Comme une lettre a la Poste : l'aleatoire en ligne

Yves Fradier[1] & Joel Williams[2]

[1] *Kantar Public France, 3 avenue Pierre Masse, 75014 Paris, yves.fradier@kantarpublic.com*
[2] *Kantar Public UK, 222 Grays Inn Road, London, WC1X 8HB, joel.williams@kantarpublic.com*

Résumé. Il y a une forte demande au Royaume-Uni, et dans les pays comparables, pour une méthode financièrement abordable d'enquête en population générale utilisant les techniques d'échantillonnage aléatoire. Dans ce texte, nous décrivons une méthode alternative, déjà utilisée aux États-Unis mais nouvelle au Royaume-Uni : le Sondage en Ligne Basé sur les Adresses. Ce texte couvre tous les aspects de la méthode, répondant aux questions sur le niveau de couverture, la qualité des données, les taux de réponse probables et les biais potentiels dans les estimations.

Mots-clés : Grandes Enquêtes, Echantillon Aléatoire, Auto-administré, CAWI

Abstract. There is a strong demand in the UK for an affordable method of surveying the general population that still employs random sampling techniques. In this paper, we describe an alternative with antecedents in the US but new to the UK: ABOS – address-based online surveying. This paper covers all aspects of the method, seeking to answer questions about its coverage levels, its data quality, likely response rates, and the potential for bias in substantive estimates.

Keywords : Large Scale Surveys, Random Sample, Self-administered, CAWI

1. Motivation for a new general population survey model

UK business and government are great consumers of information about their populations of interest. Traditionally much of this information has been derived from questionnaire surveys and, although other data forms are now becoming influential, this is still the dominant method of information gathering in the UK.

Since the late 1940s the gold standard for sample quality has been to use random sampling methods with in-person data collection. However, although response rates remain fairly high (45-75%), in-person data collection is expensive and requires long timeframes. Consequently, there has always been a demand for more affordable - and more agile - alternatives that nevertheless employ random sampling techniques.

From the early 1970s until relatively recently, RDD ('random digit dial') telephone surveying was the primary alternative. However, the modern requirement to sample mobile numbers as well as landline numbers has made the method more expensive than it was. Contact and cooperation rates have also dropped in recent years, adding to the method's cost and taking away from its quality.

Since the early 2000s, online panel data collection has been used for 'volume' research in the commercial sector. This method has much lower costs and greater timeliness compared to other data collection methods. However, these panels usually exclude everyone who is offline and random sampling methods are rarely used in their construction. This makes it hard to be *generally* confident about the accuracy and precision of their estimates, despite the occasional success.

An alternative that combines the best features of random sampling (sample unbiasedness with regard to *all* survey variables, not just those for which population parameters are available) with the convenience and (relatively) low cost of online panels would be a very popular option.

A random sample online panel has potential - several exist abroad plus one modest-sized one in the

UK[1] – but the substantial set-up costs have discouraged most research agencies from building their own. At Kantar Public (formerly TNS BMRB in the UK), we have instead developed a non-panel method – address-based online surveying (or 'ABOS') – as a way to meet demand for low-cost random sample surveying in the UK. It is not strictly new, having its antecedents in 'push-to-web' US studies[2], but its specific implementation in the UK *is* new.

2. Introduction to address-based online surveying (ABOS)

The core ABOS design is a simple one: a stratified random sample of addresses is drawn from the Royal Mail's postcode address file and an invitation letter is sent to 'the residents' of each one, containing username(s) and password(s) plus the url of the survey website. Respondents can log on using this information and complete the survey as they might any other online survey. Once the questionnaire is complete, the specific username and password cannot be used again, ensuring data confidentiality from others with access to this information.

However, this core design must be augmented with several other features to make it workable. Over the course of the last three to four years, we have carried out experiments and gathered other evidence to help us understand which features work best.

This paper is intended as an introduction to the method and as a summary of what we know and what we do not know at the time of writing (Autumn 2016). Over the course of this paper, we intend to answer a series of (hypothetical!) questions about the ABOS method, presenting evidence where we have it and giving a current viewpoint about what constitutes best practice. We will also note any plans for the future, including features that are likely be tested via already commissioned studies. The questions are:

3. Q1: If the sample is of addresses, how do you convert this into a sample of individuals?

The postcode address file (PAF) is thought to provide a highly comprehensive link to the general population of adult individuals living in residential households but the ABOS method has no interviewer to facilitate that link. Instead, we must rely on one or more residents at the address to do this job for us.

