NatCen
National Centre for Social Research

Mixing Modes within a Social Survey:

Opportunities and constraints for the National Survey for Wales

Authors : Gerry Nicolaas
Date : 31/03/2012
Prepared for: the Welsh Government

Content

Executive Summary

The National Survey for Wales is a cross-sectional survey of the adult population in Wales. A sample of addresses is selected at random from the small users Postcode Address File (PAF), stratified by local authority. Survey data are collected in face-to-face interviews with one randomly-selected adult in each selected household. The questionnaire takes about 25 minutes for most people to complete. Although the main themes of the survey are public services and wellbeing, there is scope for the questionnaire to evolve over time to ensure that it meets the needs of Ministers and policy departments.

Declining survey participation has led to an increase in the cost of conducting surveys due to the greater efforts needed to minimise non-contacts and to persuade reluctant respondents. In parallel public sector organisations, including the Welsh Government, are facing financial challenges. These developments have prompted the Welsh Government to consider alternative ways of collecting data at affordable cost, while balancing potentially negative effects on data quality.

The focus of this review is on how alternative modes of data collection could replace face-to-face interviews on the National Survey, or how alternative modes might be combined with face-to-face interviews, to reduce costs, improve coverage or increase response rates.

In section 2, I provide a brief overview of the main data collection modes that could be used to collect data for the National Survey: face-to-face interviews, telephone interviews, postal questionnaires, drop-off self-completion questionnaires and web questionnaires. Face-to-face interviewing is considered to be the superior mode because of its flexibility and potential to achieve high quality data. However, face-to-face interviewing is the most expensive data collection mode, mainly due to the travel costs of interviewers. A move towards using cheaper modes has failed to materialise in the UK because of the lack of adequate sampling frames, under-coverage of the population, non-response bias and measurement error. Given the specific limitations of each of the main modes, it is therefore not surprising that attention has been drawn towards the use of mixed modes in order to compensate for the weaknesses of each individual mode while reducing data collection costs.

It is often claimed that mixing modes of data collection has the potential to improve coverage or increase response rates. However, I found no evidence that a mixed mode design for the National Survey will improve its coverage or response rate compared to its current design. The National Survey currently uses the PAF as a sampling frame which delivers complete coverage of the Welsh household

population, and face-to-face interviewing which produces higher response rates than any other data collection mode. There is also no evidence to suggest that combining different modes will yield higher response rates than face-to-face interviewing.

Also the claim that mixing modes has the potential to significantly reduce data collection costs is questionable given our current dependency on the PAF for selecting a sample with complete coverage of the Welsh household population. Unlike Scandinavian countries and the Netherlands, the UK does not have a population register. Directories and sampling methods that provide adequate coverage of telephone numbers and email addresses for the general population do not exist. Dependency on the PAF means that face-to-face interviewers will be required to travel to the sampled addresses in order to randomly select a respondent. Travel expenses are a large part of the total fieldwork cost of a survey, so potential cost savings are restricted to a reduction in the amount of travel after the first contact and the amount of time spent on face-to-face interviewing.

The most likely mixed mode design for the National Survey would be a sequential design with face-to-face interviewers making contact at sampled addresses and administering the household questionnaire, dropping off self-completion questionnaires with a web option for the selected respondent, followed by face-to-face interviews with those who do not complete the self-completion questionnaire (described more fully in section 5.1). This mixed mode design will achieve similar coverage and response to its current design. The risk of differences in measurement across modes can be reduced by carrying out development work (e.g. uni-mode questionnaire design), field experiments and pre-testing to identify which questions can (and cannot) be asked in different modes. Also the use of show cards and self-completion modules in the face-to-face interviews will reduce the risk of mode effects. But the scope for reducing data collection costs is limited because interviewers are still required to travel to the sampled addresses in order to carry out random respondent selection and achieve interviews with the most reluctant respondents. Furthermore, cost savings will depend on a sufficient proportion of the sample completing the self-completion questionnaires. Multiple reminders using different modes (e.g. telephone and email) and incentives are likely to be required to increase self-completion rates.

The other options described in section 5 have the potential to significantly reduce data collection costs but at the cost of coverage, response and data comparability. The feasibility of these alternative designs depend on finding solutions to self-selection bias in postal surveys; the exclusion of Welsh mobile-only households in Random Digit Dialling surveys; relatively low response rates for self-completion

and telephone surveys compared to face-to-face surveys; and measurement differences between interviewer-administered and self-completion modes. The option of an online panel based on a probability sample (see section 5.5) offers the most forward-looking solution to survey data needs of government departments and academics. This option, however, will require considerable upfront investment and buy-in from multiple users.

A switch from face-to-face interviewing to a mixed mode design could affect the data being collected because different types of people are more likely to participate in different modes (selection effect) and people may answer questions differently depending on the mode of data collection (mode effect). Although several studies have found that many estimates are not affected by mode effects, there remains a risk that some estimates will be. If so, this could result in biased survey estimates (single year as well as aggregate estimates), observed changes that do not correspond with real changes in the Welsh population, and reduced comparability with other surveys. Consequently if the National Survey is to change its design from face-to-face interviews to a mixed mode design, then it will be important to manage this change.

One approach is to run the current face-to-face design in parallel with the new mixed mode design for a period of time. Recent work in this area by Vannieuwenhuyze et al (2011) has demonstrated that this approach could be used to produce comparable measurements across modes by statistically adjusting for measurement differences. Furthermore, it is a safe method of transition.

Another strategy that is often suggested is the use of embedded experiments to quantify the mode effect. It has been claimed that embedded experiments with randomised allocation of respondents to modes can be used to disentangle the mode effect from differences in sample composition across the modes (de Leeuw and Hox, 2011). Embedded experiments are integral to the plans for mixed mode data collection at Statistics Netherlands and the Office for National Statistics. It is hoped that the information can be used to statistically adjust the data for mode effects. However, research in this area is still in its infancy and its feasibility has been questioned.

The Welsh Government is not on its own. Many national governments and their statistical institutes are eager to embrace mixed mode data collection to address increasing concerns about cost and quality. For example, the Office for National Statistics (ONS) is participating in the ESSnet (a network of European Statistical Societies organisations) project entitled “Data collection for social surveys using multiple modes” (DCSS). As part of this project, ONS is planning to use web data collection in parallel with the ‘usual’ LFS. A number of issues that have been

mentioned in this review will have to be resolved as part of this project. It is therefore strongly recommended to follow ONS' progress in this area.

1 Context

1.1 The National Survey for Wales

The National Survey for Wales is a cross-sectional survey of the adult population in Wales. It is the successor of the Living in Wales Survey which was carried out from 2004 to 2008.

A sample of addresses is selected at random from the small users Postcode Address File (PAF), stratified by local authority. Survey interviews take place face-to-face with one randomly-selected adult in each selected household. The fieldwork period began in January 2012 and fieldwork will run continuously. Each year, the aim is to achieve a successful sample size of approximately 14,500 addresses across Wales (660 per local authority). The questionnaire takes about 25 minutes for most people to complete. Although the main themes of the survey are public services and wellbeing, there is scope for the questionnaire to evolve over time to ensure that it meets the needs of Ministers and policy departments.

1.2 Rationale for mixed modes of data collection

Declining survey participation has led to an increase in the cost of conducting surveys due to the greater efforts needed to incentivise respondents, to minimise non-contacts, and to persuade reluctant respondents. In parallel, survey clients (including the Welsh Government) are facing financial challenges creating an urgent need to explore cheaper methods of collecting data. Moreover, technological advances are changing how people communicate, with some modes becoming less viable as a stand-alone method for collecting survey data (e.g. telephone) and other new modes emerging (e.g. the web).

These developments have prompted survey clients, practitioners and methodologists to seek alternative and increasingly innovative ways of collecting data at affordable cost, while balancing potentially negative effects on data quality. The focus of this review is on how modes of data collection could replace face-to-face interviews on the National Survey, or how alternative modes might be combined with face-to-face interviews, to reduce costs, improve coverage or increase response rates.

1.3 Overview of requirements

Any proposed mixed mode design for the National Survey should meet the following requirements:

- The data must support generalisations to the adult population of Wales as well as all households in Wales;
- The survey must provide indicative national figures on a quarterly basis and detailed national figures (i.e. capable of sub-analysis) annually;
- The survey must provide data for indicative estimates at the LA level annually, and more robust estimates at LA level by combining 2 years' worth of data;
- The minimum confidence intervals from the survey are +/- 3%. These levels will allow relevant change over the longer term to be detected;
- Harmonisation with questions used in other surveys where possible, to provide comparable data.

2 Modes of data collection in social surveys

In this chapter I provide a brief overview of the different data collection modes that could be used to collect data for the National Survey. The advantages and disadvantages of each mode are described. A table summarising this section is provided in Appendix A.

2.1 Face-to-face interviews

As for most national government funded surveys of the general population in the UK, the current data collection mode for the National Survey is face-to-face interviewing. Face-to-face interviewing tends to be seen as the superior data collection mode because of its flexibility and potential to achieve high quality data.

Face-to-face surveys of the general population in Wales tend to use the Postcode Address File (PAF) as the sampling frame, thus ensuring almost complete coverage of the household population[1]. In theory it is possible for face-to-face interviewers to visit every residential address in Wales. Interviewers must then establish the eligibility of the address (e.g. occupied and residential), enumerate the households at the address, and randomly select respondents within households.

Response rates tend to be higher for face-to-face interviews compared to other data collection modes. Interviewers are able to tailor their calling patterns and doorstep introductions based on information that they can obtain from their visits to the address and their interaction with household members. It is generally assumed that people find it more difficult to refuse when a request to participate in a survey is made face-to-face.

Compared with other modes, face-to-face interviews offer the greatest scope for collecting complex and large amounts of data. It is generally assumed that the maximum length of a face-to-face interview is about 60 minutes. The presence of an interviewer who can motivate and assist the respondent provides the opportunity to ask more difficult questions, more detailed questions, and more open-ended questions. Visual aids can be used in a face-to-face interview to help the respondent answer the questions. It is also possible for interviewers to collect other types of "data", such as observations of the area and address, physical measurements and biomarkers.

[1] A very small number of addresses are not listed correctly, and others (especially new developments) may not be listed at all for a period of time. Sampling from the Postcode Address File excludes people who live in institutions.

However, face-to-face interviewing is the most expensive data collection mode, mainly due to the travel costs of interviewers. Samples for face-to-face surveys are often clustered to control travel costs but this leads to a reduction in effective sample size (a clustered sample is less varied than a simple random sample of the same size). Data collected using face-to-face interviews may also be subject to interviewer effects such as social desirability bias and interviewer error.

2.2 Telephone interviews

Telephone interviewing enables survey data to be collected far more cheaply than face-to-face interviewing. The cost ratio will vary depending on sample design and interview length, but a rough estimate of the data collection cost of a telephone interview is about 40%-50% of a face-to-face interview. Because travel costs are not an issue, it is not necessary to cluster the sample thus avoiding a reduction in effective sample size. There is no problem in carrying out interviews in remote areas and areas in which the interviewer might feel threatened. And it is possible for the researcher to have tight control over the interviewing process.

However, response rates to telephone surveys are lower than to face-to-face surveys[2] and interviews need to be shorter (a maximum of about 20 minutes for telephone interviews compared with 60 minutes for face-to-face interviews). Questionnaires for telephone surveys have to be simplified because visual aids cannot be used (e.g. show cards), unless these have been mailed to the respondent in advance of the interview.

The main obstacle for telephone interviewing is the lack of a suitable sampling frame or method that will produce a random probability sample with complete coverage of the Welsh general population. The PAF is not a suitable sampling frame for telephone interviews because the file does not include telephone numbers. Telephone numbers can only be matched for a small proportion of addresses[3] because there is no comprehensive list of names that can be matched to the addresses[4] and no comprehensive list of telephone numbers that can be matched to addresses with matched names[5]. Random Digit Dialling (RDD)[6] can be

[2] For illustrative purposes only: If we assume that the response rate for a face-to-face survey is 60%, then the response rate for a telephone survey would be about 50% (all other things being equal).
[3] According to Elliott et al (2006), telephone numbers could be matched for only one quarter of households selected for the UK Labour Force Survey.
[4] In the past the Electoral Register (ER) has been used to match names to addresses. However, the edited version of the ER which is for sale excludes not only those who are ineligible to vote but also those who are eligible but have opted out of their details being made publicly available. The average opt-out rate for the UK was 40% in January 2012 but this rate varies greatly by Local Authority; e.g. 92% in the Vale of Glamorgan compared to only 12% in City & County of Swansea. Those who opt out are more likely than usual to be single, middle-aged, wealthy and health conscious. (Callcredit Information Group, 2012)
[5] Telephone directories exclude all ex-directory numbers and mobile numbers. Commercially available lists of telephone numbers may include some ex-directory numbers and mobile numbers but these lists still suffer from under-coverage.

used to generate numbers for landlines but current methods tend to exclude mobile only households. About 19% of adults in Wales were living in mobile only households in 2011. Mobile-only households are more prevalent in urban (20%) than in rural areas (12%), and are most common among those aged 16-34 (37%). (Ofcom, 2011) There are some new developments to include mobile only households in RDD surveys; most of these involve a dual frame approach (discussed in section 4.2.1).

