.

After weighting, we would expect some narrowing of the differences in the results if sample bias is responsible for all or part of the differences between the two sets of results.

Across the same 157 categorical variables cited above, 58% of the differences were now statistically significant at the 5% significance level, only slightly lower than 68% without any weighting. The distribution of size of differences was only slightly more skewed to the smaller end compared with when no demographic weighting was applied:

- 55% were less than two percentage points (compared with 46% before weighting)
- 24% were greater than two but less than five percentage points (compared with 33% before weighting)
- 21% were greater than five percentage points (same as before weighting)

In summary, weighting was only marginally effective at narrowing the differences between the two sources of data. On average, the expected difference *after* weighting is approximately five sixths of the difference *before* weighting.

In order to summarise the nature and direction of differences in survey findings between the two designs, Table 5 below provides an illustration of the differences between the two survey designs (after weighting) for some of the key Community Life measures.

Overall, there is no clear direction or pattern of differences. It would seem that:

- Self-completion respondents are more likely to report participation in activities such as volunteering and giving to charity which may indicate a link between propensity to do self-completion surveys and willingness to help others
- Self-completion respondents are less likely to: feel that they belong strongly to their area; that they can influence decisions; interact with their neighbours; be very satisfied with their local area; and to consider their local area to be cohesive. They are also less likely to “never” feel lonely. These differences may be linked to a tendency in interviewer-facing surveys to over-state socially desirable characteristics (see Chapter 9 for further discussion about the mode differences in relation to questions which might be susceptible to this type of bias).

Table 5: Largest differences between interview and self-completion questionnaire estimates (after weighting)

	Face-to-face interview	Web/postal survey (Stage 2)
	n=5,105	n=10,205
	%	%
Any formal volunteering in the last 12 months	41	45
Any informal volunteering in the last 12 months	64	57
Given to charity in the last 4 weeks	75	82
Agree that “I can influence decisions made about the local area”	34	26
Feel that you belong to your local neighbourhood either very or fairly strongly	70	58
Very satisfied with area as a place to live	43	36
Definitely agree that local area is a place where people from different backgrounds get on well together	25	18
Agree that many of the people in your	39	48

neighbourhood can be trusted,		
Definitely agree that: I borrow things and exchange favours with my neighbours	19	13
How often do you feel lonely: never	30	23

The fact that the face-to-face and online/postal models produced significantly different results on key measures even when the samples were aligned in terms of key demographics left a key unanswered question: were these residual differences in results due to i) the different mode of data collection (i.e. people giving different answers due to the different presentation of the questions and/or the presence/absence of an interviewer) or ii) residual systematic differences between the samples.

This question was addressed by Stage 4 and is covered by the next chapter.

9. If the online/interview results are still not aligned after weighting what is the main reason for this: residual sample bias or data collection mode?

Summary

- A further set of experiments was devised to estimate the relative contribution of data collection mode on the one hand and sampling/fieldwork methods on the other, to explain the differences in results between the web/postal and interview designs.
- The evidence suggests that the difference in data collection mode is responsible for the bulk of the mismatch observed between the results.
- The wider evidence suggests that a self-completion mode is usually a better tool for questions which are sensitive or subject to social desirability bias. As these types of questions make up a large proportion of the Community Life survey, it is likely that the online survey will produce more accurate results than the interview survey for most (but not all) survey questions.

At the end of Stage 2 it was concluded that demographic profile differences between the two samples did not account for the bulk of the difference in results between the self-completion and interview modes. Even after the results were aligned on core demographics, substantial and significant differences in results persisted.

The question that arose was this: was this residual 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 sample recruitment methods (despite significant efforts to statistically align the two post-recruitment)? To answer this question it was vital to disentangle sample and mode effects in order to determine which has the strongest influence on the results.

9.1 What are sample effects and mode effects?

- **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. If there *are* sample effects, then it is reasonable to assume that the interview-based estimate is the more accurate of the two and this would throw into question the long-term validity of the online approach.