The first thing to note is the fact that a small fraction (probably 2-3% in England) of addresses contains more than one household and there is no way to 'sample' one in a controlled manner. Whoever picks up the letter effectively self-selects their household into the sample. While a weakness, this departure from random sampling is – in our view - small enough to be accommodated in most cases.

Accepting this uncontrolled conversion from a sample of addresses into a sample of households, the question is how to get from here to a sample of individuals while respecting random sampling principles. As part of a test of European Social Survey methods, Park and Humphrey (2014) used a variant of the ABOS method in which the first adult to read the letter was asked to log on to the survey website and complete a short questionnaire on household composition. At the end of this, the survey software randomly selected one resident adult and requested that the initial respondent facilitate a transfer to this selected person (if different). Subsequent analysis suggested that, in many households, this selection stage was ignored and that the goal of a random within-household sample was not obtained.

In our early tests of an ABOS version of the Cabinet Office *Community Life* survey, we instead tested the quasi-random 'birthday' selection method in which the adult resident with the last – or next – birthday[3] is asked to complete the questionnaire. This is not a true random sampling method but, if implemented accurately, should provide functional equivalence. Its theoretical advantage

[1] This has been developed by NatCen Social Research.

[2] See Messer (2012) for a summary.

[3] Best practice is to allocate a random half of the addresses to 'last' birthday and the other half to 'next' birthday, in an effort to minimise 'season-of-birth' sample effects that affect some survey variables.

over the ESS selection method is its simplicity: it does not take much thought to work out who has the last/next birthday and there is no two-stage responding process. However, simplicity does not guarantee compliance. Do households – or more accurately the individuals picking up the letters – bother with this part?

To test compliance, we included a question on month of birth of each adult resident in the household. We hypothesised that those ignoring the 'birthday' sampling instruction would nevertheless provide this data (where known). We could then use the date of questionnaire completion to work out which individual *should* have been selected, or at least identify the majority of 'wrong' respondents. If we had sent no sampling instruction - or alternatively every household ignored the sampling instruction – we would expect the 'right' respondent every n^{th} time where n is the number of eligible individuals in the household. This is the baseline against which to measure the success or otherwise of the sampling instruction. We concluded from our test that the success rate was not a great deal above baseline and that this success rate was lower for larger households than for smaller ones. Overall, c25% of respondents were identified as 'wrong' respondents.

There are a number of possible ways forward from this. Option 1 is to use the birthday selection method but accept a significant level of non-compliance. Option 2 is to use the birthday selection method but to identify and exclude 'wrong' respondents from the analysis base, effectively converting the problem from one of sampling error to one of non-response error. Option 3 is to ask *all* eligible individuals in the household to participate in the survey, eliminating the flawed within-household sampling process altogether but introducing other challenges in its place.

Option 1 has been used for one of our most recent ABOS studies, motivated by the fact that the *Community Life* data did not suggest any systematic difference between 'right' and 'wrong' respondents with respect to demographic profile or to the substantive variables in that survey. However, not all research commissioners will think this sufficient evidence and some will find the non-compliance with the sampling instructions fundamentally problematic. Option 2 – the exclusion of 'wrong' respondents – might be more palatable but would lead to a reduction of c25% in the analytical base, and a corresponding increase in costs to maintain the intended sample size. Furthermore, it would produce a responding sample biased towards one person households where the probability of a 'wrong' respondent is zero.

In the end, we recommended to the Cabinet Office a test of option 3 (Williams (2014)), in which *all* adult residents are asked to complete the questionnaire. This is achieved by supplying four sets of login details (with more available on request) which can be used in any sequence. However, although this solves the 'wrong respondent' problem, it introduces others.

Although there are several *general* risks when surveying multiple individuals in the same household, our main concern with option 3 was the risk that one individual would complete the questionnaire multiple times, especially if each completion was incentivised. Of course, the incentive could be dropped or limited to one per household - removing (most of) the motivation for proxy completion - but doing so would replace the risk of proxy completion with a risk of lower response motivation. Ideally, we would *identify and exclude* proxy completions rather then remove the incentive to respond. The question is how? This forms part of a larger question about how to verify ABOS data so that it is of sufficient quality for users (see Q2).

One final option – but one untested by us - is to use a hybrid selection mechanism. In Sport England's new *Active Lives* survey[4] – which uses a variant of the ABOS method – *any two* residents are invited to take part, a compromise between random sampling principles and the desire to limit the maximum incentive available per household (and thus the motivation for proxy completion). This method only departs from true random sampling in households with three or more adults (c18% of households in England) which might be a minor enough exception for research commissioners to accept. However, households with three or more adults are distinctive in numerous ways. With this method there is a small risk that real differences between these types of

[4] See www.sportengland.org/research/about-our-research/active-lives-survey for more details.

household and other types of household are confounded with differences due to the sampling mechanism.