2.3 Postal questionnaires

Conducting a survey by post is a relatively cheap method of collecting data because it does not involve interviewer fees nor other fieldwork costs (e.g. travel). Consequently it is possible to have a larger sample size for a given cost than when using other data collection methods, such as face-to-face methods and telephone methods. Difficulties related to contacting people in remote areas and those who are difficult to contact at home are avoided. It is also not necessary to cluster the sample geographically to reduce costs. Another potential advantage of postal surveys is that they give respondents the opportunity to reflect on the questions and give more precise answers.

However, there are also several disadvantages associated with postal surveys. Response rates to postal surveys are substantially lower than to face-to-face and telephone surveys[7]. The researcher has limited influence over the return of questionnaires. Additionally, the questionnaire has to be brief (generally no more than twelve A4 pages) and easy to complete. The absence of an interviewer to motivate and help the respondent can result in poorer quality data; e.g. more item non-response, less complete answers, routing errors. Postal questionnaires can be problematic for those who are less literate and those with a limited command of English or Welsh.

The only suitable sampling frame for postal surveys of the Welsh general population is the PAF. As mentioned before, PAF is a list of addresses and does not include names. Consequently postal questionnaires have to be sent to the household without a name (e.g. addressed to the resident or the occupier) which will increase the risk that the questionnaire is mistaken for junk mail and thrown away. There is also no control over respondent selection which could result in biased data, particularly if the respondent's decision to complete the questionnaire is correlated with the topic of the survey (e.g. the person with the strongest opinion

[6] There are various permutations of the RDD method but all involve attaching randomly generated digits to a sample of valid area codes and prefixes. RDD was developed to overcome the exclusion of unlisted numbers that would be missed if the telephone numbers were selected from a phone book.

[7] For illustrative purposes only: If we assume that the response rate for a face-to-face survey is 60%, then the response rate for a postal survey would be about 40% (all other things being equal).

about the topic may be more likely to complete the questionnaire). Instructions on who should complete the questionnaire (e.g. last birthday) can be provided but they are unlikely to be always followed.

Other sampling frame options which include names such as the edited Electoral Register (ER) and commercial lists have more serious shortcomings. For example, the edited ER has severe coverage error with the exclusion of those who are ineligible to vote (e.g. UK residents who are not citizens of the EU, the Republic of Ireland or qualifying Commonwealth countries) and those who fail to register, as well as those who have requested that their details are not shown on the edited ER (average opt-out rate in the UK was 40% in January 2012[8]). Commercial lists tend to supplement the edited ER with data pooled from other sources but they still suffer from substantial under-coverage (for an evaluation of the coverage of a commercial list compared to the edited ER, see Tipping and Nicolaas, 2006).

2.4 Drop-off self-completion questionnaires

In this approach, an interviewer goes to the sampled address and hands the questionnaire to a respondent for them to complete at their own convenience. In some cases, the respondent is asked to mail the questionnaire back or the interviewer returns to pick it up at a later date. Drop-off self-completion questionnaires are used on the Welsh Health Survey.

This approach shares some positive features with face-to-face interviewing: good coverage of the population when using PAF as a sampling frame, interviewer involvement in persuading the respondent to take part in the survey, and control over the respondent selection technique. This approach would be expected to produce a higher response rate than a postal survey, particularly if the interviewer has to return to the household to collect the completed questionnaire.

Similar to postal questionnaires, the self-completion questionnaire has to be short and easy to complete. Generally, the data is likely to be of poorer quality compared to data collected in a face-to-face interview (e.g. more item non-response, less complete answers, routing errors). Self-completion questionnaires can be problematic for those who are less literate[9] and those with a limited command of English or Welsh. If an interviewer is expected to return to the household to collect the questionnaire, there is some scope for the interviewer to

[8] This figure was not available for Wales.

[9] Twelve per cent of the adult population in Wales had literacy levels below Level 1 in 2010 (Miller and Lewis, 2011). Those with lower literacy levels were more likely to be renting from local authorities or housing associations, and more likely to be claiming housing benefit or council tax support, sickness /disability benefit or unemployment benefit than the sample as a whole. There are also links between low levels of literacy and self-reported health status, employment status and average household income. It is therefore likely that the underrepresentation of those with low level literacy skills in a survey will result in biased estimates of wellbeing and use of public services.

check the questionnaire and help the respondent to complete it. Although we expect the response rate to be higher than for a postal survey, it is likely to be lower than the response rate for an equivalent face-to-face survey (all other things being equal).

Although the drop-off approach will reduce interviewer fees, it may only be marginally cheaper than face-to-face interviews because this approach still involves interviewer travel. Interviewer travel can be reduced by getting the respondent to return the questionnaire by post rather than the interviewer collecting it (but this may reduce the response rate). Using the drop-off approach rather than face-to-face interviews is likely to produce more cost savings for those surveys that require more than one person in the household to complete a questionnaire. Even in the absence of significant cost savings, the drop-off questionnaire may be a solution to a shortage of available interviewers in the sampled areas.

2.5 Web questionnaires

There has been reluctance to use web questionnaires as a stand-alone data collection mode for nationally representative surveys of the household population. The main concern has been the limited coverage of the household population; in 2011 almost one quarter of UK households did not have access to the internet at home (Ofcom, 2011)[10]. Furthermore, those without access to the internet at home are known to be older on average and are more likely to be on benefits or have a low income.

Even when we limit the study population to those with home access to the internet, there is no comprehensive list of such households from which a random sample can be selected. RDD samples and samples selected from PAF can be used with screening for internet access but that involves contacting sample members using a different mode (e.g. post, telephone) and inviting them to go online to complete a web questionnaire. This extra stage in gaining cooperation will reduce the response rate and increase the risk of non-response bias. For example, at NatCen Social Research we carried out a study which involved sending letters to Omnibus survey respondents who had agreed to be recontacted, and we asked them to go online and complete a short questionnaire. Only 5% completed the online questionnaire after the initial postal invitation and a postcard reminder. Sending an additional postal reminder with a £5 gift voucher significantly boosted the response rate but it was still only 24%. Higher response rates can be obtained by using the same mode for contact and data collection (e.g. postal, telephone).

[10] A figure for Wales is not available.

In a recent study, Millar and colleagues (2009) found that even when respondents had access to the internet, certain groups were less likely to respond via web than mail; e.g. older respondents, those on lower incomes, those with lower qualification levels, female respondents and those who were not married. Although the internet may be a more acceptable and accessible mode for young people compared to older people, when given a choice it would appear that there is still a significant proportion of young people who prefer other modes to the internet.

3 Mixing modes of data collection

Face-to-face interviewing continues to be the dominant data collection mode for most surveys of the general population in the UK. A move towards cheaper modes such as telephone interviewing, postal questionnaires and more recently web questionnaires has failed to materialise because of under-coverage of the population, lack of adequate sampling frames, non-response bias and measurement error.

Nonetheless, the increasing cost of fieldwork for face-to-face surveys coupled with declining survey budgets is pushing survey clients and survey practitioners to look for cheaper ways of collecting survey data. Given the specific limitations of each of the main modes, it is therefore not surprising that attention is drawn towards the use of mixed modes. As noted by de Leeuw (2005, p. 235), "Survey designers choose a mixed-mode approach because mixing modes gives an opportunity to compensate for the weaknesses of each individual mode at affordable cost. The most cost-effective method may not be optimal for a specific study. By combining this method with a second more expensive method the researcher has the best of both worlds: less costs and less error than in a unimode approach."

This chapter provides a summary of what is known about mixing modes of data collection and how mixing modes could affect the quality of data collected for the National Survey.

3.1 How to mix modes of data collection

There are two main designs for mixing modes of data collection: concurrent and sequential mixed-mode designs. In a concurrent mixed mode design, two or more modes are implemented in parallel within a certain time period. For example, respondents could be offered a choice of data collection modes. In a sequential design the different modes are implemented in sequential order during the data collection period. For example, non-respondents to the first data collection mode could be followed up using a different data collection mode.

These two designs can be used in the following four scenarios (de Leeuw, 2005):

1. Different parts of one questionnaire in different modes, with the same sample, during one time period.

   For example, sensitive questions are included in a self-completion form while the remainder of questions are asked by an interviewer. Mixing modes in this way is not a problem because overall data quality is improved while maintaining comparability of all data items across respondents.

2. One questionnaire, with the same sample, during one time period.

   For example, respondents can be offered a choice of data collection mode or non-respondents to one data collection mode could be followed up using a different data collection mode. In this scenario the same data items are being collected from different respondents using different modes. Data comparability could be compromised if respondents give different answers to the same questions depending on the mode of data collection.

3. One questionnaire, with the same sample, over multiple time points.

   For example, a longitudinal survey in which the same respondents are interviewed at different points in time using different data collection modes. In this scenario it may not be clear whether an observed change in survey estimates over time is a real change or has been caused by the change in data collection modes.

4. Different samples, different modes.

   This scenario applies to between-survey comparisons (including international and regional comparisons) where different surveys have used different data collection modes. I would also include in this scenario repeated cross-sectional surveys that use different data collection modes and/or different combinations of modes from one year to the next. True differences over time and differences between countries/regions are possibly confounded by differences in the mode of data collection.

For the purpose of assessing the feasibility of using mixed modes for the National Survey, I am mainly focussing on scenario 2 which applies to the use of multiple data collection modes for any given year of the National Survey. However, I will also consider scenario 4 in relation to possible time trends (although time trends are not the primary focus of the National Survey) and comparisons with other surveys which are relevant to the users of National Survey data.

3.2 Mixing modes and data quality

3.2.1 Coverage of the study population

The target population for the National Survey is the adult population[11] of Wales. There is a risk of coverage error when segments of this population are not included on the sampling frame or they do not have access to the mode of data collection. If these segments of the population are different from those who are included and their differences are relevant to what the survey is trying to measure, then it is very likely that this will bias the survey estimates.

[11] Adult population is currently defined as aged 16 and over.

A mixed mode design is unlikely to increase the coverage of the Welsh population in the National Survey. The current design already uses the most comprehensive and up-to-date sampling frame (i.e. the small users Postcode Address File) in conjunction with face-to-face interviewing to obtain almost complete coverage of the adult household population. Only a small number of addresses that are new developments are not listed on the PAF. In theory face-to-face interviews should be adaptable enough to ensure that all adults can take part, but in practice a small proportion of adults may be excluded due to language difficulties, ill health (mental or physical) or disability. These exclusions can be avoided by using translators, proxy respondents and other adaptations but these methods will increase data collection costs and may reduce data quality.

Although it is unlikely that the population coverage of the National Survey can be improved, there is a risk that a mixed mode design could reduce the population coverage if this design requires a different sampling frame and/or a combination of modes is used that excludes segments of the population. For example, RDD sampling methods that are currently in use in the UK will exclude households without a fixed telephone line; this is about 20% of adults living in Wales[12]. Coverage error would be catastrophic if the edited ER was used as a sampling frame; at least half of the population would be excluded because they are either not eligible to vote or they have failed to register, or they have registered but requested that their details be excluded from the edited ER.

Dual-frame sample designs are sometimes used to improve population coverage. This involves taking independent samples from each of two frames and combining these two samples to represent the population of interest. The dual-frame sampling approach is only useful when the amount of under-coverage from using a single frame is substantially improved by the introduction of second frame. This is unlikely to be the case when using the PAF as a sampling frame because its coverage of the population is very good. However, dual frame designs are being explored to improve the population coverage that can be achieved when using RDD (described further in section 4).

For random probability samples, it is not only required that every unit in the population has a chance of being selected for the survey but that we can also calculate the probability of selection for every unit in the population. Our choice of sampling frames and data collection modes can complicate or even limit our ability to control and calculate these selection probabilities. For example, for mixed mode surveys that use RDD it is essential that we can determine the total number of

[12] In 2011, 19% of adults in Wales were living in mobile only households and 1% were living in households without either a fixed line or a mobile. (Ofcom, 2011)

eligible telephone numbers for every household/respondent so that the data can be adjusted for these unequal selection probabilities. However, determining the total number of eligible telephone numbers in a household has become more difficult over time and this process can take several minutes for each household.

A mixed mode design that uses PAF as a sampling frame and then invites people to either complete a postal or web questionnaire provides the researcher with little control over who completes the questionnaire. Household members have to be instructed to select a respondent at random according to instructions (e.g. last birthday) because there are no names on the sampling frame and there is no interviewer involvement. It is unknown to what extent household members follow such instructions correctly.