Mode effects are less of a threat to the validity of an online approach. This is because it cannot be assumed that the interview data collection mode is a more reliable measurement tool than the self-completion questionnaire. If there *are* mode effects, we may make a judgment as to which is the more accurate estimate but this is not guided by such objective measures as the response rate; it is rather more subjective than that and may differ from question to question (discussed further in section 9.3).

9.2 How did we attempt to disentangle the relative effects of sample and mode?

This section summarises 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 (see Annex's).

Stage 4 involved an additional experiment which was based 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, 1,578 did so. This represents a 37% response rate among invited cases

This new dataset was then used in a study which had two parts as follows:

- *Estimating the contribution of sample effects on the results:* this was achieved by comparing the results of the re-contacted online sample with the results of the standard online sample, which ran concurrently. Differential non-response to the online re-contact survey was corrected by weighting¹². Because the measurement tool is the same, any significant differences in results is likely to be caused by systematic differences between the samples.
- *Estimating the contribution of mode effects on the results:* this was achieved by comparing the responses of the online re-contact sample with the responses of the same sample when they completed the original face-to-face interview. As the sample is the same, any differences in results are likely to be caused by differences in the mode of interview¹³.

The sample effects and mode effects were then plotted against the observed difference between the interview and online survey to see which had the greater explanatory power.

The evidence from the more detailed analysis 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. These findings raise the obvious question: is the online/paper questionnaire survey reliable?

¹² A response propensity model was used to calculate these weights

¹³ As the interview survey was completed earlier (around ten months had elapsed between the two surveys) the effects of time were controlled for by using a statistical technique called "difference-in difference" to remove the effects of time and to isolate the effects of mode.

9.3 If data collection mode is the principal cause of differences between online and face-to-face survey designs does the web survey produce reliable findings?

If sample effects *are* minimal then essentially the answer to this question is 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 wider literature tends to produce 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. 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 many 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. For questions that are sensitive to social desirability bias 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 mid-categories 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. Having said this, the evidence presented in Chapter 7 does not indicate that this is a particular problem for this survey although there may be isolated questions which are more sensitive to this.

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. Based on the wider literature it is *likely* that the *Community Life* self-completion questionnaire will 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 sample bias to some estimates.

On balance, TNS BMRB expects the *net* error to be lower with the online/paper questionnaire survey¹⁴ than with the interview survey but this assertion is very much an informed opinion rather than a fact.

¹⁴ The much lower cost of the online/postal survey compared to an interview survey means that the effective sample size can be much larger for a fixed budget. This will decrease sampling variance, reducing net error.

10. Conclusions and recommendations

The online/postal methodology continues to evolve and develop over time as additional methods are tested and new evidence is collected. However at this particular point in time, the online/postal data collection model TNS BMRB recommends, should it be adopted as the primary source of data collection, is:

- A random probability stratified sample of addresses drawn from the Postcode Address File (PAF) with some minor variations in sample fractions between strata to help optimise the responding sample profile.
- Survey invitations issued by letter; at each address all adults (up to a maximum of four) invited to take part
- Up to four mailings sent to each address to maximise response
- Postal questionnaire available as an alternative to ensure inclusion where the household lacks internet access
- Targeted postal questionnaires included with 50% second reminder packs (currently being tested on the 2015-16 Community Life Survey and introduced based on strong evidence from other unpublished studies)
- £10 incentive to each person completing the survey to maximise response
- Strong preventative and validation measures in place to minimise the risk of one adult completing multiple surveys and to identify such cases in the data so they can be removed from the data.

Throughout the development stage, a number of improvements have been made to the design to improve the reliability of the estimates – most notably the adoption of the ‘all adult’ design, to overcome the problem of non-random within-household sampling.

The evidence presented in this report also provides some reassurances about the differences observed between the face-to-face and online/postal survey estimates. In particular, it 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. TNS BMRB expects the *net* error to be lower with the online/paper questionnaire survey than with the interview survey, but this assertion is still more of an informed opinion than a fact (see Annex’s for further detail)

It is also worth considering the original aim of testing an alternative approach to a face-to-face design – to explore cost effective methods for future survey years. The cost per completed online/postal questionnaire is lower, which means that much larger samples can be obtained for a

fixed budget than would be possible with a face-to-face methodology, allowing for more detailed sub-group analysis.