In sum, there is no perfect way of converting a sample of addresses to a sample of individuals, only a set of imperfect ways. For most of our newer ABOS studies, we have used option 3 (all adults can take part) combined with algorithmic weeding out of probable proxy completions. However, we consider this a live topic for research as the empirical evidence for each method is still rather thin.

4. Q2: How do you verify that the data is from the sampled individual(s)?

With interview-based surveys we have confidence that almost all the data is collected in a controlled manner and from the right individual. Interviewers ensure that the survey protocol is followed and they themselves are monitored by survey operations staff to minimise the risk of departures from protocol and to catch the occasional cheat.

With ABOS and most other self-completion survey methods, there is no interviewer to do this work so it must be accomplished via other methods. For a start, respondents should be made aware that we expect them to supply data in good faith. This can be partly achieved through (e.g.) asking the respondent to confirm the conditions of questionnaire completion (non-proxy, in some privacy), asking him/her to 'sign' it as their own work, and by asking for additional contact details to facilitate post-fieldwork verification checks. All these methods make it clear to the respondent that we take data quality seriously and this in itself may deter some proxy or careless completions of the questionnaire.

However, these design features ought to be combined with a programme of post-fieldwork verification. This can take two forms. The first is to re-contact respondents by telephone to check that the named person completed the questionnaire and (if so) to confirm a few characteristics that ought to be known only to the individual. The second form of verification is to use an algorithm to identify poor data *post hoc*. The implicit assumption underpinning the use of this algorithm is that proxy data will usually be of poor enough quality to be detectable – and discarded if desired.

As it stands, the first form of verification has been implemented only once for one ABOS study we know of (*Community Life*). No problems were found on that occasion but the low re-contact agreement rate - typical of self-completion surveys – is a major limitation to this form of verification. Furthermore, for cost reasons, this kind of verification can only be applied to a sample of cases so it is a far from *sufficient* method of verification.

Consequently, we are largely reliant on the second form of verification – the bad data detection algorithm – and must do so without strong evidence of its efficacy for ABOS studies. Instead, the algorithm has been built based upon a more generic understanding of measurement error in a self-completion context.

Our algorithm varies slightly between different ABOS studies but at its core it utilises a variety of classic indicators of proxy/careless completion and if a small number of these indicators light up, the case is removed from the data file. This seems to us a proportionate approach to data verification given that no one indicator is *certainly* a sign of invalid data. For the record, this approach led us to remove less than 5% of cases from the 2015-16 *Community Life* survey, a rate that may prove to be indicative for ABOS studies in general and a rate that seems low enough for us to be largely confident of the data's veracity.

Although each ABOS questionnaire is different, there are a number of indicators that we use across all studies. These include (i) inconsistencies in household data when multiple completed questionnaires have been received from the same household, (ii) use of the same email address by multiple respondents when providing the necessary details to receive the e-incentive, (iii) suspiciously short completion times, (iv) only a few minutes between one questionnaire being completed and another being started within the same household, and (v) excessive missing data rates.

We also pay special attention to households where the maximum number of questionnaires has been completed (four in *Community Life*). From the development work, we know these questionnaires tend to have been completed more quickly than average (a median of 28 minutes v 38 minutes in one *Community Life* test) and that respondents also tend to select fewer than average items from multiple response lists. However, the missing data rates are average, as is the length of open-ended text, and there is no additional primacy effect that we can detect. For the most part, these completed questionnaires do not look particularly different from others so we take the view that four completions from a single household does not *necessarily* mean proxy/careless completion in order to obtain a larger incentive. Nevertheless, to be on the safe side we tend to discard these cases based on fewer lit indicators than are required to discard other cases.

It is an open question whether this combination of 'nudging' respondents to complete the questionnaire truthfully (and with care) together with an algorithmic method of post-fieldwork case-removal is *sufficient*, even if it is *proportionate* and this is certainly an area for further development. Nevertheless, the ABOS method is intended as a low-cost way of obtaining a random sample of the general population; some level of proportionality – some level of compromise - is necessary to ensure that the cost of data verification does not transform ABOS from a low-cost to a high-cost survey model.

5. Q3: How do you cover 'offline' individuals?

According to weighted Crime Survey of England & Wales data from 2012-15[5], 17% of the adult population in GB has either never used the internet or uses it so infrequently that they are effectively not covered by an online survey method. This group is shrinking slowly over time, but more due to its demographic decline than due to a change in behaviour among the group. This group is particularly distinctive with respect to birth cohort and educational level, tending to be older and, controlling for age, disproportionately likely to have no academic qualifications.

