

Dadansoddi ar gyfer Polisi



Analysis for Policy



Llywodraeth Cymru
Welsh Government

SOCIAL RESEARCH NUMBER:

27/2022

PUBLICATION DATE:

07/04/2022

Review of options for the National Survey for Wales

Mae'r ddogfen yma hefyd ar gael yn Gymraeg.

This document is also available in Welsh.

OGL © Crown Copyright Digital ISBN 978-1-80364-004-4

Title: Review of options for the National Survey for Wales

Author(s): Peter Cornick, Jo d'Ardenne, Olga Maslovskaya,
Sierra Mesplie-Cowan, Gerry Nicolaas, Paul Smith

Full Research Report: Cornick, P; d'Ardenne, J; Maslovskaya, O; Mesplie-Cowan, S; Nicolaas G; Smith, P; (2022). *Review of options for the National Survey for Wales*. Cardiff: Welsh Government, GSR report number 27/2022.

Available at: [National Survey for Wales: development work | GOV.WALES](#)

Views expressed in this report are those of the researcher and not necessarily those of the Welsh Government

For further information please contact:

Chris McGowan

National Survey for Wales team

Welsh Government

Cathays Park

Cardiff

CF10 3NQ

Email: surveys@gov.wales

Table of contents

List of tables	3
List of figures	4
Executive Summary	5
1 Background	8
1.1 Report structure	9
2 Design parameters and quality criteria	10
2.1 Initial design parameters.....	10
2.2 Feedback from the stakeholder consultation	11
3 What should the sampling frame be?	12
3.1 Access panels	12
3.2 Random digit dialling (RDD)	12
3.3 Address-based sampling	13
4 Mode of contact and survey representation	16
4.1 First contact options using address-based sampling	16
4.2 Mail-only contact.....	16
4.2.1 Sampling considerations.....	17
4.2.2 How should within household selection work?.....	18
4.2.3 How can we maximise response in mail-only contact modes?	19
4.2.4 Survey response rates and sample composition	20
<i>Push-to-telephone</i>	20
<i>Push-to-web</i>	21
4.3 Knock-to-nudge	22
4.4 CAPI as secondary mode (sequential CAPI)	23
4.5 CAPI as primary mode.....	24
5 Data collection modes and measurement	26
5.1 Alternative modes of data collection considered.....	26
5.2 Issues associated with web modes of data collection.....	26
5.2.1 Barriers in converting CAPI questions to web questions	26
5.2.2 Can web questionnaires be 45 minutes?	28

5.2.4	Could a paper alternative be offered to the web survey?	31
5.3	CATI questionnaire issues	32
5.3.1	Barriers in converting CAPI questions to CATI questions	33
5.3.2	What can be done about the lack of visual presentation?	33
5.3.3	How could the self-completion aspect be managed?	33
5.4	Web-CATI questionnaire issues	35
5.5	New data collection modes	35
5.5.1	Computer Assisted Video Interviewing	35
5.5.2	Self-completion via tablets	37
5.5.3	CAPI plus	38
5.6	Data linkage and recontact consents	38
6	Data issues	40
6.1	Mode effects	40
6.2	Discontinuities	41
6.3	Small area estimation	42
6.4	Longitudinal designs and estimation of change	42
7	Broader priorities and implementation	44
7.1	Timely reports	45
7.2	Accessibility and inclusivity	45
7.3	Future proofing against Covid-19	46
7.4	CO ₂ reductions	47
7.5	Viability: novelty, complexity and risk	50
8	The strengths and weaknesses of different designs	52
9	Indicative costs of different approaches	59
9.1	Approximate costs for a web-CAPI design	59
9.1.1	Summary of design	60
9.2	Approximate costs for CATI-CAPI design	61
9.2.1	Summary of design	61
10	Reference list	63
Appendix A: Within-household respondent selection methods		67

List of tables

Table 7.4: Survey stage and potential ways to reduce CO ₂ emissions	48
--	----

List of figures

Figure 8.1: Trade-offs associated with mail-contact only designs with web or CATI data collection	53
Figure 8.2: Trade-offs associated with knock-to-nudge designs with web or CATI data collection	54
Figure 8.3: Trade-offs associated with sequential CAPI designs	56
Figure 8.4: Trade-offs associated with CAPI only and CAPI plus designs	57
Figure A1: A summary of within household selection methods	70
Figure A2: A summary of two-step approach to selecting respondents	71

Executive summary

Until March 2020, the National Survey for Wales was a face-to-face (CAPI) survey carried out in people's homes. However, the Covid-19 pandemic meant that alternative National Survey designs have now been trialled. The Welsh Government are considering what design should be deployed for the National Survey in the next contracting phase. This report summarises the different design options that could be used for the National Survey in the future. All the survey design options have their own strengths and weaknesses. This report summarises the various trade-offs the Welsh Government need to consider when deciding on which design to procure.

A set of survey design parameters was provided by the Welsh Government. These included criteria such as random probability sampling, minimum sample sizes, questionnaire length, reporting frequency, accessibility, being able to run in under lock-down conditions, the desire to reduce CO₂ emissions (compared to previous years of data collection), and the facilitation of recontact research. Requirements on practicability were also included, such whether contractors would be able to deliver the design, the complexity of analyses, and what can be achieved in the existing budget envelope.

In terms of sample frame, access panels and random digit dialling were deemed unsuitable for the National Survey as they would not provide results based on a random sample. Therefore address-based sampling remains the preferred sampling method. The Postcode Address File (PAF) and AddressBase Plus are currently the only practical options for this. AddressBase Plus may have some advantages over the PAF in that it is updated more regularly, and that it provides better characterisation of property type (for example whether an address is a business address).

Neither PAF nor AddressBase Plus include information on the name, telephone numbers or email addresses of people who could be living in the selected address. This means there are only two ways contact can be made with people at randomly selected addresses; mail-contact and/or face-to-face contact.

Using a mail-only contact strategy has some advantages. The cost of the survey is reduced, CO₂ emissions are reduced, and (in terms of the sampling design) there is no need for clustering in order to offset travel costs. A mail-only contact survey can be administered flexibly, which means that the sample can easily be distributed across the year, and the quarterly design can be nationally representative. Finally, a mail-only contact method could

allow for more frequent reporting. However, the major downside to mail-only contact is a decreased response rate, and thus an increased risk of not achieving a representative sample.

Having face-to-face contact will increase response rates and thus reduce the risk of non-response bias. If face-to-face contact methods are used there are various possible designs to consider. These are:

- Knock-to-nudge designs, where survey representatives visit addresses to encourage response but do not carry out in-home interviewing;
- Sequential CAPI, where face-to-face interviews are conducted with people who have not responded to an earlier mail invitation; and
- CAPI-first designs, where all selected households are visited by an interviewer, and all are offered a face-to-face interview as the default mode of data collection.

Knock-to-nudge designs offer some benefits over mail-only contact surveys, most notably the improved response rates and the increased likelihood of achieving a representative sample. However, it is unclear to what extent this method will remain in use post-pandemic now that in-home visits are permitted. There is a strong argument that once a survey representative has made contact on the doorstep the ideal option is to conduct a CAPI interview straight away, rather than risk drop-out by referring householders to a different mode of data collection.

Sequential CAPI modes offer most of the benefits of CAP- only modes, with the added advantage they could offer some reductions in both cost and CO₂ emissions. However, if multiple modes of contact and/or data collection are used, additional resource will be required from both the Welsh Government and contractors in terms of survey management. There will be more process documents to produce and sign off (procedures, mailings, interviewer instructions etc). Questionnaires will need to be checked and signed off for each mode of data collection. There will also be greater effort required in fieldwork monitoring, data processing and weighting. The more complicated the design the greater the resource required, from both Welsh Government and their contractors, and the higher the risk that errors could occur in practice. The mixed mode design will add some complexity to analysis.

Finally, the CAPI-only mode (as used in the National Survey pre-pandemic albeit with a paper self-completion module) was considered the 'gold-standard' until recently due to its high response rates. However, the cost of conducting CAPI fieldwork is high compared to

other modes of data collection. Retaining the CAPI only design would mean that reductions in interviewer mileage will not occur. If mileage reductions are not possible then CO₂ emissions will be more difficult to reduce. The CAPI only mode is not 'future-proofed' in the event of further lockdowns or if cost savings are required.

In terms of data collection, the alternative modes to face-to-face interviewing under consideration are web questionnaires or telephone interviews (CATI). Both modes would be suitable for National Survey data collection. Developing a mobile-friendly National Survey web questionnaire would be a more effortful process than developing a CATI questionnaire, and in web data collection randomised within household selection methods are problematic. That said a sequential web-CAPI design would offer marginal cost savings compared to sequential CATI-CAPI design, and web designs may offer some improvement in data quality for sensitive questions compared to CATI designs. The report concludes with some indicative costs for a web-CAPI sequential design and a CATI-CAPI sequential design, for the Welsh Government to consider.

1 Background

Until March 2020, the National Survey was a face-to-face (CAPI) survey carried out in people's homes. From May 2020, due to the Covid-19 pandemic, fieldwork was moved to telephone. Initially the telephone sample was drawn from previous National Survey participants. From January 2021, a new sample of respondents selected from the Postcode Address File was contacted by post and asked to provide a telephone number to take part. From July 2021, a trial of an online section (completed by respondents after the telephone section) was added.

The current contract to deliver the National Survey comes to an end with 2022-23 fieldwork, and a new contract will be awarded to deliver the survey from 2023-24 onwards. The change in mode that the pandemic has brought about is seen by the Welsh Government as an opportunity to review and revise the design and methodology of the National Survey in light of recent developments in the field, as well as an opportunity to reduce the carbon footprint of the survey.

The aim of the current project is to provide advice on the options for carrying out the National Survey from 2023-24 onwards. We did this by:

- Reviewing design parameters and quality criteria to be used in the review. We built on the initial criteria provided by the Welsh Government by using other quality standards, for instance the Total Survey Quality Framework (Biemer et al., 2014).
- Providing a qualitative questionnaire to a small group of National Survey stakeholders. The aim of this was to check what design features are most important and to gain practical information on how well new methods trialled in the pandemic have worked in practice.
- Reviewing alternative data collection approaches used elsewhere, noting where approaches have been implemented and any available information on data quality and practicality.
- Mapping the strengths and weaknesses of different approaches; and
- Providing indicative costs for two possible National Survey designs.

1.1 Report structure

The structure of this report is as follows:

- Chapter two summarises the design parameters for the National Survey;
- Chapter three summarises options for the sampling frame;
- Chapter four summarises options for the contact mode strategy for the survey, and what the impact of these decisions will be on response rates and representativeness;
- Chapter five looks at alternative modes of data collection including web, telephone, and new data collection modes;
- Chapter six looks at data issues, including mode effects, discontinuities in time series, small area estimation and whether a longitudinal design should be considered;
- Chapter seven looks at practical issues that go beyond data quality considerations;
- Chapter eight summarises the trade-offs of the different designs; and
- Chapter nine provides approximate costings for two possible alternative designs.

2 Design parameters and quality criteria

2.1 Initial design parameters

The following design parameters were provided by the Welsh Government as the key features for any future National Survey:

- The survey should use random probability sampling of households.
- A sample size of approximately 12,000 respondents a year is desired. The composition of the achieved sample should be as similar as possible to the wider adult population.
- The survey must enable the questionnaire content of around 45 minutes (on average) per respondent.
 - The questions must be able to change in response to policy needs.
 - The questionnaire should ideally include a self-completion element e.g. for questions on more sensitive topics.
 - The questionnaire should support data linkage and recontact research.
- Results should be published at least annually with fieldwork evenly spread across the year.
 - Ideally the design should minimise the time period between an information need being identified and results being delivered.
- The survey should provide estimates at different geographical levels for the 22 local authorities in Wales. A minimum effective sample size of at least 250 people a year is required in the smallest Local Authorities, to allow local area reporting.
- The survey should be accessible to all groups (i.e. it should support participation from subgroups including those who are not internet users, those who have hearing difficulties, those who have low levels of literacy, etc).
- It should ideally be capable of being delivered under lockdown conditions i.e. with minimal face-to-face contact between interviewer and respondent.
- The design should reduce the survey's carbon footprint i.e. by reducing interviewer travel compared with a face-to-face design.

2.2 Feedback from the stakeholder consultation

The above criteria did not meet with any major objections in the stakeholder consultation, although some stakeholders commented that the criterion for the survey design to work under lockdown conditions may be 'desirable' rather than essential. Stakeholders also described the following as being important:

- Making sure all documentation is available in both English and Welsh;
- Minimising the number of times changes in the design are made, in order to allow for less problematic time-series analysis;
- Having the ability to include ad hoc questions to address emerging policy needs;
- Timely reporting (some stakeholders wanted results available quarterly rather than annually).
- Having statistics at the Health Board level (this may imply boosting the sample in some local authorities)
- Considering whether it is possible to collect data from people not living in private households, e.g. in care homes.