To sum up, it is very likely that a mixed mode design for the National Survey will require the involvement of an interviewer at the sampling and contact stage in order to achieve acceptable coverage of the Welsh adult population, to ensure that respondent selection is carried out correctly and that selection probabilities can be calculated. The most likely strategy is the use of PAF as a sampling frame and the involvement of a face-to-face interviewer at the contact stage. There are two other possible strategies for the sampling and contact stage: (a) dual frame samples of landline and mobile numbers with telephone interviewer at the contact stage, and (b) sample of addresses from PAF with postal contact and instructions for respondent selection. However, both of these alternative strategies would require further development work to ensure that population coverage is acceptable.

SUMMARY: Coverage of the study population

Mixed mode design with optimum coverage for NSW:

- PAF sample
- Face-to-face interviewer at contact stage

Possible alternative designs requiring further development:

- Dual frame RDD with tel interviewer at contact stage
- PAF sample with postal contact and instructions for respondent selection

3.2.2 Response rates and non-response bias

Magnitude of non-response

Response rates tend to vary by mode with face-to-face interviews having the highest response rate, followed by telephone interviews, and then postal and web questionnaires. Meta-analytic evidence suggests lower response rates for web than other modes (Lozar Manfreda et al, 2008; Shih and Fan, 2008). However, the magnitude of difference in response rate by mode varies depending on survey topic, population group, availability of relevant contact information, respondent burden, and the amount of effort that is made to maximise response within each mode (e.g. number of contact attempts, refusal conversions and incentives). One relevant example of how response rate for modes may differ by population group is that of respondent age; there is some evidence that compared with older respondents, younger people are more likely to state a preference for web over postal surveys (Millar et al, 2009).

If different sample members are more likely to respond to one or the other mode, then it should be possible to use mixed modes to increase response rates. This has led many to explore the option of providing respondents with a choice of data collection mode (i.e. concurrent mixed mode design).

However, various experimental studies have shown a lack of evidence that a choice of mode will actually increase overall response (Dillman et al, 1995; Lozar Manfreda et al, 2001; Balden, 2004). Indeed some of the evidence suggests that offering a choice of modes can decrease response. One possible explanation for this is the 'responding through the mode at hand' principle; e.g. if sample members are sent a postal questionnaire with the option of completing the questionnaire online, they are more likely to complete the postal questionnaire than to make the extra effort to go online (Holmberg and Lorenc, 2008). Those who initially opt for the web questionnaire may in the interim drop out.

A more successful strategy for increasing response rates is the sequential mixed mode design where all sample members are asked to complete the survey using one mode before moving on to another mode for non-responders. Such designs can also be used to minimise costs without sacrificing response by using the cheapest data collection mode (e.g. web questionnaire) first, before proceeding to the use of increasingly more expensive data collection modes among the remaining non-responders (e.g. postal questionnaire, telephone interview, drop-off self-completion, face-to-face interview respectively). (Hochstim, 1967; Siemiatycki,1979; Japec, 1995; Voogt and Saris, 2005)

With regard to the National Survey, we can expect that a sequential mixed mode design that includes face-to-face interviewing should achieve a similar response

rate compared to its current unimode design that uses only face-to-face interviewing. Furthermore this can be achieved at a lower cost if the cheaper modes are applied first, leaving the more expensive face-to-face interviews for last.

Non-response bias

Survey estimates can be biased if non-respondents are different from respondents in relation to what the study is trying to measure. This non-response error tends to vary across modes to the extent that different sample members are more likely to respond to a given mode (for example, more web-literate people are more likely to respond to a web questionnaire). If different sample members are more likely to respond to one mode than another, then it should be possible to not only use mixed modes to maximise response but also to increase the representativeness of the responding sample.

On the whole, face-to-face surveys tend to achieve a more equal response rate across population sub-groups than telephone, postal and web surveys. Compared to face-to-face surveys, telephone surveys tend to have fewer respondents with low education, fewer respondents with low incomes and fewer respondents belonging to minority ethnic groups (Holbrook et al, 2003). Postal and web surveys are also less likely than face-to-face surveys to include respondents with low education. Furthermore, postal and web surveys tend to attract respondents with higher levels of literacy (including computer-literacy). Sample members who are not particularly interested in the survey topic are also more likely to be under-represented in postal and web surveys compared to face-to-face and telephone interviewers, because there is no interviewer present to persuade them to take part for other reasons.

This suggests that there is potential for reducing non-response bias on the National Survey by using a mixed modes design that includes face-to-face interviewing. The use of other modes alongside face-to-face interviews has the potential of drawing into the achieved sample people who would not have agreed to a face-to-face interview but are willing to use a different data collection mode.

To conclude, the most promising mixed mode design for the National Survey would be a sequential design that first attempts to collect data using the cheapest mode before moving on to the more expensive data collection modes. To ensure that this sequential design will not reduce the overall response rate or increase non-response bias, it is envisaged that face-to-face interviewing will be necessary to collect data from the most reluctant respondents.

SUMMARY: Response rates and non-response bias

Mixed mode design with optimum response and representation for NSW:

- Sequential mixed mode design
- Sequential order based on increasing cost of mode
- Face-to-face interviews for most reluctant respondents

3.2.3 Differences in measurement

When using a concurrent or sequential mixed mode strategy where the same questions are asked in different modes there may be a problem in comparing the data collected. It may be unclear whether answers provided in one mode are comparable to those collected in a different mode, and therefore whether they can be combined for analysis.

Causes of differences in measurement

Jäckle et al (2011) use the cognitive response model (Tourangeau et al, 1984) to describe how different modes can affect the way respondents cognitively process questions and provide answers (the model is presented in Appendix B). How a respondent answers a survey question will depend on the following three conditions: (a) how the information is presented to them, (b) the amount of effort that they are able and willing to make to thoroughly process the question and (c) their willingness to share the true answer. Respondents may shortcut the cognitive processing of the survey question (also known as survey satisficing) if they encounter difficulties with the information that is presented to them or if they are not motivated to put in sufficient effort. Reduced willingness to provide the 'true' answer to a question because the answer is not perceived by the respondent as being socially acceptable is referred to as social desirability bias.

The extent to which the above three conditions are optimised will vary by mode, resulting in differences in survey satisficing and social desirability bias by mode. Interviewers in face-to-face modes and telephone modes can motivate and help respondents to thoroughly process the question. The visual presentation of questions and response options in postal and web questionnaires may make the task of answering a survey question easier for respondents, whereas the oral presentation in telephone interviews requires the respondent to hold all relevant

information in their short-term memory. Some question formats may be more difficult to answer in one mode compared to another (e.g ranking questions). Respondents may be more inclined to portray themselves in a favourable light or to give answers that they think the interviewer wants to hear in face-to-face and telephone interviews compared to self-completion questionnaires (i.e. social desirability bias).

Evidence of differences in measurement

The literature suggests that differences in measurement between face-to-face and telephone interviewing, when conducted well, are minimal. This is because both modes are interviewer-administered and both present questions aurally while collecting a verbal response. For complex questions, the assistance of an interviewer can be useful for explaining the task, and guiding the respondent through the survey. The interviewer can also motivate the respondent to provide full and thorough answers, particularly in a face-to-face interview but also in a telephone interview.

Similarly, differences between postal and web questionnaires are considered to be small because both modes are self-administered, with questions and response options presented visually. However, visual presentation may differ between paper and web surveys, and this could cause differences in measurement (Christian and Dillman, 2004; Fuchs, 2009).The appearance of a postal questionnaire is the same for the designer and all respondents who receive it, whereas a web questionnaire is mediated by the hardware and software on which it is viewed. In addition, web design allows the flexible use of features that are seldom, if ever, used in postal surveys, including complex routing of questions, multiple colours and images, innovative question displays, split screens and sound (Dillman, 2007). In addition, there is evidence from research carried out on postal surveys that small changes in visual design, such as the space provided for open-ended questions, or the use of arrows to direct respondents to follow-up questions, may influence the answers respondents give (Christian and Dillman, 2004). Therefore the comparability of visual presentation of questions needs to be considered when postal and web modes are mixed.

Differences in measurement are more likely to occur when combining interviewer modes and self-completion modes in a mixed mode design. There is ample evidence demonstrating that sensitive questions are more likely to produce socially desirable responses in interviewer modes compared to self-completion modes (see Tourangeau and Yan, 2007; Kreuter, Presser and Tourangeau, 2008). And there is some evidence to suggest that subjective questions are more prone to mode effects than factual questions (see Lozar Manfreda and Vehovar, 2002;

Schonlau et al, 2003). Scalar questions may also produce different responses in different modes. Dillman et al (2009, p. 316-317), summarising over 2 decades of studies describe a "substantial difference in responses to scalar questions when asked by telephone versus visual modes" (i.e., mail, web and face-to-face with show card), with telephone respondents giving more extreme positive answers. Agree/Disagree scales are known to be a cognitively complex task in any mode (see Fowler, 1995; Converse and Presser, 1986; Saris et al, 2010), but are particularly problematic when used in mixed mode studies because respondents tend to cope with the demanding task in different ways in different modes. A recent experiment testing differences in measurement between face-to-face interviews, telephone interviews and web questionnaires found evidence of acquiescence bias in the face-to-face and telephone modes, positivity bias in telephone mode, and middle category bias in web mode (Campanelli et al, 2011).

Certain question formats are used in some modes and not others. This can be because some question formats are suited to a mode, or because of the design tradition that has developed for modes. For example, it is common practice to use show cards in face-to-face interviews but they are not used in telephone surveys[13]. The visual display of response options on a show card in a face-to-face interview compared to the aural delivery of response options in a telephone interview provides different stimuli to the respondents and may therefore result in differences in measurement (e.g. primacy bias with visual display versus recency bias with oral delivery). When response lists are long, it is common practice in telephone surveys to use a branched question design, where a series of questions is used first to identify the direction of response (e.g. "Are you satisfied or dissatisfied with X?") followed by the extremeness of that response (e.g. If satisfied: "Is that very satisfied or fairly satisfied?"). This limits the information that the respondent has to focus on in a single question. However, there is evidence to suggest that this may result in the selection of more extreme response options compared to the presentation of the full response scale which is more common in other modes. It is unclear which method produces the better quality data but the use of different question formats in mixed mode surveys increases the risk of differences in measurement. For this reason it is essential to review the question formats being used when mixing modes, and questions should be designed to provide an equivalent stimulus across modes (Dillman, 2000).

Campanelli et al (2011) provide an overview of the evidence of mode effects by question type/format, with recommendations on how to design these questions for mixed mode studies.

[13] If addresses are available it is possible to send the show cards in advance of the telephone interview. However, there is a risk that the show cards are misplaced or thrown away.

SUMMARY: Differences in measurement

Risk of differences in measurement greater:

- When mixing visual and aural modes
- When mixing interview and self-completion modes
- For subjective questions than factual questions

Minimising differences in measurement

Good question design is critical for minimising mode effects. The depth of cognitive processing and effort that is required to answer poorly designed questions will differ by mode, thus increasing the risk and magnitude of measurement differences. The risk of satisficing (i.e. not making the effort to produce an accurate answer) is likely to be greater in self-completion modes than interviewer modes. Furthermore, there is some evidence to suggest that the type of satisficing behaviour may differ across modes; e.g. more middle categories being selected in web mode, more positivity bias[14] in telephone mode.

Even with good question design, the absence of differences across modes cannot be guaranteed. Certain question formats are known to be less portable across modes; e.g. 'code all that apply' questions, questions with long response lists, 'agree/disagree' scales, end-labelled scales, ranking questions, rating questions. Certain question types are known to be more susceptible to mode effects; e.g. sensitive questions, non-factual questions.

There are three main strategies for minimising mode effects in terms of measurement; mode-specific design, uni-mode design and generalised mode design (de Leeuw, 2008):

1. For the mode-specific design, questions are designed separately for each mode, making use of the unique features of that mode (e.g. use of show cards in face to face interviews or supporting diagrams in web surveys). The rationale is to reduce overall error. However, this is based on the assumption that the same concept is being measured across all modes but with different levels of accuracy (i.e. no systematic bias). This may not be the case. For example, the use of branching of response options (optimum design for telephone mode) and no branching of response options

[14] Tendency to give more extreme positive answers in the telephone mode is sometimes referred to as "CATI positivity bias" (Dillman et al, 2009, p. 316-317).

(optimum design for self-completion modes) produces systematic differences in responses between modes. Therefore, this strategy for minimising mode effects comes with the burden of proof on the designer to demonstrate that there are no systematic biases.