Although the size of the offline subpopulation is shrinking, excluding a highly distinctive 17% of the adult population is not acceptable for surveys that aspire to the status of official statistics (unless the survey topic is exclusively concerned with online behaviour). Consequently, ABOS studies need to cover offline subpopulations using an offline data collection mode. We have experimented with offering paper questionnaires and telephone interviews on request and have also used paper questionnaires more directly, including one or more in some reminder packs (see Q4).

A different approach – to be tested at scale in early 2017 - is to use a dual sample design in which a standard ABOS study is combined with a smaller interview study in which sampled households are screened for individuals that are either (i) aged 70+ or (ii) have not used the internet in the last year[6]. Our analysis suggests internet-using people in their 70s and 80s are not particularly well covered by the ABOS method, hence their inclusion in both samples. The next age group down – the internet-using 60-69s - is covered as well as any other age group.

Given the additional costs of a separate interview study, we recommend that researchers consider under-sampling the target population - and then applying larger than average design weights to the data – rather than seeking absolute proportionality. Either PAF-based in-person interviewing or dual-frame RDD telephone interviewing (if suitable) might be used for the 'offline' sample. The 'offline' and ABOS samples can then be combined for analysis purposes with weights to deal with the slight overlap in target populations (the c7% of the population that is aged 70+ *and* uses the internet).

The choice between (i) using paper questionnaires as an alternative data collection mode within the ABOS sample and (ii) supplementing the ABOS sample with an 'offline' interview survey is largely determined by the complexity of the measurement objectives. Given a straightforward

---

[5] It is more usual to use Ofcom survey data to estimate online and offline population sizes but the Ofcom survey is not large enough for a reliable estimation of the age/educational level distribution of the offline population.

[6] An alternative is to use in-person interviewers to contact sampled households that have not responded to the ABOS study. However, this would enforce pre-clustering of the ABOS sample and (probably) some sub-sampling of non-respondents to control costs.

questionnaire, option (i) is a good choice. Firstly, it does not demand the complexity of a dual sample design; secondly, paper data collection is less costly than interview data collection; thirdly, there is plenty of evidence that paper and online questionnaires yield data with similar measurement characteristics (despite inevitable layout differences and the lack of control over the order in which a respondent completes a paper questionnaire).

However, paper cannot readily accommodate complex filtering, loop structures or any responsive pre-population of question and response texts. Simpler versions of the questionnaire might be produced to get around this problem but, in doing so, researchers accept offline non-coverage for the parts of the questionnaire not reproduced on paper. If such structural complexity is necessary, a separate interview survey is the only alternative[7] despite the additional costs, additional design complexity and the occasional risk to inference of combining interview data with online data.

6. Q4: What response rate does the ABOS method get and what is the impact of the design features you have tested?

The calculation of an ABOS response rate is only approximate but we can estimate it by assuming that c8% of sampled addresses will not contain a household, and that an average of c1.9 adults will be resident in each household. These estimates are robust, derived from the Census and from contemporary random sample interview surveys. For the current version of the ABOS *Community Life* survey – which asks all resident adults to complete a questionnaire – it is a simple task to divide the number of validated completed questionnaires by this estimated denominator. For variants that seek just one adult respondent per sampled household, the denominator is simply the estimated number of households. For both variants, we exclude 'rejected' completed questionnaires and partially completed questionnaires from the numerator.

The specific combination of ABOS design features – plus the identity of the sponsor and/or topic of the survey – appears to make a significant difference to the response rate. In 2015-16, the response rate for the Cabinet Office *Community Life* survey was 24% but in a contemporaneous survey for a different sponsor (a 'third sector' organisation that must remain anonymous for now), the response rate was only 9%, averaged across experimental conditions. Given this observed variation in response rates, for each new ABOS study we strongly recommend a pilot or a 'soft launch' phase to establish the likely response rate so that the cost per completed questionnaire can be estimated precisely.

Over the years, we have tested many different design features in an effort either to boost the response rate or to reduce costs. From this we know that :

- Conditional incentives increase the response rate, albeit not in a linear fashion and with some accompanying increase in costs;
- Sending a reminder can almost double the response rate without increasing the cost per completed questionnaire;
- Sending a second reminder has half the effect of the first reminder – and thus increases the cost per completed questionnaire – *but* if this reminder includes one or more paper questionnaires the impact can be greater and it can also alter the responding sample profile (and not just through inclusion of the offline population). These qualities make it a useful tool for manipulating sample composition as well as for increasing the response rate.