3 What should the sampling frame be?

The following section reviews the sampling frame options for the National Survey.

3.1 Access panels

Access panels recruit a pool of respondents who agree to participate in surveys regularly (usually online so that there can be rapid response), generally in exchange for a financial incentive. Despite the potential for rapid response from the panel, we have reservations about the use of access panels for the National Survey. Most panels are recruited opportunistically, which results in a panel which is not representative of the population, particularly excluding types of people who are harder to reach and less likely to engage with public activities. This can be offset partly by a probabilistic recruitment process. The probability selection approach means that there is a firmer basis to compensate for differences between the panel composition and the population composition by weighting, but there is still a risk of selective entry to the panel which could lead to biased results (i.e. results that differ from a truly random sample of the wider population).

There are constraints on how often members of access panels can be contacted, so as not to overburden them. These constraints do not fit well with National Survey requirements for minimum sample sizes in local authorities within Wales, which would quickly exhaust the available sample within existing access panels. One possibility might be to recruit a probability-based access panel specifically for the National Survey, but this is close to the development of a longitudinal survey (see section 6.4), with the associated challenges of keeping the sample up-to-date so that it remains representative and tracking the panel members, which would require additional resources. Therefore, we do not consider the use of access panels to be a viable option for the National Survey, and do not consider this method further.

3.2 Random digit dialling (RDD)

Random digit dialling (RDD) provides an approach to recruiting a probability sample that is distinct from the more usual address-based approaches. The common usage of both landlines and mobile phones means that two separate procedures would be needed to ensure reasonable population coverage. Landline and mobile information is available from Ofcom databases. Landlines are localised to an area, so that in principle targeted sampling for certain local authorities is more easily achieved. In

contrast, mobile phone information is not localised. In this case RDD would involve screening to identify respondents who live in Wales, and further screening for respondents living in specific areas to meet minimum sample sizes within local authorities. The screening is likely to be expensive and wasteful of contacts, in that only a small percentage will be both in Wales and in target areas.

The need for both landline and mobile RDD means that the survey would use a dual frame sample, since people have the possibility to be sampled in both parts of the design. This creates extra complexity, since the selection probabilities need to take account of both these selection routes. This requires additional questions in the questionnaire. In principle, most of the differences can be taken account of in the weighting, but the weighting procedure is considerably more complicated than designs that have only one selection route.

Recruitment of respondents through mobile phones in particular gives a *person* response, whereas some of the information collected in the National Survey corresponds to *households*. This means that households also have multiple chances of selection through different individuals. This is indirect sampling and can also be handled by weighting through the generalised weight share method (GWSM, Lavallée 2007), though it too creates additional complexity.

The conjunction of dual frame sampling and indirect sampling makes the whole methodology associated with RDD challenging, and we rule out RDD as a suitable approach on both costs (from screening) and complexity (of weighting and analysis) grounds.

3.3 Address-based sampling

Address-based sampling is a more traditional approach to probability-based recruitment for general population surveys. It offers a direct way to target sampling in specific geographic locations, since the addresses are already available, and therefore supports differential sampling. Since there are no comprehensive links between addresses and modes of contact such as telephone or email (see also chapter four), the contact must be made either by post or through an interviewer visit. Interviewer visits are relatively expensive, but we already know that the address is in the correct geographic area, so no additional screening is needed (although there is screening to identify eligible addresses). If there were an address frame with additional

characteristics, this might present further design options, but currently the only practical options are the Postcode Address File (PAF) and AddressBase Plus, which have only limited, address-specific (rather than person-specific) information.

Both PAF and AddressBase are thought to have almost complete coverage of the household population (Wilson & Elliot, 1987), though this seems not to have been formally evaluated recently (Tipping & Nicolaas, 2006). Parts of the population which do not have telephone or internet access, which may be difficult to reach through other frames, are well covered by address-based frames. The probability sampling of addresses is straightforward, involving a single- or multi-stage design, but not needing additional processing to deal with multiple frames or indirect sampling. It is possible to meet an achieved sample of 12k responses with minimum numbers of responses per LA, but the size of the issued sample needed to achieve this will depend on the response rate, which is influenced by the survey design, mode and salience of the topics.

The ONS Social Statistics Transformation Programme is considering the future replacement of PAF with AddressBase as the primary sampling frame for ONS surveys of the residential population.¹ AddressBase products include PAF but also other address-related information sources such as Local Land and Property Gazetteers (LLPGs), the Valuation Office Agency (VOA), and the Ordnance Survey. AddressBase should improve the coverage of surveys by reducing over-coverage (inclusion of addresses which are non-residential) – an eligibility rate of 95% has been cited for AddressBase products compared to 91% for PAF². It should also be more accurate than PAF with updates being released every six weeks.³ Additional benefits include a multi-occupancy indicator (potentially useful for both sampling and weighting) and a Unique Property Reference Number (UPRN). Having the UPRN on the frame will simplify the process of matching survey data to Land Registry data for the production of housing estimates. Some evidence on the quality of assignment of UPRNs is available in ONS (2019, Table 2).

AddressBase products are already being used as a sampling frame for a number of ONS surveys including the Coronavirus Infection Study, the Labour Market Survey

¹ See recommendation 3 at [Progress against the Labour Force Survey National Statistics Quality Review recommendations](#)

² [Integrated Population and Characteristics Survey \(IPACS\)](#)

³ See Table 2 at [ONS working paper series no 17 - Using data science for the address matching service](#)

(web-first prototype to replace the Labour Force Survey), and the Adult Oral Health Survey. It was also used for the 2021 Census.

The AddressBase products are available to public sector organisations under the Public Sector Geospatial Agreement (PSGA). The PSGA is a contract between the Geospatial Commission, on behalf of the public sector in England, Wales, Scotland, and Ordnance Survey and it provides the route for public sector members to access, use and share data from AddressBase products. The Welsh Government is a member of PSGA and can allow its contractors to use AddressBase data under this agreement. A contractor licence may be required. Such a licence is not needed if the data are being provided to anyone outside the Welsh Government who is working with its employees temporarily. They would be covered by the terms of Welsh Government's Public Sector Geospatial Agreement (PSGA) Member Licence.

The AddressBase products should provide a better characterisation of property types than PAF which contains only a division into small users and large users.

AddressBase includes property types including (for example) business addresses and communal establishments. This information could potentially be used to improve the weighting, but only if property type is sufficiently correlated with the outcomes of interest.

4 Mode of contact and survey representation

4.1 First contact options using address-based sampling

The lack of names, telephone numbers and email addresses on the sampling frame (either PAF or AddressBase), leaves us with two viable first-contact modes: mail and face-to-face. Although it is possible to match names and telephone numbers to sampled addresses using the edited Electoral Register and listed telephone numbers (e.g. BT OSIS database), the matching rate will be very low resulting in non-coverage bias unless combined with mail and/or face-to-face contact modes.⁴ We have therefore not included first-contact by telephone as an option in this review.

Based on using mail and/or face-to-face modes for first contact, we have identified four broad categories of survey design options for consideration:

1. Mail contact only
2. Knock-to-nudge
3. CAPI as secondary mode
4. CAPI as primary mode

We will now discuss each of these in turn.

4.2 Mail-only contact

Using mail contact only, eligible residents at sampled addresses are invited to take part in a self-completion survey and/or a telephone survey. This group of survey designs does not include any face-to-face contact or data collection.

Mail surveys involve sending paper questionnaires by post to sampled addresses. Such questionnaires need to be short, simple and should have limited question routing in order to encourage completion and reduce the risk of respondent error. Given the amount and type of data required for the National Survey, a paper questionnaire is not feasible as a stand-alone instrument and has therefore not been included as an option in this review.

Push-to-web surveys, also known as web-first and ABOS (address-based online surveying), involve sending letters to sampled addresses and inviting eligible

⁴ ONS carried out telematching for the Family Resources Survey which used a push-to-telephone approach during the pandemic. Telephone numbers were matched to sampled addresses using the edited Electoral Register (names) and BT (listed telephone numbers). The matching rate was about 10%.

household members to go online and complete a questionnaire. As the questionnaire is online, it is possible to use automated routing and other computerised features to simplify the task for the respondent, and to reduce respondent error. Respondents are therefore not exposed to the whole (potentially lengthy) questionnaire, just those questions that they are required to answer. Many push-to-web surveys will also include an option to complete a paper questionnaire, in order to ensure coverage of those without internet access as well as boosting response among those who are either unwilling or otherwise unable to complete an online questionnaire. A paper questionnaire is either sent on request or sent to all or a sub-sample of web non-respondents. Often these paper questionnaires are shortened and simplified versions of the online questionnaire, thus leading to missing data for this sub-set of respondents.

Push-to-telephone employs the same contact approach as push-to-web but the letter invites eligible household members to take part in a telephone survey. This requires the willing participant to opt-in to the survey by mail, telephone and/or online portal. Alternative modes are not needed to boost coverage because telephone ownership (landline and/or mobile) is almost universal.

4.2.1 Sampling considerations

Using mail contact only has advantages for the sampling design, because there is no need for an interviewer to travel to the location, and therefore no need for clustering in order to control travel costs. Simple random sampling of addresses can therefore be used, with appropriate stratification to deal with (a) requirements for minimum sample sizes in local authorities and (b) differential allocation by features associated with characteristics of the responses. The gains from stratification based on features are typically modest for household surveys, and are likely to be offset by differential sampling for minimum sample sizes, so the design effect⁵ may still be >1, but will be substantially smaller than for a clustered sample.

⁵ The difference in the precision of the estimates produced by a complex design relative to a simple random sample is known as the design effect (deff). The design effect is the ratio of the actual variance, under the sampling method used, to the variance calculated under the assumption of simple random sampling. This number will vary for different variables in the survey – some may be heavily influenced by design effects and others less so. (Sturgis, 2004)

A mail-only survey can also be administered flexibly, which means that the sample can easily be distributed across the year, and the quarterly design can be nationally representative.

There is no information in the available sampling frames about number of dwellings/households at each identified address but from previous surveys we know the proportion of addresses containing more than one dwelling/household is less than 1%. The risk of bias from subsampling within multi-occupancy addresses is therefore negligible and can be ignored.

There is equally no information on the frame about the number of eligible household members at each address, which means that a process for selecting eligible adults to respond is needed. Here there is a risk of self-selection bias, derived either from the relative propensity of the first contacted adult to respond, or from the incentive payment. This self-selection bias cannot be ignored, as it is likely to be substantial and correlated with survey estimates, and thus becomes a risk to be managed and adjusted for which can be problematic when relying only on postal contact (see next section).

4.2.2 How should within household selection work?

For push-to-telephone surveys, the telephone interviewer can enumerate the household and randomly select one adult to take part in the survey (e.g. using the Kish method, next / previous birthday or some other random or quasi-random approach). In some cases the selected respondent may not be the person who the telephone interviewer is speaking to, which will require careful handling by the interviewer. Furthermore, there is a risk that a willing household member rather than the randomly selected person takes part in the telephone survey.⁶

Within-household selection is less straightforward for push-to-web surveys. There is evidence showing relatively high levels of non-compliance with written instructions on how to select household members for survey participation, resulting in about 20% - 25% of respondents not being the target respondent which will introduce self-selection bias in the survey estimates. The alternative approach of allowing anyone in the household to take part will also introduce self-selection bias. The risk of self-

⁶ In a US telephone survey carried out by Lavrakas and Bauman (1993), the proportion of incorrect respondents when using the Last Birthday Method was estimated to be about 20%-25%. To the best of our knowledge more recent UK evidence is not available.

selection bias can be reduced by inviting all household members to take part in the survey but this will decrease the effective sample size due to clustering and, when coupled with a conditional incentive, could encourage the fabrication of extra household members and/or proxy responses. Subsequently, alternative methods have emerged which strive for an optimal balance between different sources of error and practicality. In the UK, the most commonly used methods are “up to four adults” and “up to two adults” selection with no random selection, i.e. any adults in the household can choose to take part up to the specified maximum:⁷ e.g. Community Life Survey (up to four), the Active Lives Survey (up to two). Some surveys have opted for “up to three” adults (e.g. Bike Life Survey, Financial Lives Survey). Typically, these tend to be any two/three/four adults rather than a random selection of adults within the household based on written instructions.

A more detailed description of within household selection methods is provided in Appendix A.

4.2.3 How can we maximise response in mail-only contact modes?

There is no silver bullet for maximising response rates when using mail-only contact. Instead attention is required for each step in the process of gaining co-operation from eligible residents at sampled addresses.