2. An alternative strategy for minimising mode effects is uni-mode design: 'the writing and presenting of questions to respondents in a way that assures receipt by respondents of a common mental stimulus, regardless of survey mode' (Dillman, 2007; p. 232). A uni-mode design seeks to produce questions that share design features across modes. Dillman (2007) has produced a list of recommendations for uni-mode design, such as making all response options the same across modes and incorporating them into the stem of the question, reducing the number of response options, using the same descriptive labels for response categories, developing equivalent instructions for skip patterns, etc. (see Appendix C for a full list). However, there is a risk when trying to standardise the features across modes that this will result in sub-optimal design for some or all modes.
3. The third strategy for minimising mode effects is the generalised mode design. Questions are purposively constructed to be different in different modes with the aim of achieving cognitive equivalence of the perceived stimuli, thereby resulting in equivalent answers across modes (de Leeuw, 2008). Although there has been some work done in this area (for example, Christian et al, 2008), it is still early days and much more empirical research is needed to understand what represents the same stimulus across different modes.

The choice of strategy should depend on the combination of modes being used and the question type/format. However, the lack of empirical evidence to support mode-specific design and the generalised mode design suggests that the uni-mode design is likely to remain the dominant question design strategy for mixed mode surveys in the near future.

Although uni-mode design should minimise the risk of differences in measurement, this may not be sufficient for all questions. It is not always possible to achieve equivalence in stimuli across modes. It is therefore recommended that questions to be used in a mixed modes design should be pre-tested[15].

[15] An overview of question pre-testing is beyond the scope of this investigation. However, readers are referred to Czaja (1998) for an overview of question pre-testing methods.

SUMMARY: Minimising differences in measurement

Methods for minimising differences in measurement:

- Uni-mode questionnaire design
- Pre-test questions in different modes

If mixing face-to-face interviews and self-completion questionnaires:

- Consider use of self-completion modules in face-to-face mode, particularly for sensitive questions
- Consider use of show cards in face-to-face mode
- Avoid use of middle categories

If mixing face-to-face and telephone interviews:

- Consider not using show cards in face-to-face mode
- Keep response lists short
- Avoid scalar questions
- Replace “code all that apply” questions with a series of “yes/no” questions in both modes

If mixing face-to-face interviews, telephone interviews and self-completion questionnaires:

- Keep response lists short
- Avoid scalar questions
- Avoid middle categories
- Replace “code all that apply” questions with a series of “yes/no” questions in both modes
- Thorough pre-testing is required

Impact of using mixed mode on the National Survey

I have so far discussed how the choice of modes can affect the data being collected because different types of people are more likely to participate in different modes (selection effect) and people may answer questions differently depending on the mode of data collection (mode effect). In this section I discuss how these differences could potentially affect the survey estimates.

As mentioned in section 3.1, there are two scenarios in which differences in data collection mode could be relevant to the National Survey: (1) the same data items are being collected from different respondents using different modes during one year of the survey; and (2) different modes are used to collect data from different samples that are being compared (e.g. comparing survey estimates for different years of the National Survey, comparing the National Survey estimates with survey estimates from other relevant surveys).

If the National Survey was to use different modes to survey different respondents for any particular year, then it is possible that differences in measurement across modes will have an impact on survey estimates. The direction and magnitude of impact on the estimates will depend on the proportion responding by each mode and the way in which that specific question is affected by the mode (as described earlier in this section). The use of different modes in this scenario could be particularly problematic for comparisons between groups when the process of allocation to different modes is non-random. For example, are older people really different in their opinions compared to young people, or is this a social desirability effect combined with older people being more likely to take part in a face-to-face interview rather than complete a web questionnaire?

Although the questionnaire content of the National Survey is likely to change over time so that it continues to meet the needs of the Welsh Government, it is very likely that the questionnaire will include a set of variables that will be repeated over two or more years of the survey. It is therefore possible that data users will want to compare estimates based on these variables from one year to the next. However, any observed changes in these estimates could be confounded by mode effects due to changing from a face-to-face mode to a mixed mode. This potential problem is not only limited to the moment of change-over, but could persist for subsequent years of the survey because the proportion responding by each mode is likely to change over time (e.g. the proportion of people responding by web).

Similarly, any comparisons between the National Survey and other relevant surveys could be affected by a change from face-to-face interviewing to mixed mode. Almost all surveys that could be compared with the National Survey are conducted face-to-face (Bryson and Huskinson, 2012). Even if these surveys

changed to a similar mixed mode design[16], the proportion and characteristics responding by each mode may be different for surveys carried out in other countries but also for other surveys carried out in Wales (due to differences in protocol such as number of reminders before switching to the next mode in a sequential design).

The above risks also apply to aggregating data across reporting years in order to increase the sample size and so allow more detailed analysis[17]. The direction and magnitude of the impact of using mixed modes on the aggregated estimates will depend on the proportion responding by each mode in each year and the way in which that specific question is affected by measurement error. Comparisons between sub-groups within local authorities, comparisons between local authorities, and comparisons between local authority level and Wales-level results could all be affected.

It is envisaged that the questionnaire content of the National Survey will change substantially for 2014-15 to ensure it continues to provide the Welsh Government with the information it needs. The risk of increased bias due to the use of mixed modes can be minimised by designing these questions to achieve measurement equivalence across the modes. As described in the previous section, the most promising method at the moment is uni-mode questionnaire design combined with pre-testing. Other methods, such as mode-specific and generalised mode designs, would require substantial development work and testing. Obviously the comparability of these estimates with similar estimates from other surveys would be reduced. Estimating and adjusting for measurement differences between surveys would be difficult, given the complications involved in disentangling selection effects and mode effects.

All of the above relates to the risk of measurement differences across modes, rather than definite measurement differences. The risk is greater when mixing modes that use (1) visual versus aural stimuli, and (2) interviewer administration versus self-completion. The risk is also greater for certain question types (e.g. sensitive questions) and question formats (e.g. scalar questions). But beyond this, it remains difficult to predict whether a variable will be susceptible to mode effects, and if so how large this effect will be and whether these differences have an impact on the substantive interpretation of the data (e.g. correlations, multivariate analysis).

[16] For example, ONS is exploring the feasibility of changing the first wave of the Labour Force Survey to a sequential mixed mode design which will include web data collection.

[17] The National Survey for Wales has been designed to give an annual sample size per Local Authority of around 600. To improve precision, local level statistics will be generated using two or more years of aggregated data.

3.2.4 Speed, cost and operational issues

As discussed in section 3.2.2, the most promising mixed mode design for maximising coverage and response while reducing data collection costs, is a sequential mixed mode design which starts with the cheapest data collection mode and finishes with face-to-face interviewing for the most reluctant respondents. However, the fieldwork for a sequential mixed mode design is likely to take longer than for a single mode design. For example, the implementation of a sequential mixed mode design for the Health Survey of Statistics Netherlands increased the total data collection period from two months to three months.

The actual cost savings associated with a sequential mixed mode design depend on other aspects of the survey design, such as the mode of first contact, number of people in the household being interviewed, and the proportion of the sample that is successfully encouraged to complete the questionnaire using the cheaper data collection modes. If an interviewer is required at the contact stage to carry out the respondent selection, then this will wipe out some of the savings that could be achieved by using cheaper modes in a mixed mode design. Cost savings are also lower for household surveys compared to surveys of individuals; i.e. failure to use the cheaper mode for even one person in the household results in the more expensive mode having to be used thus wiping out most of the cost savings that had been achieved when using the cheaper mode for the other household members. And finally, cost savings depend on a sufficient proportion of the sample actually using the cheaper modes. Several reminders and incentives may be required to achieve this.

Survey management is a lot more complicated for mixed mode surveys compared to single mode surveys. A sophisticated sample management system is required that can move sampled cases from one mode to the other. Ideally this would be a single sample management system rather than separate sample management systems for each mode. The system should allow collection of data from different individuals in a household, using different modes simultaneously in the same fieldwork period. Most national statistical institutes and survey research organisations are developing their own bespoke systems but this has proven to be quite challenging[18]. It is possible that future sample management systems will be web-based which would maximise efficiency in data collection.

Using a sequential mixed mode design with face-to-face interviewing for the most reluctant respondents makes the organisation of the fieldwork more difficult. First of all, the most reluctant respondents are likely to be geographically dispersed,

[18] Personal communications with the Office for National Statistics, Statistics Netherlands, and the Institute for Survey Research at the University of Michigan.

requiring interviewers to spend more time travelling and it is unknown in advance where these addresses will be. Secondly, the interviewers are dealing with the most reluctant respondents which will require extra interviewer effort to make contact and to gain cooperation. This will require extra training and support for the interviewers. As a result, the cost per productive face-to-face interview in a sequential mixed mode survey is likely to be much higher than in a single mode survey.

Using more than one mode can have implications for the length and complexity of the questionnaires. It is generally accepted that the questionnaire length for self-completion surveys should be shorter and less complicated than for interviewer-administered surveys, and that telephone interviews should be shorter than face-to-face surveys. Otherwise, questionnaires for any given mode that are too long and complex could reduce response rates and affect data quality. Very crude rules of thumb suggest a maximum length of 60 minutes for a face-to-face interview, 20 minutes for a telephone survey, eight pages for a paper self-completion and 20 minutes for a web questionnaire. However, the actual length that is acceptable is likely to vary greatly, depending on other factors such as the saliency of the survey topic, the sponsor, the study population, and the use of respondent incentives. This is reflected in the literature which tends to find weak and inconsistent evidence for a negative relationship between questionnaire length and response rates, particularly for telephone and face-to-face surveys (Collins et al., 1988; Bogen, 1996). Clearer evidence of a relationship has been found for postal and web questionnaires (Heberlein and Baumgartner, 1978; Dillman, 1978; Yammarino et al, 1991; Bogen, 1996; Dillman, 2000; Galesic and Bosnjak, 2009).

The current length of the National Survey interview is 25 minutes. Although this is slightly longer than the recommended 20 minutes for telephone interviews and web questionnaires, these are only crude rules of thumb. Consequently the current questionnaire length may not be a problem for a survey that mixes face-to-face interviews, telephone interviews and web questionnaires. However, a mixed mode design that includes paper self-completion questionnaires would require the questionnaire to be shortened and simplified (e.g. complex question routing can result in reduced data quality).

An alternative strategy to shortening a questionnaire is to split the questionnaire into a core section which is asked of all respondents, and then have additional modules which are randomly assigned to different respondents. This strategy will allow the Welsh Government to collect data on numerous topics within one data year and/or to collect more in-depth data on specific topics, while keeping the questionnaire length for any given respondent to a minimum. The disadvantage is

that variables in modules that have been assigned to different sub-samples cannot be analysed together.

4 Recent developments

The recent review carried out by Betts and Lound (2010) for the Government Statistical Service provides a good overview of the evidence and experience of using mixed modes in the UK and elsewhere. However, this is an area which is changing fast with new technologies being more widely adopted and applied to survey research. Also the pressure to reduce data collection costs has compelled survey sponsors and researchers to seek solutions to some of the barriers to mixing modes of data collection. In this section I provide an overview of relevant developments that were either not covered in the Betts and Lound review or have taken place since their review was published.

4.1 Advances in data collection modes

4.1.1 T-ACASI

A recent development in telephone survey methods is the use of Telephone Audio Computer-Assisted Self-Interviewing (T-ACASI). In a T-ACASI survey, a telephone interviewer is used to call and recruit eligible respondents and possibly also administer a portion of the questionnaire. The phone call is then transferred to the T-ACASI system and the remaining questions are read out by the computer. The respondent provides the answers to the questions by pressing keys on a touchtone telephone[19]. T-ACASI is considered to be the telephone equivalent of audio-CASI (audio computer-aided self-interviewing) which is a computerised self-completion questionnaire used in face-to-face interviews for asking questions which could be prone to social desirability bias. Various studies have shown that using T-ACASI increases the reporting of stigmatised behaviours (e.g. certain types of sexual behaviour and illegal drug use) and socially desirable responses to attitude questions. But there is also some evidence of increased levels of breakoff during the interview. (Gribble et al, 2000; Villarroel et al, 2008; Villarroel et al, 2011; Harmon et al, 2009)

Some T-ACASI surveys use Automated Speech Recognition (ASR) rather than touchtone data entry for recording respondents' answers. The respondent replies to the pre-recorded questions by speaking (e.g. saying "Yes" or "No" over the telephone), and the software digitises and codes the spoken answer. T-ACASI

[19] This is also referred to as touchtone data entry (TDE).

surveys using ASR are still uncommon (primarily market research), largely because of the limited capabilities of ASR systems.

To the best of my knowledge, existing research has only compared T-ACASI with interviewer-administered telephone interviewing. I could not find any research that compared T-ACASI with other self-administered modes such as audio-CASI, web or postal. It is therefore unknown whether the use of T-ACASI within a mixed mode design will eliminate differences in measurement between telephone mode and self-completion modes.