[7] An alternative is to make the interview method available 'on request' to the ABOS sample. However, in our experience, very few people will contact the research agency to arrange such an interview. Consequently, coverage of the offline population is no more than nominal if this approach is taken.

Beyond these general findings, we have some evidence from specific ABOS studies that may prove to be generalisable to other ABOS studies. One is that sending a vivid 'survey promotion' postcard (without login details) just before one of the letters can cost-effectively prompt people to take part when the detailed letter arrives. This reflects findings from name-based postal surveys in the UK and elsewhere (see Dillman et al (2014) for a thorough review). The second finding is that the identity of the sponsor can have an impact *even with an otherwise identical survey offer*. Combining this evidence with the observed variation in response rate between different ABOS studies, we conclude that sponsors with little name recognition should (if possible) link up with a partner organisation that can lend to the study greater name recognition or reputation.

7. Q5: How does the response rate vary between subpopulations, and what (if anything) can you do about it?

The postcode address file is itself a 'bare' sample frame but neighbourhood-level data can be attached via the postcode, allowing response rates to be estimated for different strata. Beyond this, we can compare the gender, age and regional profile of ABOS responding samples against the relevant ONS mid-year population totals, allowing us to estimate response rates for post-strata defined by these characteristics. Furthermore, we can also compare an ABOS responding sample against contemporary high response rate random samples to gauge relative bias on a wider range of characteristics[8].

Although we have accumulated response information of this type across multiple ABOS studies, we are wary of over-generalising findings given the present small number of studies.

Stratum level response rates

One reasonably consistent feature is that the *online* response rate is inversely correlated with the local Index of Multiple Deprivation (IMD), available at LSOA level. No other variable that can be attached to the postcode address file appears to be as strong a predictor of response (although there is some evidence that two other variables have additional predictive value: (i) the census proportion living in flats and (ii) the census proportion self-classifying into one of the black ethnic groups[9]).

This variation in stratum response rates can be reduced by selectively applying design features known to influence the response rate. For example, the incentive level can be varied between strata or, more subtly, the *proportion* of addresses that receive an incentive, receive a second reminder or receive a set of paper questionnaires can be varied between strata At the time of writing, we have only employed this kind of responsive design for the *Community Life* survey because we have more evidence about the impact of design features for this study than for any other ABOS study. Although preliminary evidence from other studies suggests some consistency in the additive effect of each of these design features, the evidence is not strong enough for us to determine a clear set of rules that would apply to *all* ABOS studies.

Socio-demographic profiles

Although we may manipulate the response rate via the design features described above, the impact of each design feature on the sample's demographic profile is much harder to detect. Furthermore, while in most ABOS studies we have included experiments to test the impact of one, two or even three design features, some features remain under-evaluated. For example, only the very first ABOS study (for the Cabinet Office) tested the impact of *not* offering an incentive. All of our ABOS studies since then have offered at least £5 in return for completing the questionnaire and most have offered £10.

As it stands, the only design feature that we *know* will change the sample profile is the inclusion or otherwise of paper questionnaires in the second reminder. However, even here, all we can say for

[8] Both of these 'indirect' methods of estimating bias are reliant on an assumption of measurement equivalence between the ABOS study and the benchmark study. This assumption should hold sufficiently well for us to safely assess sample balance with respect to demographics but not with respect to opinions or (in some cases) behaviours.

[9] Both are *negatively* correlated with ABOS response probabilities.

certain is that paper questionnaires bring in more people aged 60+ and especially those aged 75+ (who will otherwise take part in very low numbers). In addition, there are indications that paper questionnaires help bring in people aged under 60 who have long-term illnesses or disabilities and/or live in social rent accommodation.

Given our lack of robust evidence about the impact of specific design features, it is perhaps most instructive to simply compare the demographic profiles of ABOS studies with different sponsors, topics and design features to see if we can identify any consistent outcomes. For illustration, we show data from three 2016 ABOS studies, two of which must remain anonymous for the moment. Table 1 provides the demographic profiles for each of these studies although it is worth noting that slightly different demographic data was collected for each study, hence some cells are empty.

The two surveys with a substantial number of paper returns (the *Community Life* survey and the anonymous survey 2, also for a government department) have very similar demographic profiles, despite the *Community Life* survey having a much higher response rate than survey 2 (24% compared to 15%). Across the four common dimensions (gender, age group, working status and ethnic group), the mean absolute marginal error per category is 3.9% pts for the *Community Life* survey and 2.9%pts for survey 2. However, survey 2 has no data for highest educational qualification or for housing tenure, variables where we can expect higher error scores. If these variables are included, the *Community Life* survey's mean absolute marginal error per category increases from 3.9%pts to 4.6%pts. Given the similarity of the two surveys with respect to the common dimensions, it is reasonable to expect that the survey 2 sample is biased to a similar degree in these other (unmeasured) respects.