The first step is to increase the likelihood of the envelope being opened. As the sampling frame is a list of addresses, letters addressed to the “householder” or the “resident” could be mistaken for junk mail and go straight into the bin. The use of **authoritative logos** on the envelopes could help to overcome this misconception and reduce the risk that letters are ignored. There is some evidence showing that real stamps will produce a higher response rate compared to ‘postage paid’ franking but this is considerably more expensive, anything from about 25% for 1,000 first class letters. Other envelope design features such as taglines (e.g. “Urgent”, “Claim your £10”) and envelope colour have not been found to be very effective in increasing response rates.⁸

⁷ Self-selection bias is very small and considered to be less problematic than nonresponse bias when using more complicated random selection instructions.

⁸ For a recent UK example of testing different envelope colours and country-specific branding on envelopes, see Ipsos MORI (2018).

Having opened the envelope, the recipient needs to engage with the content of the letter. Letters should be **short and easy** to read, use a simple font style, include persuasive reasons for taking part in the survey, and have a clear call to action.⁹

The response rate will increase with the **number of mailings** sent to non-respondents. On push-to-web surveys, the first reminder tends to almost double the response rate achieved after the first invitation mail out. Second and third reminders still have a positive but diminishing impact on the online response rate.¹⁰

There is ample evidence showing that **incentives** will increase survey response rates on surveys that use mail-only contact. Unconditional incentives have a bigger impact on response rates than conditional incentives, but they are less likely to be cost effective given the low baseline response rate for push-to-web and push-to-telephone surveys. Typically surveys using mail-only contact tend to use conditional incentives of £5 or £10 per completed questionnaire.

Push-to-web surveys often provide **paper questionnaires**, either on request or sent with one of the reminder letters. Offering paper questionnaires as an alternative to web increases response rates and improves representativeness by drawing into the sample those without internet access and encouraging participation among those who are unable or reluctant to complete the questionnaire online. For push-to-telephone surveys, there is no recently published evidence on the impact of offering a paper or web alternative to telephone non-respondents.

4.2.4 Survey response rates and sample composition

Push-to-telephone

During the pandemic, we observed a number of high quality CAPI surveys switching to a push-to-telephone design; e.g. the Family Resources Survey (FRS), the National Travel Survey (NTS), the English Housing Survey (EHS). Response rates for this design varied greatly month by month from as low as 10% to about 20%. It is unclear what has caused these large monthly variations but this could be linked to improvements being made to the design over time, as well as pandemic-related effects (e.g. people being at home during lockdowns). Differences were also observed across surveys which are likely to be linked to survey topic, survey

⁹ See Dillman (2014) for guidelines on designing letters.

¹⁰ Analysis of return rates on NatCen push-to-web surveys.

sponsor, target population, and response maximising efforts such as the type / value of incentives and number of reminders.

These response rates are considerably lower than what was achieved pre-pandemic when using CAPI. For example, the response rate for the NTS was 54% in 2019 using CAPI and about 16% in 2020 using push-to-telephone. Such differences in response rates will increase the risk of nonresponse bias in survey estimates.

For the NTS, descriptive analysis showed differences in sample composition when using a push-to-telephone design compared to the pre-pandemic CAPI approach. At the household level, differences were found for the Index of Multiple Deprivation (IMS) and housing tenure; i.e. those households living in the most deprived areas and in rented accommodation were less likely to take part when using the push-to-telephone approach compared to the pre-pandemic CAPI approach. Differences were also observed at the individual level, with the achieved sample under the push-to-telephone approach being older, more educated, and less ethnically diverse than with the pre-pandemic CAPI approach (Cornick et al., 2021). Similar differences in sample composition have been reported for other surveys that switched from CAPI to push-to-telephone, including the National Survey and other surveys carried out by ONS.¹¹

Push-to-web

Household response rates for push-to-web surveys tend to vary greatly, ranging from as little as 5% for the Financial Lives Survey 2020 (online only) to over 30% for the British Election Study 2019 (online plus paper with a very generous incentive package including an unconditional £5 sent with invitation letter and a £25 voucher conditional on completion which was increased to £50 at the final reminder stage). It would appear that these differences in response rates are related to survey topic, survey sponsor, and survey design features such as the number of reminders, the type and value of incentives, and the use of an alternative mode.

Descriptive analysis carried out on the Community Life Survey (TNS BMRB, 2013) showed that online respondents to push-to-web surveys were more likely than CAPI

¹¹ For an overview, see [Impact of COVID-19 on ONS social survey data collection](#). Last revised 14 February 2022.

respondents to be middle-aged, degree educated, high earners, native English speakers, living as a couple, home-owning, and at least partly responsible for the care of another person. Adding a postal questionnaire improved the sample composition in some respects but not as much as expected, and it made the age profile worse.

More recently, a similar sample profile has been reported for the push-to-web component of the British Election Study 2019; i.e. compared to CAPI respondents, online respondents were younger, were more likely to hold a degree, have a mortgage, no religion, and working full-time (Bogdan, 2021). However, in contrast to Williams (2013), Bogdan concluded that the use of a paper questionnaire did improve the sample profile by drawing into the sample those who were older, retired, and did not have a degree.

4.3 Knock-to-nudge

During the pandemic when lockdowns were eased but social distancing rules were maintained, a new contact method called “knock-to-nudge” emerged. This method involves interviewers visiting sampled addresses, speaking to a household member on the doorstep, and requesting participation in a web or telephone survey. The prevailing social distancing requirement meant that no face-to-face interviewing was carried out in the home. Doorstep interviewing was not feasible for most social surveys due to questionnaire length and the need for privacy. Typically, push-to-web or push-to-telephone methods are used before issuing non-responding addresses to “knock-to-nudge”. However, NatCen has also used “knock-to-nudge” as the first contact mode for the National Travel Survey (NTS) as a means of boosting the response rate.

The sampling considerations for this method are the same as for mail-only contact methods with the exception that the sample often has to be clustered in order to constrain travel costs for the knock-to-nudge element. Clustering will reduce the effective sample size, which has a negative impact on the level of precision.

If sampled households are being contacted by mail before knock-to-nudge is attempted, then within-household selection remains problematic. However, when interviewers make contact on the doorstep it is possible for them to use the next birthday method or alternatively to switch to the Kish method of random selection to enumerate the household and randomly select one or more respondents.

An increase in response rates was observed across all push-to-web and push-to-telephone surveys that introduced a knock-to-nudge intervention among those who did not respond to the mailings. ONS (2022) has reported increases in push-to-telephone response rates ranging from 2.5 percentage points for the Living Costs and Food Survey to 10.7 percentage points for the Labour Force Survey.¹² The BARB survey¹³ doubled its push-to-web response rate of 17% to 36% by using knock-to-nudge among those households who had not responded to the invitation and reminder letters.

The knock-to-nudge approach has an even greater impact on response rates when used as the first contact mode instead of mail. For example, the push-to-telephone response rate for the NTS increased from 16% when using mail-only contact to 48% when using knock-to-nudge at first contact. This compares to the pre-pandemic CAPI response rate of 54% in 2019.

There is some evidence showing that the introduction of knock-to-nudge also improves the sample composition. For example, ONS has reported that its knock-to-nudge surveys captured a larger proportion of younger respondents who were not married, lived in larger households of three or more (possibly with their families), were renting or sharing their accommodation, and were living in the most deprived areas. This impact was observed despite the knock-to-nudge approach only being used among a certain proportion of those who did not respond to the mailings. These results demonstrate the key role that interviewers make in drawing into the sample those who are more reluctant to take part in surveys.¹⁴

4.4 CAPI as secondary mode (sequential CAPI)

As part of its “digital by default” strategy, ONS intends to offer an online self-completion mode for all of its social surveys. The model that they have adopted for the Labour Market Survey (LMS)¹⁵ is a **push-to-web approach followed by CAPI** among those who do not respond online. This method was used for the 2021 Census for England and Wales, and it is envisaged that this method will also be used for the future Integrated Population and Characteristics Survey (IPACS).

¹² For an overview, see [Impact of COVID-19 on ONS social survey data collection](#). Last revised 14 February 2022.

¹³ The Broadcasters Audience Research Board – a survey designed to measure the UK’s audience ratings.

¹⁴ For an overview, see [Impact of COVID-19 on ONS social survey data collection](#). Last revised 14 February 2022.

¹⁵ A prototype survey designed to replace the UK Labour Force Survey.

The development of the LMS has involved considerable testing and the survey results have been compared against the Labour Force Survey (LFS) datasets covering the same period. The findings to date have been very encouraging. The push-to-web design followed by CAPI among non-respondents produced an overall response rate of 56.5% which is slightly higher than the wave 1 response rates for the LFS in the final quarter of 2018 and the first quarter of 2019 (53.3% and 54.6% respectively). The online (28.4%) and face-to-face (28.1%) response rates for the LMS were very similar (ONS, 2020a).

There were minimal differences in sample profiles between the LMS test and the LFS in terms of the socio-demographic characteristics of the responding sample; the proportions of responding males and females were the same and only some of the age distributions varied slightly. The responding sample for the LMS test was more evenly distributed in terms of the Index of Multiple Deprivation (IMD) deciles than the LFS over the same period. Furthermore, no statistically significant differences were found between the LMS test and the comparative LFS dataset for any of the important headline labour market estimates. Significant differences were however observed in the estimates of people aged 16 to 24 years in full-time education, people who were classified as self-employed and total hours were worked by the total workforce. These differences are being investigated further. (ONS, 2020b; ONS, 2020c)

To the best of our knowledge, there are no examples of a **push-to-telephone approach followed by CAPI** among those who do not respond online. Given that the methods for making contact and gaining cooperation are the same, our assumption is that the overall response rate and sample composition is likely to be similar to that achieved on a push-to-web survey followed by CAPI and therefore similar to equivalent surveys that use face-to-face contact from the outset. However, this would have to be verified through testing.

4.5 CAPI as primary mode

Face-to-face contact and computer-assisted personal interviewing has been the preferred design for high quality surveys in the UK for many decades, including the National Survey. Generally, surveys that require face-to-face contact and CAPI data collection have the best sample coverage properties and the highest response rates (and, therefore, lower risk of nonresponse bias).

However, maintaining the current National Survey design would not meet one of the Welsh Government's criteria: a reduction in the survey's carbon footprint; i.e. by reducing interviewer travel compared with its current face-to-face design. With data collection costs increasing, it may also prove difficult to continue this design within the existing budget.

Furthermore, the National Survey could be designed to be deliverable when external factors impose limitations on face-to-face contact and face-to-face data collection (e.g. lockdowns, budget cuts). Having a questionnaire that is portable across modes would enhance the ability of the survey to be administered using different modes of contact and data collection.

5 Data collection modes and measurement

5.1 Alternative modes of data collection considered

In this section we will consider options for alternative modes of data collection to CAPI. We will discuss measurement issues associated with four main alternatives:

- Web only modes,
- Web first modes with a paper alternative,
- CATI only modes,
- Web/CATI mixed-mode designs.

We will also discuss innovative methods which have been trialed and tested during Covid-19 pandemic such as Computer-Assisted Video Interviewing (CAVI) and the provision of self-completion tablets to respondents. Finally, we will discuss the ‘CAPI plus’ approach to data collection, i.e. offering an alternative mode as part of CAPI-first data collection.

5.2 Issues associated with web modes of data collection

This section discusses measurement issues specific to web mode of data collection. We will first discuss the barriers in converting CAPI questionnaires to web questionnaires. We will then discuss issues related to the questionnaire length. Then we will address matrix sampling approach as a method of reduction of questionnaire length. We will finish this section by discussing whether a paper alternative can be offered to those who cannot complete questionnaire online and what are the advantages and limitations of this approach.

5.2.1 Barriers in converting CAPI questions to web questions

During the Covid-19 pandemic when CAPI was not permitted, despite expectations from some survey methodologists that many surveys would transition to online data collection, we observed a greater shift to CATI (or pausing data collection) rather than to moving to web modes. One barrier to moving surveys online was the lack of time for optimising questions for web self-completion e.g. for creating the necessary adaptations in terms of question length, visual design and so on. Likewise, there were concerns about the accessibility of web questionnaires for some groups, for example those with limited digital literacy. For example, in 2018, 66% of people in Wales had five/five ‘basic digital skills’ (Office for National statistics, 2019b).

When CAPI is used, not as much effort is needed to make the questionnaires respondent-friendly in terms of their interface and visual design. This is because an interviewer is always present to administer the questions. Interviewers receive appropriate training on how to navigate the questionnaire software and are experienced questionnaire software users (i.e. interviewers know how to access help screens, input data into different types of answer field etc). Even in self-completion sections (i.e. the CASI component) interviewers are on hand to assist the respondents with comprehension and data-entry if this is needed. Therefore, user-friendly design and interface are of lower priority for CAPI questionnaires.