4.1.2 M-CASI

M-CASI or mobile-CASI stands for computer-assisted self-interviewing using mobile phones. In its most simple form, respondents are invited to answer survey questions using text messaging (e.g. Yes or No). However, this mode is restricted to surveys that collect only a small number of data items. The increasing uptake of smartphones provides the opportunity to invite respondents to complete a web questionnaire on their mobile phones. However, web questionnaires that are normally administered on laptops and desktop PCs need to be adapted for smaller screens and smaller keyboards. Also a one-size-fits-all web questionnaire may not be feasible in the near future due to the variability among smartphone platforms. There has been some experimentation with questionnaire design strategies and visual design for mobile web surveys (Buskirk and Andrus, 2012; Callegaro, 2010; Peytchev and Hill, 2010) but the literature is scarce.

4.1.3 Web surveys mimicking other modes

Web surveys tend to be viewed as a single mode; i.e. a computerised self-completion mode that relies solely on the visual display of questions and answers. Although this view may correspond with the prevailing practice of web surveys, it does not reflect the true potential of web surveys. The web is an extremely versatile tool that can incorporate other communication media such as audio and video. Furthermore, it is possible to include interviewer elements in a web survey, either real or virtual. It is therefore possible to design web surveys that mimic other modes and possibly reduce differences in measurement between web and other modes (Couper, 2011). For example, it is now possible to conduct face-to-face interviews over the internet using webcams and Voice over Internet Protocol (VOIP). And a few studies have explored the feasibility of using animated interviewers (Cassell and Miller, 2008; Conrad et al, 2008; Malakhoff and Jans, 2011). Although the technical possibilities exist, it may be some time before they can be applied to general population surveys.

4.2 Developments in coverage and sampling

4.2.1 RDD incorporating mobile-only households

An alternative to sampling from PAF is Random Digit Dialling (RDD) with telephone contact. However, RDD surveys tend to exclude mobile-only households and it is very likely that this will bias the survey estimates. It is possible but not straightforward to extend RDD to cover mobile phones. A sample of working blocks of '07' numbers can be selected from the Ofcom database and the last four digits randomly generated to create a phone number. Such a method is likely to produce a very large proportion of non-working numbers which can be identified and removed using a computerised system (also known as 'pinging'). The hit rates are expected to be lower for a RDD sample of mobile phones because the average rate of working numbers in allocated blocks tends to be lower but there is no recent empirical evidence to support this for the UK. However, mobile RDD has not been used in social research in the UK, so there is no evidence on how much lower the hit rates would be. Furthermore, a RDD sample of mobile numbers for the National Survey would require extensive screening because it would be UK-wide (the prefix '07' is used for all UK mobile numbers and allocation of blocks is not based on geography).

There has been some research carried out in the USA on incorporating mobile-only households into RDD samples. Two main approaches have been identified, both supplement an RDD sample of landlines with an RDD sample of mobile phones but the treatment of the mobile sample differs. The first uses an RDD sample of mobile phones to screen for individuals in mobile-only households. Any individuals contacted by mobile who belong to households with access to a landline are screened out. The second approach attempts to interview all individuals contacted by mobile phone, regardless of their access to a landline. The AAPOR Cell Phone Task Force concluded that it could not yet recommend one approach over the other, and that the choice of approach should based on the particulars of the survey (AAPOR, 2010). RDD with dual frame sampling has not been used in the UK but there are plans to test the second approach at NatCen Social Research.

4.2.2 Probability-based online panels

Given the restrictions imposed by our reliance on the PAF for sampling the general population, the use of probability-based online web panels are worthy of further exploration. Considerable developments in probability-based web panels have been made in the Netherlands, USA, France and Germany.

The LISS panel (Longitudinal Internet Studies for the Social sciences) in the Netherlands is based on a true probability sample of households drawn from the population register by Statistics Netherlands. Those households with a listed telephone number are first contacted by phone whereas those without are contacted by a face-to-face interviewer. To ensure complete coverage of the Dutch household population, those without access to the internet are provided with a computer and broadband connection so that they can participate in the panel. The upfront investment in computers and broadband access would not be cost-effective for a cross-sectional survey. Incentives are used to maximise response rates. Panel members are invited to complete a 30 minute online questionnaire every month, including a core questionnaire which is repeated every year. The quality of the panel is regularly evaluated.

A probability-based online panel also exists in the USA (Knowledge Networks) and similar panels are now being set up in France and Germany. There is growing interest in setting up such a panel in the UK but acquiring the necessary funding is likely to be a significant barrier.

4.2.3 Population register

The Government's proposal to terminate the decennial Census may open up other possibilities for selecting an unbiased sample from the general population. A viable alternative to the Census may require a population register of some sort. If so, individuals could be selected directly from the population register thus negating the need for interviewer involvement to carry out respondent selection. Named individuals could be invited by post to complete a postal or web questionnaire. Those who do not complete a postal or web questionnaire could be followed up using more expensive data collection modes in a sequential mixed mode design. Even if the UK was to adopt a population register, a possible barrier to this solution would be access restrictions to the population register for sampling purposes.

4.3 Web survey response rates

As discussed in section 3.2.2, a sequential mixed mode design starting with the cheapest mode and finishing with face-to-face interviewing for the most reluctant respondents is the most promising design for maximising response while cutting costs. The success of this strategy, however, depends on a significant proportion of the sample completing the questionnaire in the cheaper modes.

Statistics Netherlands' experience of inviting sample members by post to go online and complete web questionnaires is promising - they manage to achieve response rates between 25% and 30% to the web questionnaire after two postal reminders (discussed further in section 4.4). However, our own experience at NatCen Social

Research suggests that similar response rates may not be realised in the UK without using unconditional incentives and additional reminders. NatCen Omnibus respondents with internet access who had agreed to be recontacted but had not provided an email address were sent a postal invitation and a postcard reminder to go online and complete a web questionnaire. Although these respondents had already taken part in the Omnibus survey and expressed their willingness to be recontacted, the response rate was only 5% without an incentive compared to 27% when a £5 gift voucher was sent with the postal invitation (as well as a postcard reminder having been sent to both groups)[20]. An additional postal reminder (after the postcard reminder) increased the response rate further to 39%. Further experiments are needed to determine whether it is feasible to get a sufficient proportion of 'fresh' sample members to go online using postal invitations, reminders and incentives.

4.4 Recent developments in National Statistics Institutes

The Welsh Government is not on its own in wishing to explore the use of mixed modes in delivering social surveys. Many national governments and their statistical institutes are similarly eager to embrace mixed mode data collection to address increasing concerns about cost and quality. Betts and Lound (2010) provide an overview of the research and development being carried out in this field.

Since the publication of the Betts and Lound (2010) review, the ESSnet (a network of European Statistical Societies organisations) launched a new project in January 2012 entitled "Data collection for social surveys using multiple modes" (DCSS), with a workshop being held in advance in September 2011. The objective of the project is to facilitate web data collection for social statistics in European Statistical Societies. The scope of the project has been extended to cover multi-mode data collection because web data collection is not (yet) a viable stand-alone mode for official social statistics. The programme is focussing on the Labour Force Survey (LFS) which is carried out by all participating countries[21]. The priorities include:

1. Web data collection tools covering questionnaire design, respondent access, questionnaire testing, and software;
2. Multiple mode data collection design covering the design of the multiple mode data collection system, fieldwork organisation, and processing and analysis issues;

[20] It should be noted that these results are not based on a true experiment. The study included respondents who had taken part in one of two separate rounds of the NatCen Omnibus Survey. The first round respondents were sent the postal invitation without an incentive, whereas the second round respondents received a £5 gift voucher with their postal invitation.

[21] Workshop participants included all EU countries except Greece and Romania, as well as Norway, Iceland, Turkey, Switzerland and Kosovo.

3. Development of a computer assisted coding system which will involve creating a database of occupations with ISCO-08 codes, developing algorithms for the selection of occupations to match user input, set up structures for the maintenance of the database, and creating shared IT tools.

The UK Office for National Statistics (ONS) is participating in the ESSnet project. ONS are planning to use web data collection in parallel with the 'usual' LFS. This will allow mode effects to be assessed without damaging the regular estimates. A number of issues need to be resolved such as the identification of non-eligible addresses, how to deal with multi-household addresses, and changing of modes during web data collection. The current proposal is to implement the sampling and estimation methods later in 2012-13 and to carry out the parallel run in 2013-14.

Statistics Netherlands has been carrying out research into mixed mode surveys for many years as part of its large-scale re-engineering programme of Dutch household surveys. It has therefore made more progress in this area than most other countries. Its Health Survey is completely mixed mode since last year, and the Dutch Labour Force Survey has switched to mixed mode this year. Both surveys use a sequential mixed mode design. For the Health Survey, a letter is sent to the named respondent (selected from the population register) inviting them to complete a web questionnaire. It is slightly more complicated for the Dutch Labour Force Survey which invites the household (all household members aged 15 years or older) by letter to complete the web questionnaire. Two reminders are sent before the unproductive cases are moved to telephone interviewing (those with matched telephone numbers) and face-to-face interviewing (those without matched telephone numbers). Response to the web questionnaire fluctuates between 25% and 30% - reasons for the fluctuation are unknown but it has been speculated that this could be caused by problems with the server. Willingness to enter the LFS panel was considerably lower for those who completed the web questionnaire compared to those who had been interviewed face-to-face (66.4% and 88.2% respectively) and preliminary results show significant mode effects for labour status. The data collection period for mixed mode data collection is three months – one month longer than it used to be. (Mars et al, 2011)

The 2011 Census was the first in England and Wales to have the option to be completed online. Overall, 13.4% of Census 2011 returns in Wales were completed online. The online completion rates varied by Local Authority, ranging from 9.5% in Blaenau Gwent to 18.2% in Cardiff. The rate for England and Wales[22] varied by age (using the age of person 1 on the questionnaire as a proxy

[22] These figures were not available for Wales only.

for who completed the questionnaire for the household), with nearly 30% of respondents aged 26-35 choosing to complete online; similar proportions for those aged 17-25 and 36-45; compared with less than 10% for those aged over 65. (ONS, 2012)

5 What makes a good mix for the National Survey?

The dominance of face-to-face interviewing for government surveys of the general population can be ascribed to the lack of suitable sampling frames that provide full coverage. Unlike Scandinavian countries and the Netherlands, the UK does not have a population register. Directories that provide adequate coverage of telephone numbers and email addresses for the general population do not exist. Only the PAF provides more or less full coverage but the sampled units are addresses rather than individuals, thereby requiring the involvement of a face-to-face interviewer for enumeration of the household(s) at the sampled address and respondent selection within the selected household(s). This is a severe constraint on the range of mixed mode options that could be considered for the National Survey.

In this section I will present three mixed mode options for the National Survey. Each has its advantages and its disadvantages, with perhaps only one of these being a 'serious' contender to a single mode face-to-face survey. In addition to these three mixed mode options, I also present two alternative single mode designs. Neither of these could really be considered viable options for the National Survey at this stage. This range of options is included in the review to provide the reader with context and comparison.

5.1 Option 1: Sequential design with drop-off questionnaire at contact stage

This mixed mode design shares features with both the 2009-10 pilot for the National Survey and the current design of the National Survey. Sample selection would be identical to both its current design and the 2009-10 pilot with a sample of addresses being selected from the PAF thus ensuring complete coverage of the household population in Wales. The mix of modes would be similar to the 2009-10 pilot survey, i.e. face-to-face and self-completion modes.

The face-to-face mode could be restricted to household enumeration and respondent selection; or it could also be used to collect household data from any responsible adult in the household. The latter approach would be preferable to ensure the collection of basic information about the household (and preferably a few key items for the Welsh Government) in case of non-response to the individual questionnaire. If the contacted household member is also the selected respondent,

there is the option of allowing the interviewer to administer the individual questionnaire in a face-to-face interview.

A self-completion questionnaire for the selected respondent could be left at the address to be completed by the respondent at a time convenient to them (if not already administered by the interviewer). Interviewers could arrange a date to pick up the completed questionnaire, or the respondent could be asked to return the questionnaire by post. Interviewer pick-up would produce a higher response rate than postal return, but is more expensive.

In addition to providing paper self-completions it would also be possible to provide a URL address and password for online completion of the questionnaire. Evidence suggests that only a small proportion of respondents will complete an online questionnaire when provided with a paper self-completion, and this concurrent approach is unlikely to produce a higher response rate overall than not offering an online option. The online option could be made more attractive by offering incentives, which be justified on the grounds that the incentive is less costly than sending an interviewer to the address to collect the completed paper questionnaire (i.e. more efficient use of public money).