Survey 3 (for a 'third sector' organisation) has a distinctively different profile, largely due to the fact that no paper questionnaires were included in the second reminder package. Consequently, paper completions make up only 2% of the total responding sample. For survey 3, the mean absolute marginal error per category is a much higher 6.0%pts across all dimensions. In particular, the sample is younger than it should be and too highly educated. It shares these traits with the online-responding subsets of the *Community Life* survey and of survey 2. However, the most notable bias is in the gender profile: only 37% of respondents were male. This gender bias is *not* found in the online-responding subsets of the *Community Life* survey (47% male) or of survey 2 (49% male) so it has nothing to do with the almost online-only nature of survey 3.

As a point of comparison, the 2015-16 *Community Life* in-person interview survey has a mean absolute marginal error per category of just 1.4%pts. Clearly, the ABOS method produces a less accurate demographic profile than the face-to-face interview method but that is to be expected given its lower response rate. In fact, the accuracy of the ABOS profile is similar (albeit with a different error distribution) to that of a mid-fieldwork in-person interview survey after two or three visits to each sampled address. The accuracy of the profile is also similar to that of contemporary dual-frame RDD surveys, for which 2-5 percentage points of error per category is typical.

Table 1: Demographic profiles for three 2016 ABOS studies plus a contemporary benchmark

Variable	Survey 1: *Community Life, 2015-16 (ABOS version)*	ABOS Survey 2	ABOS Survey 3	Post-stratified benchmark survey estimate (*Community Life, 2015-16, in-person interview version*)
% Responding online	74%	77%	98%	
Responding sample size	3,016	1,170	968	3,027
Response rate	24%	15%	9%	61%
Gender				
*Male	46.1%	48.4%	37.2%	48.8%
*Female	53.9%	51.6%	62.8%	51.2%
Age group				
*16-24	8.5%	8.4%	15.2%	14.4%
*25-34	12.9%	15.8%	16.4%	16.8%
*35-44	14.0%	15.2%	13.8%	16.8%
*45-54	18.1%	15.8%	17.8%	17.2%
*55-64	18.2%	17.5%	18.4%	14.0%
*65-74	18.3%	18.0%	13.5%	11.2%
*75+	9.9%	9.4%	4.9%	9.6%
Working status				
*Working	54.2%	56.5%	52.4%	57.8%
*Not working	45.8%	43.5%	47.6%	42.2%
Ethnic group				
*White British	86.2%	86.7%*	n/a	79.6%
*Other	13.8%	13.3%	n/a	20.4%
Highest qualification (if aged <70)				
*Degree or higher	33.2%	n/a	41.4%	28.3%
*Other qualification	59.0%	n/a	46.7%	57.9%
*No qualifications	7.8%	n/a	11.9%	13.8%
Housing tenure				
*Outright ownership	37.7%	n/a	32.4%	25.5%
*Mortgaged	27.5%	n/a	27.4%	39.3%
*Renting/other	34.8%	n/a	40.3%	35.2%

For all of these variables, benchmark values are available so marginal calibration methods can reduce category-level errors to approximately 0.0%pts so long as a sufficient number of responding cases are present for each category. (At the time of writing) both the *Community Life* survey and survey 3 responding samples have been calibrated in this way, although not to exactly the same set of marginal totals.[10] In both cases, the design effect due to calibration is modest: 1.30 for the *Community Life* survey and 1.46 for survey 3. However, it should be noted that cell level errors may

[10] *Community Life*: gender/age, education/age, housing tenure, region, household size and ethnic group; Survey 3: gender, age, age/working status, education/age and housing tenure.

persist even after marginal calibration, as will non-response errors that are uncorrelated with the variables used in the calibration procedure. Inevitably, the question remains: what level of non-response bias can we expect *after* calibration?

8. Q6: What evidence do you have for non-response bias?

Non-response bias can be identified with some confidence with respect to a sample's demographic profile but with much less confidence with respect to the substantive data, since benchmarks are usually unavailable. Our evidence in this respect is rather limited but in one of our development phases for the *Community Life* survey we were able to shed some light on this, at least for that particular study.

For several years (2012-16), at least one ABOS variant of the *Community Life* survey ran alongside the standard in-person interview survey that was used to produce official statistics. The two designs produced significantly different results even when the samples were weighted to the same population parameters. The question that arose was this: was the difference in results due primarily to (i) *measurement effects* related to the two different modes of data collection (online and paper self-completion questionnaires vs. in-person interviews) or (ii) residual *selection effects*, despite weighting the two samples to the same population parameters? To answer this question it was vital to disentangle selection and measurement effects in order to determine which had the strongest influence on the results.