However, for transitioning the questionnaire online, this aspect becomes of a high importance as an interviewer is not there any longer to guide the respondents through the process of questionnaire completion. Unfortunately, a “lift and shift” approach towards web questionnaire design does not work effectively (personal communications with colleagues from ONS). Therefore, the process of preparation of the instrument for online self-completion requires effort, resources, and time. The whole questionnaire might need to be re-programmed in different software allowing for user-friendly visual designs for specific questions, help screens, and other useful and necessary features.

All these concerns led some survey data collection organisations to use telephone mode of data collection during the pandemic because translating the CAPI questionnaire into a CATI questionnaire could be done with less investment of effort and time than transitioning to web questionnaires.

The barriers to transitioning to online mode of data collection can be addressed by ensuring that enough time and resources are available for the conversion process. One key consideration is to ensure that all questions render well on smaller screens i.e. mobile devices. Having a ‘mobile-first’ approach to questionnaire design is recommended. This means being mindful of how much text can appear on a mobile screen and writing questions that require minimal scrolling. This approach then allows easy use of questionnaire on other devices with larger screens, and it is considered best practice and helps minimise measurement difference between devices respondents might choose for survey completion.

In order to address accessibility issues, simple language should be used (for example, ONS recommends the rule of writing any questions for the average

reading age of the UK public¹⁶). Simple presentation of fonts, colours, and layouts should also be implemented. These measures are also crucial for the minimisation of measurement error and item nonresponse. For the National Survey, it is important to allocate additional time for development of the instrument in both English and Welsh languages to make sure the presentation of questions is mobile optimised in both languages.

It is also important to consider trade-offs between optimising and re-designing questions to improve accessibility while also preserving time-series of questions when needed.

There is a misconception about online surveys being cheaper in comparison to other modes of data collection. Although cost-savings can be made the savings are not huge for probability-based online surveys, especially when additional efforts are made to convert reluctant respondents and to get the highest response rate possible (for example by using repeated mailings, high incentives and knock-to-nudge methods/ CAPI follow-up: see section 4). Online surveys might be only slightly cheaper when everything is prepared and set up, including completion of re-designing and re-programming of the questionnaire, when compared to interview-administered modes or paper-based approaches. However, some of the re-design and programming costs associated with the move online would be a ‘one-off’ given that much of the content of the National Survey is repeated each wave.

5.2.2 Can web questionnaires be 45 minutes?

Existing literature and best practice guides suggest that the ideal web questionnaire length should be around 10-20 minutes (Galesic & Bosnjak, 2009; Revilla & Ochoa, 2017; Lugtig 2021). Longer online questionnaires have been associated with lower data quality, for example, higher nonresponse rates, higher break-off rates, higher frequency of “don’t know” answers. A summary of existing literature conducted by Lugtig (2021) suggests that when surveys are longer than the recommended 10-20 minutes, small effects on nonresponse and breakoffs are observed as well as some effect on measurement error. However, literature suggests that the incentive value and subject salience in both cross-sectional and longitudinal surveys (as well as participants’ loyalty to the surveys in longitudinal contexts) have more influence on

¹⁶ [Designing the future of ONS surveys](#)

response rates and other data quality indicators than questionnaire length (Mavletova & Couper, 2016; Lugtig, 2021).

During the Covid-19 pandemic in 2021 the European Social Survey (ESS) successfully trialled longer web surveys in a cross-sectional context. An experiment was conducted in Austria and response rates were compared by questionnaire length and different level of incentives (Hanson, 2021). They introduced five experimental groups:

- A. 20 minutes with €5 of unconditional incentive and €10 of conditional incentive;
- B. 35 minutes with the same incentive structure as group A;
- C. 35 minutes with €5 of unconditional incentive and €25 of conditional incentive;
- D. 50 minutes with group A incentive structure;
- E. 50 minutes with group C incentive structure.

No significant difference in response rates was identified between 20 and 35 minute questionnaires, and only marginal difference of around three percentage points was observed in response rates between 20 and 50 minute questionnaires. The results of a more detailed investigation of data quality indicators for longer questionnaires are very positive¹⁷. There was no significant difference observed in the average number of missing responses between the shorter and longer questionnaire conditions. There was also no significant difference observed in the percentage of respondents selecting the same response for all items (nondifferentiation), and no significant difference observed in the respondent standard deviation between the questionnaire length conditions. When examining a battery of questions where agreement would be a contradiction, the association was observed to be negative, as expected, and this relationship was consistent across conditions.

These findings are very reassuring and suggest that with the right incentive structures longer questionnaires can be used successfully and, therefore, longer web questionnaires (20 minutes+) should not be ruled out.

¹⁷ Personal communication with Tim Hanson (November 2021).

5.2.3 Should matrix sampling be used to shorten average length?

Matrix sampling refers to an approach where different respondents get a sample of questions from a longer set. This approach has several objectives. The principal one is to keep the questionnaire as short as possible for all respondents while gathering information on many topics (which can also be thought of as increasing the number of topics which can be asked within a fixed interview/response time). But it can also be used to adjust the sample size for different questions within the same survey, so that a topic that is relatively well-measured can have a reduced sample size and free up space for other topics or for increased sample sizes for topics which are less well measured.

Matrix sampling covers a wide range of complexity. At its simplest it may involve splitting the sample into two sub-samples which get different questions, and this kind of approach has been used in UK household surveys (e.g. the Wealth and Assets Survey). It involves some extra complexity in processing, as different weights are needed for the questions which are subsampled. There are methodological approaches which use the information from the fully sampled questions to reduce the variance of estimates from the subsampled variables, and if the correlation between subsampled and main variables is strong, this effect may be substantial (see for example Renssen & Nieuwenbroek 1992, Merkouris 2010, 2015).

In principle this approach can extend to complex schemes covering many question groups, but the challenges in producing a functional questionnaire with a logical flow for the respondents, and of dealing with the reweighting of many different patterns, can become substantial. This approach can however be an efficient design choice as long as the complexity can be handled (Chipperfield & Steel, 2011). Thus there are few examples (see Chipperfield & Steel 2013 for one), though it has been suggested as the basis for a system of integrated social surveys which is not unlike the merged National Survey (Ioannidis et al., 2017). The National Survey has used subsampling since 2012-13, accounting for differences in weights from 2016-17.

In the case of the National Survey, the additional complexities of administering a split questionnaire design, and the consequent additional processing of weights, which also has an effect on the ability of the users of the survey microdata to undertake analyses correctly accounting for the design, all suggest that this

approach is not ideal. A more detailed evaluation of this approach might be worth considering in the future if resources for the National Survey are particularly constrained.

5.2.4 Could a paper alternative be offered to the web survey?

Some surveys use a mixed-mode design combining web and paper modes. In this design online data collection is offered first, but offline respondents can request the use of a paper version of the questionnaire (for example, Active Lives Survey, Community Life Survey, British Election Study (BES)). This approach was also successfully used for Census 2021 in England and Wales (the respondents were able to request a paper version of the questionnaire for Census completion).

There may be practical issues to offering paper alternatives for the whole National Survey questionnaire. The first barrier is the questionnaire relies on electronic routing (where some questions are only asked of some groups dependant on their earlier responses). This type of routing can be difficult to visually render on paper and would require an element of redesign work. Due to the length of the interview the resultant paper questionnaire would be a long document, which may be initially off-putting to respondents.

It is also important to remember that mixed-mode designs can introduce mode effects, and in some contexts which are more prone to mode effects, it is important to make necessary adjustments for them during data analysis stage. In the context of mixing web and paper designs, the risk of mode effects is not high as both modes are self-administered and visual, and can be designed to be as equivalent as possible (Nicolaas, 2012). Conversely, the risk of mode effects is higher when mixing web and CATI. Some of the approaches discussed in literature on the adjustment for mode effects are (Klausch et al., 2017a, b; Kolenikov & Kennedy, 2014; Park et al., 2016; Schouten et al., 2022; Vannieuwenhuyze et al., 2014):

- The regression modelling adjustment approach: adjustments are computed by regressing survey responses on mode, demographics, and other relevant variables. This approach assumes that one of the modes is the gold standard and the other mode(s) is measured against it. The approach works best when one of the modes in the mixed-mode design is face-to-face, therefore might be less suitable for other mixed-mode designs;

- The multiple imputation adjustment approach (conceptualised as missing-data problem): the method is applied to create imputed values of the unobserved counterfactual outcome variables in the mixed-mode surveys;
- The re-interview method, when respondents are re-approached under a second mode, where relevant questions from the survey are repeated.

All of the above adjustment methods require a high level of technical expertise to implement and/or have various limitations. Due to complications linked to the adjustment for mode effects, it is advisable to invest in design of the instruments and focus on prevention and minimisation of mode effects rather than on adjustments where possible.

It is also important to remember that in mixed-mode designs the sample profiles for different modes often differ. For example, it was found in BES that those who completed the survey online were younger, had higher educational qualifications, were mortgage owners and were more likely to be working full time, whereas those who completed the survey using a paper version of the questionnaire were older, more likely to be retirees, were home owners, had lower educational qualifications among other characteristics (Bogdan, 2021).

A paper alternative can be offered to supplement web data collection and to include offline respondents, but it is important to remember the issues associated with mixed-mode designs and to address them appropriately at the data analysis stage if required and appropriate.

5.3 CATI questionnaire issues

As mentioned above, during the Covid-19 pandemic many surveys transitioned from CAPI to CATI. CATI was an attractive alternative as it still had an interviewer who was able to guide respondents through the process of survey completion and, therefore, not many changes had to be introduced to the questionnaires that had been designed for CAPI. However, despite this option being the preferred design for numerous surveys, there are still challenges in converting a CAPI questionnaire to a CATI one. These barriers will be discussed below. We will then discuss the option of introducing a self-completion element to the CATI questionnaire and the issues associated with this.

5.3.1 Barriers in converting CAPI questions to CATI questions

Three main barriers in converting CAPI questions to CATI questions were identified:

1. No visual presentation of questions is available in CATI mode such as showcards.
2. There is an increased risk of measurement error for questions that rely on longer lists or other visual prompts. CATI suffers from recency effects when a respondent tends to select an option which they heard most recently. It might be difficult to retain all the options when the lists of possible categories are long.
3. Telephone interviews do not support self-completion elements. If self-completion elements are offered these are typically done using another mode, which introduces the risk of drop-off and item nonresponse.

5.3.2 What can be done about the lack of visual presentation?

In the National Survey, the sports participation questions are at higher risk of having increased measurement error due to long lists of activities which cannot be presented visually in CATI. Therefore, other options should be considered when these questions are re-designed to improve data quality. These options could include shortening the lists or presenting them under a series of sub-headings or clusters. If the changes are not implemented, it should be expected that item nonresponse for these questions could be increased reducing data quality as a result.

5.3.3 How could the self-completion aspect be managed?

As mentioned above, CATI on its own does not support a self-completion element. If self-completion elements are not introduced, it should be anticipated that social desirability bias will be present for some questions and data quality will be reduced. Some questions in the National Survey are at risk of social desirability bias: for example, the mental wellbeing scale (WEMWBS) questions, health and lifestyle items (alcohol consumption, physical activity, fruit and vegetable consumption, and smoking), gambling questions, and environmental behaviours questions. The lack of a self-completion element not only negatively affects the respondents' willingness to complete these specific items honestly, but also reduces comparability with other

data sources because these questions are often asked in other surveys using self-completion modes of data collection.

These issues could be overcome by introducing a self-completion element to the CATI survey for a specific set of questions. This element could be introduced in a form of a short web questionnaire (or a postal alternative for offline populations). The introduction of an additional self-completion element has already been trialled by the National Survey during the pandemic, with a web self-completion element being added to the CATI. During the CATI interview, respondents are provided with a specific web address to enable them to answer some questions privately. The questions asked online were those which would normally be administered via CASI. If CATI respondents were unable (or unwilling) to do the web component they are offered the option to do them via telephone instead. In this trial approximately 90% of respondents completed the former CASI module, 78% online and 22% continuing to use the CATI mode.

Therefore, having CATI as the primary mode of data collection, supplemented by a web self-completion element, is technically feasible. However, data users should have a clear understanding of which questions were collected using what mode of data collection. When a self-completion element is introduced in the form of a web questionnaire, we might anticipate the increased risk of nonresponse among specific groups in the population and accessibility issues for the self-completion part. There is a risk that older people and people with specific disabilities would have increased accessibility issues and might become non-respondents for this specific component of the survey.

There is a possibility of introducing a postal alternative for a self-completion element for offline respondents. There was a lot of experience in Welsh Health Survey (when it was used as a standalone survey) to collect health questions using paper questionnaire which might help with the transitioning of these specific questions to this mode of data collection if needed. It is important to mention that web self-completion element offered during the interview might have lower nonresponse rate in comparison to a postal alternative which would need to be completed after the interview.

It is difficult to draw a general conclusion on which approach would offer better data quality (a postal alternative with lesser mode effects but higher nonresponse; or a

telephone alternative with greater mode effects but lower nonresponse): which is best would be topic- and survey-specific.