Response could be boosted further by pursuing face-to-face interviews among those who have failed to complete the paper self-completion and online questionnaire. The interview could be carried out when the interviewer returns to pick-up the questionnaire or, if self-completions are being returned by post, a separate visit by the interviewer.

At the first contact stage when the interviewer carries out household enumeration and respondent selection, telephone numbers and email addresses could be collected which can be used for reminders, arranging pick-up times, email invitations for online completion (which will boost response to the web questionnaire), and even telephone data collection among those who have not responded via other modes.

If this design is restricted to face-to-face interviews for the household data and paper self-completion for the individual questionnaires, then there is no risk of differences in measurement. Including other modes to boost response increases the risk of differences in measurement. Offering the option of online completion of the individual questionnaire should be acceptable, provided good questionnaire design principles have been followed and questions have been pre-tested for measurement equivalence. The risk of differences in measurement across modes is far greater when allowing face-to-face or telephone interviewing at the drop-off stage or at the refusal conversion stage. Show cards and self-completion modules can be used in the face-to-face interview to minimise differences in measurement.

The scope for cost savings is limited for this mixed mode design because of the involvement of a face-to-face interviewer to carry out household enumeration and respondent selection at the first contact. Additional involvement of interviewers at the refusal conversion stage and possible pick-up stage will further reduce the cost savings. Cost savings will also depend on getting a substantial proportion of the sample to complete the web questionnaire.

If the Welsh Government opted for this mixed mode design for the National Survey, I would recommend the following development work:

- Testing methods for encouraging respondents to provide email addresses and telephone numbers at the contact stage;
- Testing methods for boosting response to the web questionnaire;
- Experimentation with the number, timing and mode of reminders;
- Development and testing of a questionnaire that will produce comparable data in paper self-completion, web questionnaire, telephone interview (if being used) and face-to-face interview (if being used).

5.2 Option 2: Sequential design with postal questionnaire at first contact

It has been noted that cost savings are limited with Option 1 because of the necessity to use a face-to-face interviewer for household enumeration and respondent selection. Considerable cost savings could be made if it was possible to send postal questionnaires to sampled PAF addresses with instructions for respondent selection.

The Kish method, which is the well-established respondent selection technique used on face-to-face surveys, requires a full listing of eligible household members and then random selection of one household member using a respondent selection table known as the Kish grid (Kish, 1965). This technique would be too complicated for a postal survey. Allowing any eligible household member to complete the questionnaire will introduce bias because it is most likely to be completed by the person who opens the post. Requiring all eligible household members to complete the questionnaire is burdensome and logistically difficult (e.g. how many questionnaires should be sent to sampled addresses?). An easier and less intrusive technique is the “Next Birthday” method (or alternatively the most recent birthday) which is a quasi-probability method. It is commonly used for telephone surveys and it has been shown to produce similar sample compositions as the Kish method (Tipping and Nicolaas, 2001). Although this technique has occasionally been used for postal surveys based on PAF samples, there is no evidence of its effectiveness in producing a random sample.

The correct application of the birthday technique in postal surveys depends heavily on the person who opens the post to read the selection instructions. Furthermore this person must understand the concept of random selection and the need to comply with the instructions. The risk is that the person with the next/last birthday is not willing to complete the questionnaire or not available, and the questionnaire is then completed by another household member who is available (more likely to be the person who opened the mail) or has a specific interest in the survey topic. Nonetheless, it would be worthwhile to explore the effectiveness of the next/last birthday technique for postal surveys based on PAF samples because the cost savings could be considerable if this technique is proven feasible.

As for Option 1, a sample of addresses is selected from the PAF. Postal questionnaires with respondent selection instructions are sent to sampled addresses, addressed to the resident or occupier. It would also be possible to provide a URL address and password for online completion of the questionnaire. As noted before, evidence suggests that only a small proportion will complete an online questionnaire in a concurrent design. Response to the postal questionnaire can be boosted using postal reminders and incentives. A larger incentive can be offered to those who complete the questionnaire online. For example, an unconditional incentive could be sent to all with the postal questionnaire plus the promise of an additional incentive if the questionnaire is completed online. However, the use of a promised incentive will increase the risk that other household members will complete the questionnaire if the selected respondent is unwilling to do so. Response can be increased further by sending face-to-face interviewers to postal/web non-respondents. Risk of measurement differences between postal/web modes and face-to-face modes can be minimised by using show cards and self-completion modules within the face-to-face interview.

Compared to Option 1, this design will achieve similar population coverage but at a much lower cost. The response rate should also be similar if face-to-face interviewing is used among the postal/web non-respondents. Risk of measurement differences should be more or less similar (indeed this risk for Option 2 would be lower if telephone interviewing is used for refusal conversion in Option 1) but it should be noted that the proportion completing a self-completion questionnaire is likely to be higher in Option 1 compared to Option 2 – particularly if interviewers collect the self-completion questionnaires in Option 1. But Option 2 has the additional risk of self-selection bias.

The following development work should be carried out before this design is adopted for the National Survey:

- Evidence gathering on the impact of self-selection at postal/web stage on sample composition and survey estimates;

- Testing methods for dealing with multiple households at sampled PAF addresses;
- Testing methods for boosting response to the web questionnaire (without increasing the risk of self-selection);
- Experimentation with the number, timing and mode of reminders;
- Development and testing of a questionnaire that will produce comparable data in paper self-completion, web questionnaire, and face-to-face interview (if being used).

5.3 Option 3: Sequential design with telephone (RDD dual frame) at contact stage

At the moment, the PAF is the only sampling frame that provides more or less full coverage of the Welsh household population, which seriously restricts the mixed mode options that could be considered for the National Survey. The only other method that could possibly be used to select a probability sample is Random Digit Dialling (RDD) with dual frame samples for incorporating mobile-only households (see section 4.2.1 for further details). This method has not yet been used in the UK but there are plans to test this approach at NatCen Social Research.

If RDD dual frame sampling proved to be feasible, then it would be possible to use telephone interviewers at the contact stage for household enumeration and respondent selection. The telephone interviewer could encourage the selected respondent to complete the questionnaire online or alternatively conduct a telephone interview. To encourage completion of the web questionnaire, respondents could be promised an incentive. For those who agree to complete the web questionnaire, the telephone interviewer would collect the respondent's email address so that invitations and reminders can be sent by email with embedded links to the web questionnaire. It is very likely that a significant proportion of those who agree to complete the questionnaire online fail to do so. They could either be sent a postal questionnaire (if postal address was collected at first contact) or the telephone interviewer could call in order to complete the questionnaire by phone.

The advantages of this mixed mode design include good population coverage (99% of Welsh adults live in households with either a fixed landline or a mobile number) and it would be a lot cheaper than mixed mode designs that involve face-to-face interviewers. But the response rate for this design is likely to be a lot lower than mixed mode designs that include face-to-face interviewing[23]. Furthermore, the combined use of web questionnaires (visual stimuli, no interviewer) and

[23] For illustrative purposes only: If we assume that the response rate for a mixed mode design which includes face-to-face interviews is 60%, then the response rate for the RDD/web design would be about 40%-45% (all other things being equal). Note that it is not possible to send advance letters on RDD surveys, so we would expect the response rate to a RDD survey to be lower than a telephone sample of named individuals with a postal address.

telephone interviewing (aural stimuli, interviewer) carries a higher risk of mode effects than other combinations. And of course this design depends completely on the feasibility of using RDD with dual frame sampling to include mobile-only households in Wales. As mentioned in section 4.2.1, a RDD sample of mobile numbers for the National Survey would require extensive screening because it would be UK-wide. Detailed information about telephone ownership also needs to be collected in order to calculate selection probabilities and to weight the data accordingly. The weighting of RDD samples, particularly those that include mobile phone numbers as well as landlines, is complicated and there is no consensus on how this should be done (AAPOR, 2010).

Prior to this mixed mode design being considered for the National Survey, the following work should be carried out:

- The feasibility of RDD dual frame samples for Wales needs to be determined, including the feasibility of screening a RDD sample of mobile numbers for Welsh households;
- Testing methods for encouraging respondents to provide email addresses (and possibly postal address) at the contact stage;
- Testing methods for boosting response to the web questionnaire;
- Testing methods for increasing response to telephone interview (e.g. incentives, telephone introductions, call scheduling);
- Experimentation with the number, timing and mode of reminders;
- Development of optimal call scheduling;
- Development of questions that will provide info about selection probabilities;
- Development and testing of a questionnaire that will produce comparable data in web questionnaire and telephone interview;
- Development of an appropriate weighting strategy.

5.4 Option 4: Single mode design with RDD (dual frame)

This option does not involve mixing modes within a particular year of the survey, but it does involve changing the mode from one year (face-to-face) to the next (telephone). This relates to scenario 4 in section 3.1.

As for Option 3, the population coverage for this design should be good, conditional on the feasibility of incorporating mobile-only Welsh households. It is likely to be more expensive than Option 3 (depending on the proportion of the sample that completes the online questionnaire in Option 3) but still cheaper than mixed mode designs that include face-to-face interviewing. A single mode design is a lot easier to design and manage than a mixed mode design. Potential differences in measurement between face-to-face and telephone interviewing are only relevant for time series and comparisons with other surveys. The response

rate would be similar to that for Option 3, but a lot lower than the response rates that can be achieved with a mixed mode design that includes face-to-face interviews. As for Option 3, mobile numbers that belong to people who live outside Wales will have to be screened out as well as other ineligible numbers (e.g. non-residential numbers). The calculation of selection probabilities is complex because of the likelihood of households being contacted using a variety of household-based and individually owned phone numbers.

Development work required for a single mode design with RDD dual frame sampling:

- The feasibility of RDD dual frame samples for Wales needs to be determined, including the feasibility of screening a RDD sample of mobile numbers for Welsh households;
- Testing methods for increasing response to telephone interview (e.g. incentives, telephone introductions, call scheduling);
- Experimentation with the number and timing of reminders;
- Development of optimal call scheduling;
- Development of questions that will provide info about selection probabilities.

5.5 Option 5: Online panel based on probability sample

This survey design involves the greatest departure from the current design of the National Survey. It would entail changing the design from a repeated cross-sectional survey into a longitudinal panel.

A random probability sample would be selected from the PAF with initial contact being made by a face-to-face interviewer. The interviewer would carry out household enumeration, respondent selection (although this design would also be suitable for including all adults and even children within the household) and recruitment to a panel. The primary data collection mode for the panel would be web. There are two main strategies for including those without access to the internet: (1) those without access to the internet could be asked to complete questionnaires in a different mode (e.g. postal, telephone), and (2) those without access to the internet could be provided with a computer and broadband connection.

The latter is very similar to the LISS model and Knowledge Networks Panel mentioned in section 4.2.2. This model is preferable to the former in that all data are collected using one mode, thus avoiding differences in measurement as well as being easier to manage.

The advantages of an online panel include speed of data delivery; low unit costs and low marginal costs (Dillman and Bowker, 2001); avoidance of interviewer

effects (Chang & Krosnick, 2010); relatively low burden on respondents and potential for higher response rates particularly among some sub-groups (Couper, 2000; Couper 2008; Smyth & Pearson, 2011). But it involves considerable upfront investment to provide computers and broadband access to those households without. Incentives are also required at the recruitment stage and throughout the lifetime of the panel to maintain acceptable response rates. Refreshment samples would have to be added to mitigate the effects of attrition.

The investment and frequency of data collection exceeds what is currently required for the National Survey but an online panel of this type could serve as a survey platform for multiple clients, including academics as well as Welsh Government policy teams. It could provide answers to questions of immediate interest more quickly and provide information on topics that do not require a full survey. In addition to providing data for cross-sectional research at a reduced cost, the panel will also provide longitudinal data that can be used to measure change at the level of the individual, to measure and understand trends, and to assess impacts.

5.6 Conclusion

Option 1 (sequential design with drop-off paper questionnaire) is the most feasible mixed mode design for the National Survey. This mixed mode design will achieve similar coverage and response to its current design. But the scope for reducing data collection costs is limited because interviewers are still required to travel to the sampled addresses in order to carry out random respondent selection. The extent of the cost savings will depend on the proportion of the sample that can be persuaded to complete the questionnaire online or to return a paper self-completion by post, thus reducing the amount of interviewer travel after first contact and interviewer fees for face-to-face interviews with non-respondents to web/postal questionnaire. The risk of differences in measurement can be minimised by restricting the modes to the concurrent offer of paper self-completion questionnaire and web questionnaire, followed by face-to-face interviews with show cards and self-completion modules for those who do not complete the paper or web self-completion. Questions should be designed according to uni-mode principles and pre-tested.

Considerable cost savings could be achieved by replacing the interviewer with postal instructions for respondent selection (Option 2). However, this design is likely to suffer from self-selection bias and will achieve a lower response rate. A move to this design should be based on empirical evidence that self-selection bias and marginal non-response bias (relative to its current design) are within acceptable limits.