Adopting a solely online/postal survey going forward does however have some drawbacks:

- It would represent a break in time series as the differences observed between modes mean that some of the data are not closely compatible.
- Despite the evidence presented in this report, some data users may find the lower response rate difficult to accept as it is traditionally regarded as a marker of quality¹⁵.

¹⁵ It is worth noting that telephone Random Digit Dialling (RDD) methods *have* been used to produce Official Statistics with lower levels of population coverage and similar responses rates

Appendix A: What weighting strategy is recommended for an online/postal Community Life Survey?

The weighting plan adopted for the 2014-15 survey does not require data from the face-to-face interview survey before it can be deployed and represents our current recommendation for weighting the online/paper survey data.

The weighting plan has three steps.

Step 1 is to use a classification tree algorithm to combine sample strata with similar questionnaire completion rates into a small number of classes. The class questionnaire completion rate can be used in the calculation of base weight w_1 .

$$w_1 = 1/((n/N)*p_i)$$

n = number of serial numbers issued for the Community Life Survey (= number of issued addresses multiplied by 4)

N = total number of addresses in England multiplied by 4

p_i = % of issued serial numbers in stratum group i that resulted in a completed questionnaire

Stratum-specific versions of n and N should be used if sample fractions vary between strata.

In step 2, the base weight w_1 should be further modified into weight w_2 to compensate for sampling a maximum of four individuals at each sampled address:

$$w_2 = w_1 * (\max((N_e/4), 1))$$

N_e = number of resident adults at address e

Base weight w_2 should be used as a start point for calibrating the respondent sample to population totals. If a household weight is required (as has been the case up till now), base weight w_1 should be used as a start point for calibrating the residents of sampled households to population totals. However, the latter cannot be computed if the respondent completed the questionnaire on paper so is available only for online participants.

In step 3, we recommend that a generalised linear regression method is used to calibrate the respondent sample and that the following variables are used to construct population totals:

1. Gender by five year age group

2. Region
3. Housing tenure
4. Highest qualification crossed by age group (25-64s only)

These four variables have been selected based on an analysis of the base-weighted 2014-15 sample profile against a wide set of population benchmarks. Population totals should be drawn from ONS mid-year population estimates (1 & 2) or from robust large sample government surveys such as the Annual Population Survey or the Crime Survey of England & Wales (3 & 4).

Although steps 1 to 3 represent our current recommendation, there is some flexibility over the choices for steps 1 and 3.

At step 1, a respondent-specific version of p_i may be estimated using regression methods instead of a classification tree algorithm and with a selection of other variables as predictors rather than sample strata. However, if regression methods are employed, we recommend dividing the respondent set into five equal-sized classes, each defined by a specified range of respondent-specific p_i values. The class mean of p is then used as p_i in the step 1 equation. Using this value rather than the respondent-specific estimate of p_i guards against model-misspecification whereby unrealistically high (or low) questionnaire completion probabilities are estimated for a small number of cases, lowering the effective sample size with no compensatory reduction in bias.

At step 3, other variables may be used in the calibration matrix but care should be taken to avoid classes with very small respondent sample sizes (<50) and to avoid variables that are (i) sensitive to the mode of data collection or (ii) reliant on data series that are not guaranteed for at least the following three years.

Because there are numerous mode-specific measurement effects within the Community Life survey, we do not recommend using the parallel interview survey data (2012-16) to develop a general 'propensity score' model that could be incorporated within the weighting plan¹⁶. In theory, this might further reduce the risk of sample-related bias but, in practice, the different measurement characteristics of online/paper self-completion questionnaires and in-person interviews mean that most variables are 'off limits'. In any case, our analysis suggests that the different sampling and fieldwork methods of the two surveys have little or no impact on the survey estimates. Consequently, a propensity-score adjustment to the weights might lower the effective sample size with no compensatory reduction in bias.

¹⁶ The propensity score would be equal to the relative odds of responding to the online/paper questionnaire against the interview survey benchmark (as estimated from the data collected during the period of parallel running).