5.4 Web-CATI questionnaire issues

It would be possible to implement a mixed-mode design which combines web and CATI modes of data collection for the entire questionnaire, and not just for the self-completion element of it.

The main advantage of a mixed-mode approach is that it increases accessibility in comparison to the use of either mode alone and consequently it may reduce unit nonresponse. The availability of an interviewer-administered CATI mode allows respondents with lower level of literacy and/or a lower level of digital literacy to take part. The availability of a self-administered web mode does not restrict respondents with, for example, hearing impairment from taking part. However, we need to keep in mind a potential introduction of measurement differences between modes and the potential need for adjustment for measurement differences at the stage of analysis which was discussed in section 5.2.4. Specifically, there is a risk of recency effects for CATI and primacy effects for web questionnaires for questions with long lists, and other potential issues are associated with presence or absence of an interviewer and presence or absence of visual aids, which may increase measurement error and measurement differences by mode. The measurement differences could be partially prevented and reduced by careful design of an instrument that is suitable for different modes. However, adjustments might still be needed, and it is important that data users are aware of the issue and able to perform adjustment during the data analysis stage.

5.5 New data collection modes

In this section we will briefly discuss recent innovations in data collection that have been trialled as a result of the Covid-19 pandemic. It should be noted that, as these are relatively new methods, most of the evidence related to these has been gathered from relatively small pilots, and full results of the trials have yet to be published. Therefore, there is an element of uncertainty (and risk) regarding how well these methods would work if they were applied to the National Survey.

5.5.1 Computer Assisted Video Interviewing

Computer Assisted Video Interviewing (CAVI) refers to interviewers being able to complete a questionnaire remotely with respondents, via a live video link. This

method had been trialled experimentally pre-pandemic as a theoretical alternative to web and telephone questionnaires (Schober, 2020) although not, so far as we are aware, as an alternative to CAPI. During the pandemic there was increased interest in the method due to restrictions on face-to-face fieldwork.

CAVI offers some theoretical advantages as a mode of data collection, in cases where face-to-face interactions are not permitted or practical:

1. Interviewers and respondents can see each other, and interviewers can provide assistance or motivation if they pick up on non-verbal cues that respondents are struggling with questions;
2. Unlike in CATI, interviewers can share visual materials with participants via screen-sharing technology e.g. showcards;
3. Having an interviewer visible may help prevent breakoff associated with survey fatigue.

It should be noted that this last point is not something that has been empirically established. It should also be noted that there are downsides of having an interviewer visible i.e. we would anticipate that CAVI conditions could create social desirability bias and other negative interviewer effects in a similar way to a CAPI interview.

Self-completion elements are possible within a CAVI interview to help with the administration of questions that are most prone to bias. For example, participants can be sent a link to a web-survey element that they can complete privately. This web-survey can be completed live (i.e. the interviewer is available to help if needed but the default is that interviewers cannot see the web-questionnaire window). This approach acts as the equivalent of a CASI module and has been trialled in some pilots already, for example a 2020 British Cohort Study 1970 pilot, and a National Child Development Survey pilot (Cole, 2021).

There are some obvious negatives to the CAVI mode. Most notable is the issue of internet access / use. As with web modes there are significant minority of people who do not have internet access or who do not use the internet regularly. In Wales 8% of households don't have internet access, and 7% of adults don't use the internet (National Survey, 2021). Of adults who do use the internet, 4% only use it weekly or less often (National Survey, 2021). Although video calling popularity has

increased dramatically compared to pre-pandemic levels, three in ten people do not use this technology regularly (OfCom, 2021). This suggests that a ‘CAVI’ option should only be seen as a supplementary mode for the National Survey, rather than a primary mode of data collection.

A 2021 pilot of the National Survey of Sexual Attitudes and Lifestyles (Natsal) trialled offering both video and telephone interviews as alternative modes to people who declined to take part in a face-to-face interview. Uptake of the video option was relatively low, accounting for just 7 of the 131 total completed interviews (publication forthcoming). The video alternative offer has now been dropped for the next stage of Natsal data collection.

5.5.2 Self-completion via tablets

Another novel mode of data collection trialled in the pandemic is a knock-to-nudge approach, where participants are asked to complete a self-completion questionnaire on a tablet. The European Social Survey is currently trialling this approach as an alternative to their CAPI survey (d’Ardenne, 2021). They are also trialling web and paper options (Hanson, 2021).

Tablet self-completions are not a new technology. Market research companies have been using this method in “intercept” surveys for a number of years. In intercept surveys participants are recruited from public locations (such shopping centres and public transports hub) and are asked to complete a relatively short series of questions on a given topic (usually less than 10 minutes of questions). Tablet self-completions are not usually used as the primary mode of data collection in address-based random sample surveys, as the questionnaires tend to be longer and more complicated.

The advantages of using knock-to-nudge with tablet self-completions are as follows:

- Tablets can work even when the device is offline, so they have some benefits of a web questionnaires (easy routing etc) without relying on the respondent having an internet connection;
- Interviewers are able to demonstrate the device on the doorstep to encourage, and to provide some basic training for people with lower digital literacy;
- As all questionnaires are completed on standardised devices there is greater control of questionnaire rendering and no risk of device effects;

- This mode of administration can work under lockdown conditions.

However, the downside of this approach is that it has not been fully trialled for longer PAF surveys and we have no quantitative evidence yet of how well it will perform, including what response rates using this method would be. Moreover, it is not clear what the benefit of this approach would be if CAPI in-home interviewing is permitted. The main theoretical benefit is, if web modes are being used as the first mode of data collection, that the tablet alternative acts as the closest possible equivalent for offline groups without introducing measurement effects. That said, the ESS pilots have demonstrated that some less digitally literate people may encounter issues using the tablets despite training given on the doorstep, and further work is required to quantify the size of this group (unpublished).

5.5.3 CAPI plus

The final mode of data collection that we have observed emerging post pandemic is what we are calling ‘CAPI plus.’ In this design CAPI is the default first mode of data collection, but an alternative mode is available for people who decline to take part, or for those who have concerns about letting interviewers in their homes (including those who are self-isolating). We have found that research commissioners are requesting information on offering a secondary mode alongside CAPI. The reasons for this are two-fold:

- There are concerns that the higher response rates normally associated with CAPI could decline further post-pandemic, for instance if people continue to have concerns about letting interviewers into their homes.
- There is a desire to have a secondary mode of data collection available in case of future regional or national lockdowns.

As this point CAPI plus modes are still being piloted, and there is only limited data available about their overall success and about the uptake of any secondary modes used.

5.6 Data linkage and recontact consents

Currently data linkage consent is not required for National Survey, as linkage is based on providing fair processing information. However, if the situation changes in the future, it is important to consider that consent for data linkage is usually lower for self-administered modes when compared to interviewer-administered ones (Jackle et al.,

2021). For web modes of data collection, where respondents can use different devices for survey completion, there is some evidence suggesting that the likelihood of consent to data linkage is higher for laptops and PCs when compared to smartphones (Maslovskaya et al., 2020). These findings are context specific and might not hold in other surveys or when linkage is required to other types of administrative data. The lower likelihood of consent to data linkage can reduce quality of linked data sources and should be addressed appropriately.

Consent to recontact would be especially important if the survey moved to a longitudinal format. The results from CATI testing in National Survey suggested that the consent to recontact has been equivalent or higher in the test when compared to CAPI across all age groups, which is an encouraging and reassuring result.

6 Data issues

This chapter looks at data issues present in the different design options.

6.1 Mode effects

It is well-known that responses to the same question asked in different modes may be systematically different, and these differences are known as mode effects. They consist of a selection effect, which captures how the characteristics of respondents on known variables vary according to what mode they choose to answer in, and a measurement effect, which captures the differences to the response to the substantive question engendered by the mode in which it is asked (and which could for example include primacy and recency effects in telephone vs other modes).

There are two risks to the results of a survey that uses multiple modes from these sources. The balance of responses in the different modes may mean that the measurement and the weighting interact in such a way that the estimates from the survey are different from the target estimate, leading to a bias in the cross-sectional estimates. And the balance of responses between the two modes may vary from survey occasion to survey occasion (which may be particularly important in times of unusual change such as that imposed by the Covid-19 lockdowns, but may also arise due to random variation in response mode from period to period), so that the time series is less well estimated.

Selection effects are generally worse for mail-only methods, because reluctant respondents are nudged only by postal reminders and not by interactions with an interviewer. There are additionally within-household self-selection issues for mail-contact only, knock-to-nudge and sequential CAPI, which rely on some activity of the respondent to undertake or assist with the within-household sampling procedure.

Schouten et al. (2022, chapter 9) summarise the approaches to dealing with these issues. The correction of the mode effect relies on an assumption that one mode is free of measurement error, a measurement error model, and an assumption that all the selection differences are accounted for by calibration to known auxiliary variables. The correction of the evolution of the time series can be done through a calibration process which treats each mode separately.

In specific situations it may also be possible to make adjustments for mode effects using time series models, particularly where there is a panel element with follow-up of existing cases (van den Brakel et al., 2021), but this seems less likely to be useful in the National Survey where sampling is cross-sectional (but see next section).

Although the methodology for making adjustments is available, the application of these methods requires some investment, first from the survey organisation to produce a set of estimates which account for the selection and mode effects as far as possible, and second from the data users (especially microdata users) to undertake principled analyses accounting for the differences in modes. It has not been found practical to adjust data directly, in order that users don't have to do further adjustment themselves. We know from experience that some users won't adjust for mode effects, for a variety of reasons, and that there is a risk that some analyses will therefore be based on biased information if the survey consistently uses a mixed mode approach.

6.2 Discontinuities

Changes in modes (or the balance between modes), questions and sampling all potentially lead to discontinuities in the time series of estimates produced from the survey. Such discontinuities are estimable. In an ideal situation where the discontinuity is planned (for example from a change in the questions to be included), the before and after situations can be used together in a parallel run, and this provides good information with which to estimate the effect of the change. It is also possible in a situation where the change is not planned and can only be observed. Both situations can be analysed using time series models which account for the correlation between measures from the National Survey at different times (van den Brakel et al., 2021), and an example of this kind of analysis was undertaken when the National Survey moved to its existing design (Smith et al., 2017). This example also included a combination with small area estimation methods to produce estimates of discontinuities for small domains, though with reduced precision.

For any substantial changes, it would be advisable to use the available methodology to make an estimate of the size of the differences, as an aid to users in interpreting time series of key estimates.

6.3 Small area estimation

Small area estimation is a term applied to a range of different approaches which make use of models to make improved estimates for domains where there are relatively few sample observations. For a detailed text see Rao & Molina (2015). Small area estimation is useful in specific situations where there are good predictor variables available which relate to the small domains/areas of interest, for example from administrative datasets. Where such variables exist there can be substantial gains in accuracy, but for at least some of the variables included in the National Survey there may be few predictors and little gain. The ONS has historically had success using small area estimation for unemployment variables, where the claimant count is a good predictor available at a detailed level. By contrast, several attempts at modelling adult psychiatric morbidity failed to identify suitable predictors to create satisfactory small area estimates.

In the National Survey, advantage could be taken of variables for which there are strong predictors, in which case the planned sample size could be reduced (perhaps in connection with a matrix sampling approach) as a means to control costs. It is not, however, without challenges. The need for modelling would likely increase the time needed to produce these estimates (though they need not necessarily be issued contemporaneously with the direct estimates from the survey). Interpretation of differences between small area estimates is not straightforward for users, and some users do not have the statistical background and training to draw appropriate inferences accounting for the quality and properties of the estimates. The results of the procedures are also model-dependent (though an objective description of the modelling approach and diagnostic checking is generally enough to reassure users). It is better to design the sample such that direct estimates have the required quality (and that is the aim of the minimum sample sizes within local authorities). But small area estimation can be considered as a supporting methodology within an overall design.

6.4 Longitudinal designs and estimation of change

The National Survey is designed as a cross-sectional survey, and therefore the estimates produced at each time point have independent errors. Comparisons between time points have variances which are the sum of the variances of the cross-sectional estimates, which makes inferences relatively poor. However, the same independence property means that cases can be accumulated over time to create

larger sample sizes, which allow direct estimation for smaller domains, but with a more blurred time interval.

A true longitudinal design would allow better (i.e. with smaller variance) estimation of changes and would also provide estimates of gross flows between states (e.g. of employment or health). It would reduce the benefit from accumulating cases over time (although an element of this could be retained by a judicious choice of longitudinal design), and has some additional operational costs, because it would need some activities for sample maintenance and keep-in-touch exercises between waves. The weighting and analysis procedures for longitudinal data are also considerably more complex.