Options 3 (telephone and web) and 4 (only telephone mode) replace the PAF sample frame with RDD sampling but this approach will only achieve acceptable population coverage if it is feasible to include Welsh mobile-only households. Even if this is possible, this design will suffer from response rates that are a lot lower than for Options 1 and 2. Furthermore, Option 3 carries higher risk of mode effects due to the combined use of telephone (aural stimuli and interviewer) and web modes (visual stimuli and no interviewer).

Option 5 (online panel based on a random probability sample selected from PAF) is a far more radical and forward thinking approach to the problem of collecting data from a representative sample of the Welsh population while controlling data collection costs. This option requires considerable upfront investment and it therefore would be more feasible as a longer-term solution for survey data collection for multiple clients.

6 Managing change from unimode to mixed mode

A switch from face-to-face interviewing to a mixed mode design could affect the data being collected because different types of people are more likely to participate in different modes (selection effect) and people may answer questions differently depending on the mode of data collection (mode effect). This could result in more biased survey estimates (single year as well as aggregate estimates), observed changes that do not correspond with real changes in the Welsh population, and reduced comparability with other surveys. Consequently if the National Survey is to change its design from face-to-face interviews to a mixed mode design, then it will be important to manage this change.

There are two main strategies for managing a change to a mixed mode design: (1) minimise error, and (2) quantify and possibly adjust for error.

As discussed in sections 3.2.1 and 3.2.2, non-coverage error and non-response bias can be avoided by retaining PAF as a sampling frame and using a sequential mixed mode design which includes face-to-face interviews. Methods for minimising differences in measurement across modes have been discussed in section 3.2.3. I have suggested that at the moment the most promising method for minimising differences in measurement is uni-mode design. However, it may not always be possible to design questions in a way that assures measurement equivalence across survey modes. This can be particularly problematic when adapting existing questions designed for one particular mode so that they can be used in other modes. It is therefore recommended that a switch from face-to-face interviewing to mixed mode data collection on the National Survey should be preceded by development work, field experiments and pre-testing to identify which questions can (and cannot) be asked in different modes.

Despite best efforts to reduce the risk of measurement differences, it is unlikely that this risk will be eliminated altogether. It is therefore desirable to collect information that can be used to quantify and adjust for measurement differences. One approach is to run the old face-to-face design in parallel with the new mixed mode design for a period of time. Recent work in this area by Vannieuwenhuyze et al (2011) has demonstrated that this approach could be used to produce comparable measurements across modes by statistically adjusting for measurement differences. Furthermore, it is a safe method of transition.

Another strategy that is often suggested is the use of embedded experiments to quantify the mode effect. It has been claimed that embedded experiments with randomised allocation of respondents to modes can be used to disentangle the mode effect from differences in sample composition across the modes (de Leeuw and Hox, 2011). Embedded experiments are integral to the plans for mixed mode data collection at Statistics Netherlands and the Office for National Statistics. It is hoped that the information can be used to statistically adjust the data for mode effects. However, research in this area is still in its infancy and its feasibility has been questioned.

References

AAPOR Cell Phone Task Force (2010) New Considerations for Survey Researchers when Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers.

Balden, W. (2004). Multi-mode data collection: Benefits and downsides. Paper presented at the 2004 Conference of the Great Lakes, Chapter of the Marketing Research Association, Cancun, Mexico.

Betts, P. and Lound, C. (2010). The Application of Alternative Modes of Data Collection in UK Government Social Surveys. Literature Review and Consultation with National Statistical Institutes. United Kingdom: Government Statistical Service.

Bogen, K. (1996). The effect of questionnaire length on response rates - a review of the literature. Proceedings of the Survey Research Methods Section, Alexandria: American Statistical Association, pp. 1020-1025.

Bryson, C. and Huskinson, T. (2012). The National Survey for Wales: question comparability study. A report prepared for the Welsh Government.

Buskirk, T. and Andrus, C. (2012). Smart Surveys for Smart Phones: Exploring Various Approaches for Conducting Online Mobile Surveys via Smartphones. Survey Practice, February: www.surveypractice.org.

Callegaro, M. (2010). Do You Know Which Device Your Respondent Has Used to Take Your Online Survey? Survey Practice, December: www.surveypractice.org.

Callcredit Information Group (2012). A Change in the Tides: The 2012 Electoral Roll.

Campanelli, P., Nicolaas, G., Jäckle, A., Lynn, P., Hope, S., Blake, M. and Gray, M. (2011). A classification of question characteristics relevant to measurement (error) and consequently important for mixed mode questionnaire design. Presented 11 October 2011 at the Royal Statistical Society, London, UK. Last accessed on 19 March 2012 at http://www.surveynet.ac.uk/SDMI/mixedmodes2011/handoutFinal.pdf

Cassell, J. and Miller, P. (2008). "Is It Self-Administration If the Computer Gives You Encouraging Looks?" Envisioning the Survey Interview of the Future, edited by Conrad, F.G. and Schober, M.F. , John Wiley & Sons., pp. 161-178.

Chang, L. and Krosnick, J. (2010). Comparing Oral Interviewing with Self-Administered Computerized Questionnaires: An Experiment. Public Opinion Quarterly; 74:154-167.

Christian, L. and Dillman, D. (2004). The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Questions. Public Opinion Quarterly; 68: 57-80.

Christian, L., Dillman, D., and Smyth, J. (2008). "The effects of mode and format on answers to scalar questions in telephone and web surveys". in J. Lepkowski, C. Tucker, M. Brick, E. de Leeuw, L. Japec, P. Lavrakas, M. Link, R. Sangster (Eds.) Advances in Telephone Survey Methodology, John Wiley, New York, pp. 250-275.

Collins, M., Sykes, W., Wilson, P. and Blackshaw, N. (1988), Nonresponse: The UK Experience, in: Groves et al. Telephone Survey Methodology, Wiley and Sons.

Conrad, F., Schober, M., Jans, M., Orlowski, R., Nielsen, D., and Levenstein, R. (2008). Features of Animacy in Virtual Interviewers: Dialogue Capability and Visual Realism. American Association for Public Opinion Research Annual conference, New Orleans.

Converse, J. and Presser, S. (1986). Survey Questions: Handcrafting the Standardized Questionnaire. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-063. Thousand Oaks, California. Sage Publications.

Couper, M. (2000). Web Surveys: A Review of Issues and Approaches. Public Opinion Quarterly; 64: 464-494.

Couper, M. (2008). "Technology and the Survey Interview/Questionnaire.'' In Envisioning the Survey Interview of the Future, edited by Michael F. Schober and Frederick G. Conrad. New York: Wiley, pp. 58–76.

Couper, M. (2011). The future of modes of data collection. Public Opinion Quarterly; 75: 889–908

de Leeuw, E. (2005). To mix or not to mix? Data collection modes in surveys. Journal of Official Statistics ; 21(2): 1-23.

de Leeuw, E. (2008). "Choosing the Method of Data Collection.'' In Social Research and the Internet, edited by Marcel Das, Peter Ester, and Lars Kaczmirek. New York: Taylor and Francis, pp. 113–35.

de Leeuw, E. and Hox, J. (2011). Internet surveys as part of a mixed-mode design. In M. Das, P. Ester and L. Kaczmirek (Eds.). Social and Behavioral Research and the Internet. NY: Routledge, pp. 45-76.

Dillman, D. (1978). Mail and telephone surveys: The total design method. New York: Wiley-Interscience.

Dillman, D. Brown, T. Carlson, J. Carpenter, E. Lorenz, F. Mason, R. Saltiel, J. and Sangster, R. (1995). Effects of category order on answers to mail and telephone surveys. Rural Sociology; 60: 674–687.

Dillman, D. (2000). Mail and Internet Surveys: The Tailored Design Method (2nd ed.). New York: Wiley

Dillman, D. and Bowker, D. (2001). "The Web questionnaire challenge to survey methodologists." In: Dimensions of Internet science, edited by U.-D. Reips and M. Bosnjak, Lengerich: Pabst Science Publishers.

Dillman, D. (2007). Mail and internet surveys: The tailored design method. Hoboken, NJ: John Wiley and Sons.

Dillman, D., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., and Messer, B. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response and the internet. Social Science Research; 39: 1-18.

Elliot, D., Rainford, L. and Eldridge, J. (2006). Interview Mode Effects in the UK Local Labour Force Survey. Paper presented at the Second International Telephone Survey Methodology Conference Miami, January 2006.

Fowler, J. (1995). Improving survey questions: Design and evaluation (Vol. 38). Thousand Oaks, CA: Sage Publications.

Fuchs, M. (2009). Asking for Numbers and Quantities: Visual Design Effects in Paper and Pencil Surveys. International Journal of Public Opinion Research; 21: 65-84.

Galesic, M. and Bosnjak, M. (2009). Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. Public Opinion Quarterly; 73(2): 349-360.

Gribble, J., Miller, H., Cooley, P., Catania, J., Pollack, L., Turner, C. (2000). The impact of T-ACASI interviewing on reported drug use among men who have sex with men. Substance Use and Misuse; 35(6-8):869-90.

Harmon, T., Turner, C., Rogers, S., Eggleston, E., Roman, A., Villarroel, M., Chromy, J., Ganapathi, L. and Li, S. (2009). Impact of T-ACASI on Survey Measurements of Subjective Phenomena. Public Opinion Quarterly; 73: 255-280.

Heberlein, T. and Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. American Sociological Review; 43: 447-462.

Hochstim, J. (1967). A critical comparison of three strategies of collecting data from households. Journal of the American Statistical Association; 62: 976-989.

Holbrook, A., Green, M. and Krosnick, J. (2003). "Telephone versus face-to-face interviewing of national probability samples with long questionnaires: comparisons of respondent satisficing and social desirability bias." Public Opinion Quarterly; 67:79-125.

Holmberg, A. and Lorenc, B. (2008). Understanding the Decision to Participate in a Survey and the Choice of the Response Mode. Paper presented at the European Conference on Quality in Official Statistics (Q2008) Rome, 8-11 July 2008.

Kish, L. (1965). Survey Sampling, New York, NY: John Wiley and Sons.

Jäckle, A., Campanelli, P., Lynn, P., Nicolaas, G. and Hope, S. (2011) How and when does the mode of data collection affect survey measurement? Presentation at “Mode Effects on Survey Measurement, one day conference” at the Royal Statistical Society, London, UK, 11 October 2011.
http://www.surveynet.ac.uk/SDMI/mixedmodes2011/pres1.pdf

Japec, L. (1995). Issues in Mixed mode Survey Design, Stockholm: Statistics Sweden, March 1995.

Kreuter, F., Presser, S. and Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. Public Opinion Quarterly; 72:847–65.

Lozar Manfreda, K. and Vehovar, V. (2002). “Mode effect in web surveys“, In the proceedings from The American Association for Public Opinion Research (AAPOR) 57th Annual Conference, 2002, http://www.amstat.org/sections/srms/Proceedings/y2002/files/JSM2002-000972.pdf

Lozar Manfreda, K., Bosnjak, M., Berzelak, J., Haas, I. and Vehovar, V. (2008). Web Surveys versus Other Survey Modes: A Meta-Analysis Comparing Response Rates.'' International Journal of Market Research; 50:79–104.

Malakhoff, L. and Jans, M. (2011). Toward Usage of Avatar Interviewers in Web Surveys. Survey Practice, June: www.surveypractice.org.

Mars, G., Chadli, R., Janssen, B. and Cuppen, M. (unpublished). Introducing web interviewing in the Labour Force Survey at Statistics Netherlands: a pilot.

Millar, M., O'Neill, A. and Dillman, D. (2009). Are Mode Preferences Real? Technical Report 09-003, Washington State University Social and Economic Sciences Research Center. Washington State University: Pullman.

Miller, N. and Lewis, K. (2011). National Survey of Adult Skills in Wales 2010, A report prepared for the Welsh Government Social Research, Number 27/2011.

Ofcom, (2011). Communications Market Report: Wales. Last accessed on 31 March 2012 at http://stakeholders.ofcom.org.uk/binaries/research/cmr/cmr11/CMR_2011_Wales.pdf

Office for National Statistics (2012). 2011 Census Update: Providing the online Census, February 2012. Last accessed on 29 March 2012 at http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/the-2011-census-project/2011-census-updates-and-evaluation-reports/index.html

Peytchev, A., and Hill, C. (2010). Experiments in mobile web survey design: Similarities to other modes and unique considerations. Social Science Computer Review; 28: 319-335.

Saris, W., Revilla, M., Krosnick, J. and Shaeffer, E. (2010). Comparing Questions with Agree/Disagree Response Options to Questions with Item-Specific Response Options. Survey Research Methods; 4(1): 61-79.