Williams (2015) describes the investigation in detail elsewhere but, in summary, the evidence suggested that the difference in data collection mode (i.e. measurement effects) was responsible for the bulk of the mismatch observed between the results. Selection effects appeared to be small in comparison.

Naturally, the study has some limiting assumptions and there are questions it could not answer. For example, sample size constraints limited analysis to total population estimates only. Findings might be different if sub-groups were assessed separately so it is possible that selection effects are meaningful for some parts of the population even if not in aggregate for the total population. Absence of evidence for selection effects does not imply that none exist.

Chart 1 plots the estimated measurement effects against what we might call 'system effects': the difference in results between the ABOS version of the *Community Life* survey and the contemporary in-person interview version of the survey. The correlation between the estimated measurement effects and these system effects was very strong (R =.86), leaving only a small amount of residual variance that might be explained by selection effects. Furthermore, as table 2 shows, the distribution of estimated measurement effects - in terms of magnitude - almost exactly matched the distribution of the observed system effects.

Chart 1: Measurement effects (called 'mode effects' here) plotted against system effects ('web – F2F') in the *Community Life* survey

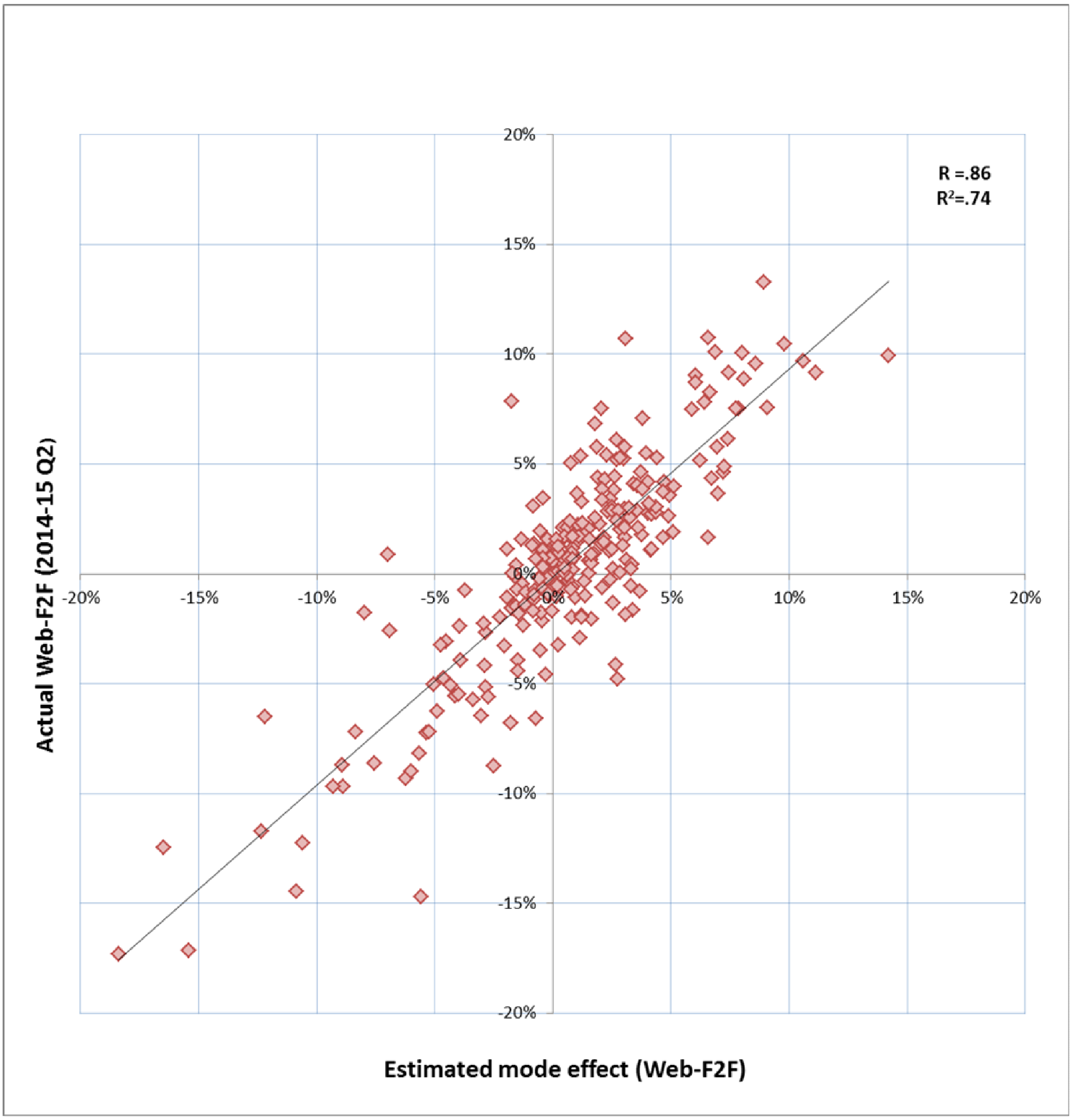

Table 2: Aggregated analysis of estimated measurement effects (online/paper self-completion vs in-person interview) against observed differences between ABOS and in-person interview survey results (July-September 2014)

	Observed difference between data collection systems (July-Sept 2014)	Estimated measurement effect
Mean absolute difference	3.4pp	3.0pp
Median absolute difference	2.1pp	2.1pp
% of differences <1pp	26%	32%
% of differences <2pp	48%	49%
% of differences <3pp	60%	63%

% of differences <4pp	69%	74%
% of differences <5pp	75%	82%
% of differences 5pp+	25%	18%
% of differences 10pp+	5%	4%
% of differences that are statistically significant (null expectation = 5%)	40%	38%

The same kind of method was also used to disentangle selection and measurement effects causing differences between the RDD telephone interview and ABOS versions of Sport England's *Active People* survey. That study also found that measurement effects were stronger than selection effects but concluded that modest selection effects were probably still present in the data. However, in this case, the benchmark – an RDD landline-only telephone-interview survey - was not of gold standard quality so the presence of selection effects was not the cause for concern it would have been had the benchmark been an in-person interview survey.