7 Broader priorities and implementation

Most of the issues discussed in previous chapters have been on design factors related to data quality issues. Although data quality is of paramount importance to survey designers, there are other factors that are important to consider. For instance, the Total Survey Quality Framework (Biemer et al., 2014) describes how data quality considerations such as accuracy need to be weighed up against other factors including:

- **Cost-effectiveness:** It is important that the survey offers value for money both now and in the future. It is important that all spending decisions for the National Survey have a rational justification. At present the Welsh Government has a budget of £1.8 million plus VAT per annum to spend on the National Survey, and all designs considered need to fit within this budget.
- **Timeliness:** There can be distinct advantages to data users if survey results can be produced in shorter windows of time. The benefits of having results sooner may outweigh the need for the highest standards of data quality in certain circumstances.
- **Credibility:** Official statistics are subject to a relatively high degree of public scrutiny. It is important that the methods employed are demonstrably robust for statistics to be seen as trustworthy. Public and professional perception is important especially if the National Survey opts to use more innovative methods that are less ‘tried and tested.’ There could be reputational risks if new methods or technologies do not work as anticipated.

There are further ethical and practical issues to consider in order to ensure the National Survey design aligns with the Welsh Government’s broader policy objectives. These include:

- **Reducing the National Survey’s overall CO₂ emissions:** Stakeholders felt that this objective is essential given the Welsh Government’s declaration of a Climate Emergency and policies related to this.
- **Maximising the accessibility of the survey.** By this we mean ensuring that it is theoretically possible for all selected respondents to take part in the survey, *with minimal burden*, regardless of their personal circumstances. The survey design

should support participation from everyone including those who are not internet users, those who have health conditions or disabilities, and those who have lower levels of literacy.

- Having all survey materials available in both English and Welsh.

Finally, after their experiences in the pandemic, the Welsh Government are naturally keen to have a survey design that could work if lockdown conditions are ever necessary again in the future. Based on feedback from the stakeholder consultation this feature was considered to be 'highly desirable' but not as important as the other criteria raised.

In the following sections we will consider how the modes of contact and the data collection modes have an impact on:

- Timeliness of results;
- Accessibility;
- Whether the survey could work in the event of future lockdowns;
- CO₂ emissions; and
- Logistical considerations, such as whether the survey could easily be delivered by contractors and the level of risk involved.

Cost considerations will be addressed in Chapter 9, where indicative costs for two alternative designs are provided.

7.1 Timely reports

During the consultation phase of this project one re-occurring comment from data users was that it would be beneficial to have more frequent data outputs (i.e. quarterly reports rather than annual reports). There was also a desire for more flexibility to add 'ad hoc' questions mid-fieldwork i.e. to address emergent research questions based on current policy needs. The designs best suited for fast turnaround and ad hoc additions are those that do not involve face-to-face contact or fieldwork i.e. mail only contact modes followed by web or CATI data collection.

7.2 Accessibility and inclusivity

In addition to wanting to ensure representativeness, there is an ethical principle behind wanting to make surveys accessible to as many groups as possible. The Government

Statistical Service endorses user-centred design principles, with the aim of making surveys less burdensome for respondents. One important principle is that of inclusivity (Wilson, 2020); i.e. the principle of designing surveys that can be completed by anyone, regardless of their health, education levels or other personal circumstances. The benefit of designing for inclusivity is that in doing so you improve experience for all users, not just for the groups with additional needs.

Arguably, one mechanism for inclusive design is offering multiple modes of data collection. For example, web-only surveys may exclude offliners, those with lower levels of literacy, and those with lower levels of digital skills. In Wales, only 66% of people have five/five 'basic digital skills' (Office for National Statistics, 2019b). In contrast CATI only surveys can exclude those with hearing difficulties. Approximately 4% of adults have objective moderate-severe hearing loss (defined by being unable to hear a mid-pitch tone, in a quiet room, at the volume of 35 Db hl; Scholes & Mindell, 2014). Combining the two modes of data collection, therefore, should increase the inclusivity of the survey.

CAPI modes of data collection are viewed as a more accessible mode, as interviewers can tailor their approach based on the needs of the respondent. For example, interviewers are trained to read out showcards to those who have visual impairments and to those who cannot read well. Offering an additional mode to CAPI (i.e. CAPI plus) would also promote inclusivity, by allowing people to take part even if they have concerns about face-to-face interactions.

7.3 Future proofing against Covid-19

Future proofing against future lockdowns was considered a desirable (but not essential) feature. A mail-contact only design and a knock-to-nudge element would both be able to operate under lockdown conditions, although under the most rigorous lockdown conditions knock-to-nudge might have to be suspended. The CAPI element of sequential CAPI designs (and CAPI plus designs) would not be able to operate under lockdown conditions. However, as both designs include a supplementary data collection mode, they would be more able to pivot to just that alternative mode in the event of further lockdowns.

CAPI only designs could pivot to an alternative mode, but there may be issues with converting some questions to a new mode if this has not been done in advance. There

is an argument that going forward the Welsh Government should ensure that new questions are designed to be 'unimodal' i.e. that they could function equally well regardless of mode of administration. In order to fully future proof the questionnaire against any future mode changes then the Welsh Government could also consider redesigning existing questions to be unimodal, taking into account that this may involve trade-offs (e.g. breaks in time series, benchmarking with other studies, data quality).

7.4 CO₂ reductions

It is a priority for the Welsh Government to reduce CO₂ emissions. In the following section we will describe the various sources of emissions produced in surveys. Please note that this report is not designed to be an exhaustive account of emissions, as this task goes beyond the current scope of work.

Broadly speaking surveys contribute to CO₂ via three main mechanisms:

1. **Electricity usage.** This would include energy usage from heating offices, running electrical devices and running computer servers / cooling facilities. The exact extent of emissions produced via electricity is difficult to quantify. Electricity usage will occur across a range of third parties, including contractors and server providers. Survey agencies will share their facilities/server rooms across various customers so it is not possible to say how much electricity is being used for the National Survey in isolation from other projects. It should be noted that all the survey designs discussed in this report would generate these types of emission, regardless of the contact mode and mode of data collection. Emissions from electricity usage could be reduced if more parties committed to using renewable energy sources, however this may not be practical to enforce.
2. **Resource usage.** This would include any emissions created in the production, distribution and disposal of the physical resources required for the survey, including IT equipment, mailings, other printed documents, physical incentives (gift cards) and so on. The resources used would vary to some extent by survey design. For example, in mail-only contact surveys many more addresses are selected, so many more letters are sent compared to a CAPI survey (to compensate for the lower response rate for mail only surveys).
3. **Vehicle usage:** This would include the emissions produced via cars as part of survey fieldwork. Interviewer contact modes (knock-to-nudge and CAPI) all

involve interviewers travelling to randomly-selected addresses, and a relatively high amount of interview time is spent on travel. Interviewer travel is a major source of emissions for surveys.

It should be noted that although mail-only surveys use more of some physical resources (i.e. mass mailings), interviewer vehicle usage is associated with particularly high emissions. For example, newer models of vehicle (excluding electric vehicles) emit on average 122 grams of CO₂ per kilometre travelled¹⁸. In contrast, estimates of emissions per letter (cradle to grave, including impact on forestry) are between 20 and 50 grams, including the travel required for delivery.¹⁹

Table 7.4 below describes the different sources of emissions in various parts of the survey process, and how these emissions could be potentially reduced.

Table 7.4: Survey stage and potential ways to reduce CO₂ emissions

Sampling	<ul style="list-style-type: none"> • Reduce travel required for knock-to-nudge and CAPI by clustering the sample – the aim is to sample addresses in such a way that it improves fieldwork efficiency, • NOTE: tighter geographical clustering will reduce travel requirements for individual assignments but will make the sample less statistically efficient. • Refresh clustering algorithms (i.e. those used by statisticians for clustering addresses) on a more regular basis to improve efficiency of allocating assignments to interviewers.
Recruitment	<ul style="list-style-type: none"> • Reduce the number of letters sent to ‘deadwood’ addresses if possible. • Reduce the volume and weight of printed recruitment materials as far as possible. • Use fewer/lighter materials (e.g. smaller envelopes, lighter paper, fewer items in envelope). However, care would need to be taken to ensure new materials do not have a negative impact on response rates. • Use monochrome designs over colour designs (black ink produces fewer emissions than colour ink). Again care would need to be taken to ensure that monochrome designs do not influence response rates.

¹⁸ [New car carbon dioxide emissions](#)

¹⁹ [The Environmental Impact of Mail: A Baseline](#)

-
- Adopt simple design approaches (i.e. easy read documents, avoiding sending different versions to different groups etc). Keeping designs simple helps to avoid checking errors and wastage (e.g. re-prints).
 - Avoid pre-printing approaches where letters are printed in bulk for the whole survey period. Letter content is always susceptible to change over time (avoids potential for wastage)
 - Reduce the number of reminder letters by filtering out completed cases and opt-outs.

Briefings

- Try to avoid the need for interviewers to travel to briefings and/ or re-briefings. Adopt virtual briefing sessions instead.
- Avoid the number of hard copy documents produced for briefings (instruction booklets etc)
- Having a simple survey design will contribute to simple briefings that can be conducted remotely.

Fieldwork

- Reduce the number of trips to deadwood addresses if possible.
 - Make e-vouchers the default option for incentives. When physical vouchers are needed, use paper/card alternatives over plastic.
 - Having a larger pool of interviewers will reduce travel miles. Ask potential contractors to disclose the size of their fieldwork panel and consider pooling interviewers from more than one service provider to increase fleet.
 - Acknowledge that, for CAPI options, a significant amount of travel miles are generated trying to recruit the most hard to reach cases. Mileage increases disproportionately towards the end of an interviewer assignment, as interviewers make multiple trips to hard-to-contact households with lower success rates. There is a need to strike a balance between reducing emissions and methodological concerns (i.e. better survey estimates, representativeness of the achieved sample). There is evidence that demonstrates that making large numbers of calls to sampled addresses may increase overall response rates but it does little to change final survey estimates (Sturgis, 2016).
 - Minimise or avoid re-issuing of cases. Adopt focused re-issuing practices to recruit specific groups/areas that are underrepresented. There is some evidence to suggest that reissues have little impact on survey estimates once weighting for nonresponse has been applied.
-

For example, Hutcheson et al. (2020) found that general reissuing did not have any meaningful impact on survey estimates for the Scottish Household Survey.

Monitoring

- Fieldwork agencies collect information on interviewer mileage for pay and performance purposes. Consider requesting that number of mailings and overall mileage (and approximate carbon emissions) are presented alongside other key metrics such as the number of interviews achieved / response rates. This will make the carbon impact of projects more visible to stakeholders and will allow greater understanding of the trade-offs between increasing response rates and reducing emissions.
 - Having closer monitoring of trips to maximise efficiency (i.e. not repeating the time of day of unsuccessful trips etc)
-

7.5 Viability: novelty, complexity and risk

A final point to consider is practicability; that is the extent to which any new design can be successfully managed by the Welsh Government and any contractors appointed to run the survey on their behalf. Some of the survey designs discussed in this paper have a long-standing track record, namely mail-only contact methods and CAPI only methods. Some methods were gaining traction pre-pandemic, for example the Office for National Statistics (ONS) have been investing in trialling web-CAPI for some time for their Labour Market Survey (LMS). Other designs have been piloted mainly as a direct result of the pandemic e.g. knock-to-nudge methods, CAVI and CAPI plus. For newer designs less is known about performance, and contractors may be less able to deliver these methods at optimal efficiency.

If multiple modes of contact and/or data collection are used additional resource will be required from both the Welsh Government and contractors. There will be more process documents to produce and sign-off (procedures, mailings, interviewer instructions etc). If mixed modes of data collection are used, versions of questionnaires will need to be checked and signed off for each mode of data collection. There will also be greater effort required in fieldwork monitoring, data processing, checking and weighting. The more complicated design the greater the resource required, from both Welsh

Government and their contractors, and the higher the risk that errors could occur in practice.

8 The strengths and weaknesses of different designs

This review has indicated that there is no one single survey design that is the best fit for all of the quality criteria. Instead, each design has its own merits and its own trade-offs.

Which design is preferred will depend on which criteria the Welsh Government choose to prioritise. For example, designs involving face-to-face methods of contact (knock-to-nudge, sequential CAPI or CAPI first) will yield higher response rates and, potentially, achieve a more representative sample composition. This will help ensure that the data collected are seen as robust and therefore credible. That said overall response rates are not as important as having an unbiased sample. Changes in response rates do not always equate to meaningful changes in survey estimates (Sturgis et al., 2016).

Any designs with face-to-face contact will have higher running costs, will produce more CO₂ emissions and have less timely data outputs. Designs that involve multiple different modes of data collection (for example web-CATI, or sequential CAPI designs) may be viewed as beneficial as they reduce running costs and CO₂ emissions. At the same time, they provide potential respondents with more choice and control regarding how to take part (i.e. promoting accessibility) and will future-proof the survey in case of lockdowns or other triggers of a mode change. But mixing modes between respondents can introduce inadvertent biases to data collected, can make analyses more complicated, and introduces practical constraints on how easy it will be for Welsh Government and their contractors to run the survey.