Schonlau, M., Zapert, K., Simon, L., Sanstad, K., Marcus, S., Adams, J., Spranca, M., Kan, H., Turner, R. and Berry, S. (2003). A comparison between responses from a propensity-weighted web survey and an identical RDD survey. Social Science Computer Review; 22(1): 128-138.

Shih, T-H. and Fan, X. (2008). Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis, Field Methods; 20(3): 249–271.

Siemiatycki, J. (1979). A comparison of mail, telephone, and home interview strategies for household health surveys. American Journal of Public Health; 69: 238-45.

Smyth, J. and J. Pearson (2011). “Internet survey methods: a review of strengths, weaknesses, and innovations.” In: Social and Behavioral Research and the Internet.

Tipping, S. and Nicolaas, G. (2001). Respondent selection procedures for telephone surveys. Survey Methods Newsletter; 21(1): 4-7, NatCen Social Research.

Tipping, S. and Nicolaas, G. (2006). In search of a population sampling frame for UK postal surveys. Survey Methodology Bulletin, Special Edition No. 58 August 2006. Office for National Statistics.

Tourangeau, R. (1984). “Cognitive science and survey methods”. In T. B. Jabine, M. Straf, J. Tanur and R. Tourangeau (Eds.), Cognitive aspects of survey methodology: Building a bridge between disciplines (pp. 73-100). Washington, DC: National Academy Press.

Tourangeau R. and Yan, T. (2007). Sensitive Questions in Surveys, Psychological Bulletin; 133(5): 859-883.

Vannieuwenhuyze, J., Loosveldt, G. and Molenberghs, G. (2011). A Method for Evaluating Mode Effects in Mixed-Mode Surveys. Public Opinion Quarterly; 74(5):1027–45.

Villarroel, M., Turner, C., Rogers, S., Roman, A., Cooley, P., Steinberg, A., Eggleston, E., Chromy, J. (2008). T-ACASI reduces bias in STD measurements: the National STD and Behavior Measurement Experiment. Sexually Transmitted Diseases; 35(5):499-506.

Villarroel, M., Turner, C., Eggleston, E., Al-Tayyib, A., Rogers, S., Roman, A., Cooley, P. and Gordek, H. (2011). Same-gender sex in the United States impact of T-ACASI on prevalence estimates. Public Opinion Quarterly; 70: 166–196.

Voogt, R., and Saris, W. (2005). Mixed mode designs: Finding the balance between nonresponse bias and mode effects. Journal of Official Statistics; 21: 367-387.

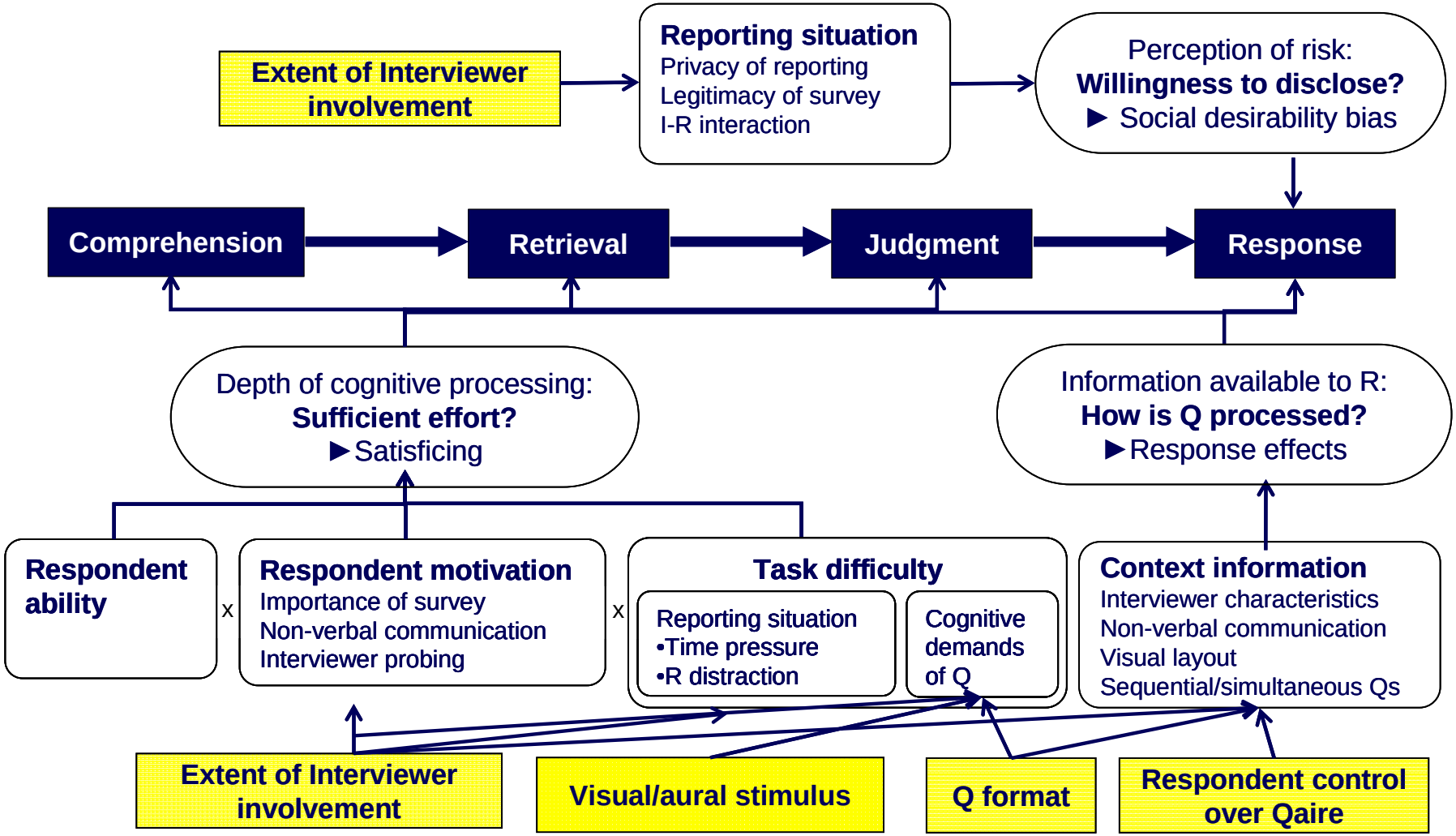
Yammarino, F.J., Skinner, S.J., and Childers, T.L. (1991). Understanding mail survey response behavior: A meta-analysis. Public Opinion Quarterly; 55: 613-639.

Appendix A Pros and cons of different modes

Design Parameter	Face to face	Telephone	Postal	Web
Cost of data collection	Usually most expensive method.	Usually around 40%-50% of face to face cost for same interview.	Relatively cheap	Cheapest
Amount and type of resources required	Specialised fieldworker skills and field-force management resources needed.	Specialised interviewer skills and management resources needed.	Operational resources for managing mail-outs and returned questionnaires.	Programming and web hosting resources needed.
Timetable considerations	May require several months unless respondents are easily accessible or 'captive'.	A potentially fast mode of data collection, but depends on respondent availability.	With response reminders, may require two months.	Usually the fastest mode of data collection, but likely to require postal/email reminders to achieve acceptable response.
Operational control	Best for control of field sampling and data collection.	Good for interviewer supervision, but respondent tolerance may be limited.	Few means of controlling how questionnaires are completed.	Question routing and ordering can be controlled by programming.
Amount/complexity of data to be collected	Best/mandatory for long and complicated questionnaires.	Limitations on length and data collection complexity compared with face to face.	Weaker for groups with poor literacy or motivation, but can be good for experts.	Requires computer and language literacy. Complex routing can be programmed into web questionnaires.
Data Quality	Best for complex topics and issues. Computer assistance improves quality. May incur interviewer effects.	Good for simple factual and attitudinal questions. Computer assistance improves quality. Interviewer effects less likely.	Worst for missing data, routing errors, misunderstandings.	May include prompts if questions are missed and data validation can be programmed into web questionnaires.
Statistical efficiency	Less efficient clustered samples needed for national surveys to reduce fieldwork costs.	Does not require clustered samples.	Does not require clustered samples.	Does not require clustered samples, but may have sampling problems (i.e. coverage).
Expected response rate	Usually gets highest rate.	Likely to be 10-30% lower than face to face.	Can be well below 50%.	Limited evidence, but likely to be a low response rate: may be higher among computer literate and young respondents.
Adapted from the Magenta Book: Guidance notes on policy evaluation http://www.nationalschool.gov.uk/policyhub/magenta_book/)				

Appendix B Causes of mode effects

Source: Jäckle et al, 2011



Appendix C Principles for Uni-Mode Question Design

1. Make all response options the same across modes and incorporate them into the stem of the survey question.
2. Avoid inadvertently changing the basic question structure across modes in ways that change the stimulus.
    - e.g. individual questions versus matrix, 'yes/no' versus 'tick all that apply'.
3. Reduce the number of categories to achieve mode similarity.
4. Use the same descriptive labels for response categories instead of depending upon people's vision to convey the nature of a scale concept.
5. If several items must be ranked, precede the ranking question with a rating question.
    - i.e. the rating task provides the means to get the respondent to focus on each individual item which will help with the subsequent ranking task
6. Develop equivalent instructions for skip patterns that are determined by answers to several widely separated items.
7. Avoid question structures that unfold.
    - These should be avoided because in self-completion questionnaires they lead to more questions being skipped.
8. Reverse the order in which categories are listed in half the questionnaires.
9. Evaluate interviewer instructions carefully for unintended response effects and consider their use for other modes.
    - e.g. routine interviewer answers to respondent queries about the survey question, such as "Please answer it in terms of what it means to you", cannot be used in the same way in self-administered modes and could affect how respondents answer the question in different modes.

Source: Dillman, 2000.

Appendix D Mixed mode designs for the National Survey

Mixed Mode Options for the National Survey for Wales	Population Coverage	Response rate	Measurement	Cost
Option 1 (section 5.1) Sequential design with drop-off questionnaire at contact stage: • Advance letter • Face-to-face contact – household questionnaire, respondent selection, drop-off paper self-completion with web option • Pick up of completed questionnaire or postal return • Face-to-face interviews for self-completion non-respondents	Complete coverage of Welsh household population.	Should achieve similar response rate to single mode face-to-face when using face-to-face interviews among those who do not complete self-completion questionnaire.	Where possible, uni-mode question design and pre-testing. Show cards and self-completion modules can be used in the face-to-face interview to minimise differences in measurement.	Scope for reducing costs is limited because interviewers are required at contact stage for respondent selection and to achieve interviews with the most reluctant respondents. Cost savings will depend on proportion of sample completing the self-completion forms.
Option 2 (section 5.2) Sequential design with postal questionnaire at first contact: • Postal questionnaire with instructions for respondent selection plus web option • Face-to-face interviews for self-completion non-respondents	Complete coverage of Welsh household population but risk of self-selection bias.	Should achieve similar response rate to single mode face-to-face when using face-to-face interviews among those who do not complete self-completion questionnaire.	Where possible, uni-mode question design and pre-testing. Show cards and self-completion modules can be used in the face-to-face interview to minimise differences in measurement.	Considerable cost savings could be achieved by not using face-to-face interviewers at the contact stage. Cost savings will depend on proportion of sample completing the self-completion forms.
Option 3 (section 5.3) Sequential design with telephone (RDD dual frame) at 1st contact: • Telephone contact – household questionnaire, respondent selection, request to complete web questionnaire or alternatively telephone interview • Telephone interview for those who those who agree to web but do fail to do so	Coverage depends on inclusion of Welsh mobile only households	Response rate for this design is likely to be a lot lower than sequential mixed mode designs that include face-to-face interviewing	Mix of web (visual stimuli, no interviewer) and telephone (aural stimuli, interviewer) carries a higher risk of mode effects than other combinations.	A lot cheaper than mixed mode designs that involve face-to-face interviewers. Cost savings will depend on proportion of sample completing the web questionnaire.
Option 4 (section 5.4) Single mode design with RDD (dual frame): • Telephone contact – household questionnaire, respondent selection, telephone interview	Coverage depends on inclusion of Welsh mobile only households	Response rate for this design is likely to be a lot lower than sequential mixed mode designs that include face-to-face interviewing	For any given year, no differences in measurement but comparisons with previous years and other relevant surveys could be affected.	More expensive than the above mixed mode designs but cheaper than current design (face-to-face survey).
Option 5 (section 5.5) Online panel based on probability sample: • Advance letter • Face-to-face contact – recruitment to panel • Web questionnaires	Complete coverage of Welsh household population	Refreshment samples would have to be added to mitigate the effects of attrition.	If only using web to collect data, then there is no risk of differences in measurement but comparisons with previous years and other relevant surveys could be affected.	Considerable upfront investment to provide computers and broadband access. Once it is set up, low unit costs and low marginal costs.