Two studies are by no means enough to make general conclusions about the nature of ABOS samples. Although it seems fair to say that selection effects are minimal within the ABOS *Community Life* survey (once it has been calibrated to population parameters), it does not follow that they will be minimal for *all* ABOS studies. Nevertheless, it seems fair to say that the relatively low response rates obtained from ABOS studies are not necessarily indicative of strong selection effects. This conclusion aligns with studies of in-person interview surveys which have demonstrated high levels of convergence between estimates based on early data (when the response rate was low) and estimates based on final data. See for example Williams, Sturgis, Brunton-Smith and Moore (2016).

9. Q7: How much does it cost?

ABOS is primarily intended as an alternative to RDD telephone interviewing. So far, we have two examples of 'parallel runs' and have found the cost per completed ABOS questionnaire to be roughly 60-80% of the cost of a same-survey dual-frame RDD telephone interview. Naturally, the specific combination of design features that is adopted will influence this cost ratio.

10. Conclusions

Although the ABOS method has its antecedents in the US, it is a relatively new method for UK survey research and the details will undoubtedly be refined over the next few years. It appears to obtain reasonably balanced samples at response rates that are similar to those achieved with RDD telephone surveying. Selection effects seem modest - where we have been able to estimate them - but we do not yet have enough evidence to make a general statement about the relative robustness of the method compared to the gold standard of in-person interview surveys. Nevertheless, there is enough positive news to continue developing this as a genuine option for survey research studies.

Bibliographie

Dillman, Smyth and Christian (2014), *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method, Wiley*

Messer (2012), *Pushing households to the web: experiments of a 'web+mail' methodology for conducting general public surveys, Washington State University*, https://research.libraries.wsu.edu/xmlui/handle/2376/4653

Park and Humphrey (2014), *Mixed-mode surveys of the general population: Results from the European Social Survey mixed-mode experiment, NatCen Social Research*, www.natcen.ac.uk/media/541183/ess-mixed-mode-natcen-report.pdf

Williams (2014), *Community Life Survey: Investigating the feasibility of sampling all adults in the household, Cabinet Office*, www.gov.uk/government/uploads/system/uploads/attachment_data/file/466925/The_Community_Life_Survey_Investigating_the_Feasibility_of_Sampling_All_Adults_in_the_Household_FINAL.pdf

Williams (2015), *Community Life Survey: Disentangling sample and mode effects, Cabinet Office*, www.gov.uk/government/uploads/system/uploads/attachment_data/file/466922/The_Community_Life_Survey_Disentangling_Sample_and_Mode_Effects_FINAL.pdf

Williams, Sturgis, Brunton-Smith and Moore (2016), *Fieldwork effort, response rate, and the distribution of survey outcomes: a multi-level meta-analysis, NCRM working paper*, eprints.ncrm.ac.uk/3771/