Given these different trade-offs, in this chapter we will summarise the strength and weaknesses of different survey designs using a series of tables. The designs we will consider are:

1. Mail contact only designs (web only, CATI only and web-CATI);
2. Knock to nudge designs (web only, CATI only and web-CATI);
3. Sequential CAPI modes (web-CAPI and CATI-CAPI); and
4. CAPI (either as a standalone mode, or CAPI plus where an alternative mode is offered to refusers).

Figure 8.1: Trade-offs associated with mail-contact only designs with web or CATI data collection

Quality criteria	Rating			Comments
Sample size of 12,000	★★★			<ul style="list-style-type: none"> Sample size could be easily achieved within budget.
Representative sample	★			<ul style="list-style-type: none"> Lowest response rates. Most risks in terms of representation
Questionnaire length (45 mins)	★★			<ul style="list-style-type: none"> 45 minutes of content possible in web and CATI, however shorter lengths are recommended as default
Accessible	★★★ - Web	★★★ + CATI	★★★ Web-CATI	<ul style="list-style-type: none"> Web only designs not accessible to people with no internet access. They are also less accessible for those with lower levels of literacy and lower levels of digital skills. CATI only designs are more accessible to those with issues with literacy or digital literacy. However, they are less suited for those with hearing impediments. Mixed web-CATI designs are most accessible
Self-completion element	★★★ Web		★★★ CATI	<ul style="list-style-type: none"> CATI only designs can include a self-completion element online or on paper. However, having a separate component leads to more missing data.
Supports recontact research	★★★ Web	★★★ CATI		<ul style="list-style-type: none"> Lower recontact rates for web surveys/ surveys with web component
Data and analysis	★★★ Web or CATI only		★★★ Web-CATI	<ul style="list-style-type: none"> Analysis is more problematic if there is mode variation between respondents
Timely results	★★★			<ul style="list-style-type: none"> Mail only contact modes are the quickest in terms of data collection. This method could support quarterly reporting
Could work under lock-down conditions	★★★			<ul style="list-style-type: none"> Would work under all lockdown scenarios, as long as the postal system is still operating
Attempts to reduce carbon footprint	★★★			<ul style="list-style-type: none"> Largest reduction in CO₂ as no interviewer travel required
Manageable for WG / contractors	★★★			<ul style="list-style-type: none"> Relatively straightforward design for contractors to deliver

The main advantages for mail contact only designs compared to the current CAPI standard are that they can produce more timely results (e.g. it would be easier to

allow for quarterly reporting). They would see the largest reduction in mileage, and thus CO₂ emissions would be substantially reduced and would work under all lockdown conditions. However, the major drawback is the decrease in response rates and the increased risk of nonresponse bias.

Figure 8.2: Trade-offs associated with knock-to-nudge designs with web or CATI data collection

Quality criteria	Rating			Comments
Sample size of 12,000	★★★			<ul style="list-style-type: none"> Sample size could be achieved within budget
Representative sample	★★			<ul style="list-style-type: none"> Improved response rates compared to mail only methods.
Questionnaire length (45 mins)	★★			<ul style="list-style-type: none"> 45 minutes of content possible in web and CATI, but shorter lengths are recommended as the default
Accessible	★★★ Web	★★★+ CATI	★★★ Web-CATI	<ul style="list-style-type: none"> Web only designs are not accessible to offliners/people with lower levels of literacy/ digital skills. CATI only designs are more accessible but are less suited for those with hearing impediments. Mixed web-CATI designs are most accessible
Self-completion element	★★★ Web		★★ CATI	<ul style="list-style-type: none"> CATI-only designs can include a self-completion element online or on paper. However, having a separate component could lead to more missing data
Supports recontact research	★★ Web	★★★ CATI		<ul style="list-style-type: none"> Lower recontact permission rates when using web modes of data collection
Data and analysis	★★★ Web or CATI only		★★ Web-CATI	<ul style="list-style-type: none"> Analysis is more problematic if there is mode variation between respondents
Timely results	★★			<ul style="list-style-type: none"> Slower data collection compared to mail only designs.
Could work under lock-down conditions	★★★			<ul style="list-style-type: none"> Would work under nearly all lockdown scenarios
Attempts to reduce carbon footprint	★★			<ul style="list-style-type: none"> Some reductions in CO₂ as less interviewer travel is required than if all-CATI
Manageable for WG / contractors	★★			<ul style="list-style-type: none"> This is a relatively new survey design and information on performance is limited

Knock-to-nudge designs have been trialled in the pandemic in order to boost response rates to mail-only surveys, but without having to involve interviewers entering the home. They offer some benefits over mail-only contact surveys, most notably the improved response rates and the increased likelihood of achieving a representative sample. However, it is unclear to what extent this method will remain in use post-pandemic, now that in-home visits are allowed. There is a strong argument that once an interviewer has made contact on the doorstep the ideal option is to conduct a CAPI interview straight away, rather than risk drop out by referring householders to a different mode of data collection. The benefit of increased response is likely to outweigh the risk of response-mode variation (or measurement effects) especially if interviewer effects can be mitigated by offering CASI for the most sensitive items.

Figure 8.3: Trade-offs associated with sequential CAPI designs

Quality criteria	Rating		Comments
Sample size of 12,000	★ ★ ★		<ul style="list-style-type: none"> Sample size could be achieved within budget
Representative sample	★ ★ ★		<ul style="list-style-type: none"> Improved response rates compared with mail-only and knock-to-nudge methods.
Questionnaire length (45 mins)	★ ★		<ul style="list-style-type: none"> 45 minutes of content is possible in web or CATI modes, but shorter lengths are recommended as the default
Accessible	★ ★ ★		<ul style="list-style-type: none"> Offering different modes improves accessibility, as does the provision of a face-to-face interaction (i.e. interviewers can tailor assistance to meet respondent needs)
Self-completion element	★ ★ ★ Web-CAPI	★ ★ CATI-CAPI	<ul style="list-style-type: none"> CATI designs can include a self-completion element online or on paper. However, having a separate component could lead to more missing data Self-completion possible in CAPI using CASI or paper forms
Supports recontact research	★ ★ Web-CAPI	★ ★ ★ CATI-CAPI	<ul style="list-style-type: none"> Lower recontact permission rates when using web modes of data collection
Data and analysis	★ ★		<ul style="list-style-type: none"> Analysis is more problematic if there is mode variation between respondents.
Timely results	★ ★		<ul style="list-style-type: none"> Slower data collection compared to mail-only designs.
Could work under lock-down conditions	★ ★		<ul style="list-style-type: none"> CAPI element would not work under lockdown conditions. However, in this case the design could pivot to 'knock-to-nudge'.
Attempts to reduce carbon footprint	★ ★		<ul style="list-style-type: none"> Some reductions in CO₂ compared with CAPI-only as less interviewer travel is required.
Manageable for WG / contractors	★ ★		<ul style="list-style-type: none"> This is a relatively new survey design and information on performance is limited.

Sequential CAPI modes offer most of the benefits of CAPI only modes, with the added advantage they would offer some reductions in CO₂. However, as a relatively new mode of data collection there is greater risk that contractors will not be able to deliver the work or there are unanticipated fieldwork issues. The mixed mode design will add complexity to analysis.

Figure 8.4: Trade-offs associated with CAPI only and CAPI plus designs

Quality criteria	Rating		Comments
Sample size of 12,000	★ ★ ★ -		<ul style="list-style-type: none"> Sample size could be achieved. However, as CAPI is the most expensive data collection mode it is becoming more challenging to deliver this within budget
Representative sample	★ ★ ★		<ul style="list-style-type: none"> Highest response rates. Lowest risk in terms of errors of representation.
Questionnaire length (45 mins)	★ ★ ★		<ul style="list-style-type: none"> Longer questionnaires are more suited to CAPI modes
Accessible	★ ★ ★ CAPI only	★ ★ ★ + CAPI +	<ul style="list-style-type: none"> Having a face-to-face interaction improves accessibility (i.e. interviewers can tailor assistance to meet respondent needs). Offering an alternative mode may make the survey appealing to those who wish to avoid having an interviewer in their home
Self-completion element	★ ★ ★		<ul style="list-style-type: none"> Self-completion are possible in CAPI using CASI or paper forms
Supports recontact research	★ ★ ★		<ul style="list-style-type: none"> Higher recontact permission rates
Data and analysis	★ ★ ★ CAPI only	★ ★ CAPI +	<ul style="list-style-type: none"> Analysis is more problematic if there is mode variation between respondents (i.e. if an alternative mode is offered for CAPI+)
Timely results	★ ★		<ul style="list-style-type: none"> CAPI is the slowest data collection mode.
Could work under lock-down conditions	★ CAPI only	★ ★ CAPI +	<ul style="list-style-type: none"> CAPI does not work under lockdown conditions. CAPI+ would allow more options for the design to pivot if a lockdown occurs.
Attempts to reduce carbon footprint	★		<ul style="list-style-type: none"> No reductions in mileage/ CO₂ if maintaining the current CAPI design.
Manageable for WG / contractors	★ ★ ★ CAPI only	★ ★ CAPI +	<ul style="list-style-type: none"> CAPI only has a long-standing track record as a successful design. CAPI + is a relatively new survey design and information on performance is limited

The above table compares the differences between a CAPI only design, and a CAPI+ design (in which participants are approached to do a CAPI interview but are offered an alternative mode if they decline to take part). Prior to the pandemic CAPI designs have been considered the 'gold-standard' due to their high response rates and the amount

of data it is possible to collect with a trained interviewer present. However, the cost of conducting CAPI fieldwork is high compared to other modes of data collection. Retaining the CAPI design would mean no reductions in interviewer mileage and associated CO₂ emissions. Please see section 7.2. for examples of how CO₂ could be reduced across all survey designs including CAPI surveys.

9 Indicative costs of different approaches

After discussions with the Welsh Government, there was an interest in exploring two of the alternative designs further in terms of general costs. These were:

1. A sequential CAPI design; with mail contact / push-to-web as the first mode and a CAPI follow up for non-responders.
2. A sequential CAPI design; with mail contact / push-to telephone as the first mode and a CAPI follow-up for non-responders.

The costs provided in this document are indicative and exclude VAT. The purpose of showing cost estimates is so that the Welsh Government can see the general scale of fieldwork costs for surveys using the alternative methods. The costs do not present a formal and binding quote from NatCen. It should also be noted that survey costs are dependent on a wide range of factors including: length of interview, what pre-testing takes place, what engagement strategies are used response (mailings, incentives), the response rate assumptions used and so on. We have provided details of what assumptions we have used in order to generate our cost estimates.

The costs provided assume fieldwork would take place in 2023. Should fieldwork dates change, costs would need to be adjusted dependent to take account of inflation. There will be variation in quotes between different fieldwork agencies. Therefore, as with all surveys, we would recommend that the new survey is commissioned via a competitive tendering process rather than by approaching a single agency.

9.1 Approximate costs for a web-CAPI design

The approximate fieldwork cost for conducting a web-CAPI survey of 12,000 respondents would be £1.3 million.

9.1.1 Summary of design

The cost provided is based on the following design assumptions.

Sample frame: PAF. Unclustered sample.

Target completes: 12,000 individuals (10,022 households)

Selection: Two adults in web condition (any two); One adult in CAPI condition (Kish)

Response rate assumptions: Web: 22% (household level); CAPI: 25%; overall: 40.5% (household level)

Deadwood: 10.7%

Questionnaire length: 40 mins (same as 2021-22 survey)

Batching: monthly

Translation: Translation provided by the Welsh Government; programming of Welsh version required. Assume 3% of interviews are in Welsh

Incentive: £15 conditional (60% encashment rate)

Mail outs: Advance letter (A4) + colour leaflet (tri-fold A4). White C5 envelope. Postage 2nd class. Reminder postcard and letter. Called today cards where contact is not made. Conversion letter. All materials to be double-sided: English on one side; Welsh on the other side

Weights provided: a household weight, an adult weight, a population weight and a child weight.

All selected addresses would be sent a letter and leaflet and, if no response, a reminder postcard. The invitation letter would encourage respondents to complete the survey online. After a defined period, all non-responding addresses would be sent a further reminder letter, encouraging completion online and letting respondents know that an interviewer may visit. Up to two adults can complete online (no random selection, any two adults).

After a defined period, interviewers would visit non-responding addresses to encourage participation and to complete interviews in CAPI. Interviewers would make up to three visits per address. At addresses, the interviewer would make a random selection of one dwelling unit and one individual.

9.2 Approximate costs for CATI-CAPI design

The approximate fieldwork costs for conducting a CATI-CAPI survey of 12,000 respondents would be £1.6 million.

9.2.1 Summary of design

The cost provided is based on the following design assumptions.

Sample frame: PAF. Unclustered sample

Target completes: 12,000 (households and individuals)

Selection: One adult in CATI (next birthday); One in CAPI (Kish)

Response rate assumptions: CATI: 22%; CAPI: 25%; overall: 42.1%

Deadwood: 10.7%

Questionnaire length: 40 mins. (25min CATI; 15min web self-completion); 40 min CAPI (25 min CAPI + 15 min CASI).

Batching: monthly

Translation: Translation provided by the Welsh Government; programming required. Assume 3% of interviews are in Welsh

Incentive: £15 conditional (60% encashment rate)

Mail outs: Advance letter (A4) + colour leaflet (tri-fold A4). White C5 envelope. Postage 2nd class. Reminder postcard and letter. Called today cards where contact is not made. Conversion letter. All materials to be double-sided: English on one side; Welsh on the other side

Weights provided: a household weight, an adult weight, a population weight and a child weight.

All selected addresses would be sent an invitation letter and leaflet. Later a reminder postcard would be sent. The invitation letter would encourage respondents to opt-in for a telephone interview by either online portal; telephone; or email. After a defined period, all non-responding addresses would be sent a reminder letter, encouraging opt-in for a telephone interview and letting respondents know that an interviewer may visit.

For cases that opt-in, an interviewer would call and select one adult per household to take part, using the next birthday method.

After the CATI interview respondents would be required to complete a 15 minute web survey (the online element will be explained in the introduction to the telephone survey). At the end of the telephone questionnaire, the interviewer would give the respondents the same 12-digit Unique Access Code that was provided in the advance letter. The interviewer would also provide respondents with a link to complete the online element. The interviewer would give this link verbally over the phone, and also by email and text message. Where respondents do not complete the online element immediately, they would be reminded by the interviewer, first by text message, then by email, and finally by a phone call.

After a defined period, interviewers would visit non-responding addresses, to encourage participation and complete interviews in CAPI. Interviewers make up to 3 visits per address. At addresses, the interviewer would make a random selection of one dwelling unit and one individual.

10 Reference list

- Biemer, P., Trewin, D., Bergdahl, H. and Japac, L., 2014. *A System for Managing the Quality of Official Statistics*. *Journal of Official Statistics (JOS)*, 30(3).
- Bogdan, A., 2021. [Learnings from transitioning face-to-face studies to push-to-web](#). Presentation at GenPopWeb2 event, 17 September 2021. [Accessed 12 November 2021].
- van den Brakel, J.A., Smith, P.A., Elliott, D., Krieg, S., Schmid, T. and Tzavidis, N., 2021. Assessing discontinuities and rotation group bias in rotating panel designs. In P. Lynn (ed.) *Advances in Longitudinal Survey Methodology*, pp.399-423. Wiley, Hoboken.
- Castaldo A., Nikolakis, D., 2018 *Assessing the Use of an Address Based Design for the 2021 Census Coverage Survey*. Version 4, 17 May 2018 (unpublished).
- Chipperfield, J.O., Barr, M.L. and Steel, D.G., 2015. *Split Questionnaire Designs: are they an efficient design choice?* Proceedings of the International Statistical Institute 2013, Hong Kong.
- Chipperfield, J.O. and Steel, D.G., 2011. Efficiency of split questionnaire surveys. *Journal of Statistical Planning and Inference*, 141(5), pp.1925-1932.
- Cleary, A., Lynn, P., Cernat, A. & Nicolaas, G., 2017. The viability of a push-to-web survey design in 28 EU member states: the new Fundamental Rights Survey. Presentation given at the International Workshop on Household Survey Nonresponse, 30 August – 1 September 2018, Utrecht, Netherlands.
- Cole, K., 2021, Video calling as an alternative mode to CAPI for longitudinal surveys during the pandemic, Understanding Society Conference paper, September 27th 2021
- Cornick, P., Keyes, A., Swannell, B., Templeton, I., Woolfe, E., Yarde, J., 2021. [National Travel Survey 2020 Technical Report](#). Department of Transport. [Accessed 8 November 2021].
- d'Ardenne, J., Collins, D., Doušak, M., Fitzgerald, R., Martens, M. van Geisen R., 2021 Developing an electronic device for use on the ESS in absence of an interviewer by people without access to the internet, Presentation given at the ESRA conference, July 9th 2021
- Dillman, D., Smyth, J. & Christian, L., 2014. *Internet, Phone, Mail and Mixed-Mode Surveys: The Tailored Design Method*. 4th ed. Hoboken, NJ: Wiley.
- Galesic, M., & 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.
- Hanson, T., 2021. [Developing self-completion instruments for the European Social Survey](#). Presentation at GenPopWeb2 event, 14 July 2021, Online. [Accessed 12 November 2021].

- Hutcheson, L., Martin, C. and Millar, C. 2020. [Response rates, reissuing and survey quality: Does reissuing reduce nonresponse bias in the Scottish Household Survey](#), Scottish Government. [Accessed 17 December 2021].
- Humphrey, A. & Park, A., 2014. [Mixed-mode surveys of the general population: Results from the European Social Survey mixed-mode experiment](#). [Accessed 6 April 2022].
- Ioannidis, E., Merkouris, T., Zhang, L.C., Karlberg, M., Petrakos, M., Reis, F. and Stavropoulos, P., 2016. On a modular approach to the design of integrated social surveys. *Journal of Official Statistics*, 32(2), pp.259-286.
- Ipsos MORI, 2018. [Labour Market Survey Response rate experiments - Report for Test 1: Materials experiment](#). Prepared by Ipsos MORI for the Office for National Statistics. [Accessed 6 April 2022].
- Jäckle, A., Burton, J., Couper, M. P., Crossley, T. F., & Walzenbach, S., 2021. [Understanding and improving data linkage consent in surveys](#). Understanding Society at the Institute for Social and Economic Research. (No. 2021-01). [Accessed 12 November 2021].
- Klausch, T., Schouten, B., Buelens, B., & Van Den Brakel, J. (2017). Adjusting measurement bias in sequential mixed-mode surveys using re-interview data. *Journal of Survey Statistics and Methodology*, 5(4), 409-432.
- Klausch, T., Schouten, B., & Hox, J. J. (2017) Evaluating bias of sequential mixed-mode designs against benchmark surveys. *Sociological Methods & Research*, 46(3), 456-489.
- Kolenikov, S. and Kennedy, C. (2014). Evaluating Three Approaches to Statistically Adjust for Mode Effects, *Journal of Survey Statistics and Methodology*, 2 (2), 126–158.
- Lavrakas, P.J., and Bauman, S.L. (1993). The Last-Birthday Selection Method and Within-Unit Coverage Problems. Proceedings of the American Statistical Association, Section on Survey Research Methods, 2, 1107-1112.
- Lutig, P., 2021 [Moving a long survey online: Problems and some potential solutions](#). Presentation at GenPopWeb2 event, 14 July 2021. [Accessed 12 November 2021].
- Lynn, P., 2020. [Evaluating push-to-web methodology for mixed-mode surveys using address-based samples](#). Survey Research Methods, 14: 19 -30. [Accessed 6 April 2022].
- Maslovskaya, O., Smith, P.W.F., & Durrant, G., 2020. [Do respondents using smartphones produce lower quality data? Evidence from the UK Understanding Society mixed-device survey](#). NCRM Working Paper, University of Southampton. [Accessed 12 November 2021].
- Mavletova, A., & Couper, M. P., 2016. Device use in web surveys: The effect of differential incentives. *International Journal of Market Research*, 58(4), 523-544.

- Merkouris, T., 2010. [An Estimation Method for Matrix Survey Sampling](#). *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 2010, pp.4880-4886. [Accessed 12 November 2021].
- Merkouris, T., 2015. An efficient estimation method for matrix survey sampling. *Survey Methodology*, 41(1), pp.237-263.
- Nicolaas, G., 2012. [Mixing Modes within a Social Survey: Opportunities and constraints for the National Survey for Wales](#). A report prepared for the Welsh Government. [Accessed 24 February 2022].
- OfCom, 2021. [Online Nation](#). [Accessed 12 November 2021].
- Office for National Statistics. 2019 [ONS working paper series no 17 - Using data science for the address matching service](#). ONS: Newport. [Accessed 12 December 2021].
- Office for National Statistics. 2019b [Exploring the UK's digital divide](#) ONS: Newport. [Accessed 12 December 2021].
- Office for National Statistics, 2020a. [Labour Market Survey: technical report](#). [Accessed 12 December 2021].
- Office for National Statistics, 2020b, [Labour Market Survey: characteristics report](#): Characteristics Report. [Accessed 12 December 2021].
- Office for National Statistics, 2020. [Labour Market Survey: comparative estimates report](#). [Accessed 12 December 2021].
- Rao, J.N. and Molina, I., 2015. *Small area estimation*. John Wiley & Sons, New York.
- Renssen, R.H. and Nieuwenbroek, N.J., 1997. Aligning estimates for common variables in two or more sample surveys. *Journal of the American Statistical Association*, 92(437), pp.368-374.
- Revilla, M., & Ochoa, C., 2017. Ideal and maximum length for a web survey. *International Journal of Market Research*, 59(5).
- Scholes, S., & Mindell J., 2014, [Health Survey for England 2014, Chapter 4: Hearing](#). [Accessed 6 April 2022].
- Schober, Michael F., Frederick G. Conrad, Andrew L. Hupp, Kallan M. Larsen, Ai Rene Ong, and Brady T. West., 2020. "[Design Considerations for Live Video Survey Interviews](#)." *Survey Practice* 13 (1). [Accessed 6 April 2022].
- Schouten, B., van den Brakel, J., Buelens, B., Giesen, D., Luiten, A. and Meertens, V., 2022. *Mixed-Mode Official Surveys: Design and Analysis*. CRC Press, Boca Raton.
- Smith, P.A., Tzavidis, N., Schmid, T., Rojas-Perilla, N. and van den Brakel, J., 2017. Identifying potential discontinuities in the new National Survey for Wales. Commissioned report to the Welsh Government.

- Sturgis, P., 2004. [Analysing Complex Survey Data: Clustering, Stratification and Weights](#). Social Research Update, Issue 43, University of Surrey. [Accessed 6 April 2022].
- Sturgis, P., Williams, J., Brunton-Smith, I. and Moore, J. 2016. Fieldwork effort, response rate, and the distribution of survey outcomes: a multi-level meta-analysis. *Public Opinion Quarterly*. 81.
- Tipping, S. and Nicolaas, G., 2006. In search of a population sampling frame for UK postal surveys. *Survey Methodology Bulletin*, 58, pp.34-46.
- TNS BMRB, 2013. Community Life Survey: Summary of web experiment findings. Prepared by TNS BMRB for the Cabinet Office.
- Vannieuwenhuyze, J.T.A., Loosveldt, G and Molnberghs, G. (2014). Evaluating Mode Effects in Mixed-Mode Survey Data Using Covariate Adjustment Models, *Journal of Official Statistics* 30(1), 1-21.
- Williams, J., 2015. [Community Life Survey: Investigating the feasibility of sampling all adults in the household](#). TNS BMRB Report. [Accessed 6 April 2022].
- Wilson, L. 2020 *User Centred Design approach to Surveys*: GSS publication: User-centred design approach to surveys – GSS (civilservice.gov.uk) Retrieved 10.11.2021

Appendix A: Within-household respondent selection methods

Random selection of one respondent at each address is not straightforward for push-to-web surveys. As the sampling frame contains no information about who lives at the sampled address, the letter needs to include an instruction on who in the household is selected to take part in the survey. If only one adult is being selected, then a prescribed method for selecting the one adult is crucial as most households (about 65%) consist of more than one adult and self-selection is likely to bias the survey results.

A commonly-used approach for self-completion surveys is the **previous or next birthday** method which will produce a quasi-random sample of household members. However, experiments carried out on the Community Life Survey and the European Social Survey have shown that about 20% to 25% of respondents in online survey designs are not the target respondent (Humphrey & Park, 2014; Williams, 2015). This equates to about one in three households which contain more than one adult not carrying out the selection correctly. It is very likely that the importance of following the instructions is not understood or accepted by many people, particularly when household members other than the target respondent are ready and willing to take part.

Given that non-compliance with written instructions is relatively high, an alternative approach is to not provide instructions in the letter and to allow **any adult** to complete the questionnaire. This may not be an issue when asking questions about the household which can be answered fully and accurately by any household member (e.g. number of rooms functioning as a bedroom). However, for most National Survey estimates, there is a risk of bias as the respondents who choose to participate will not fully represent the wider population.

Another approach is to ask **all adults** in the household to take part in the survey. This approach has been used on a small number of push-to-web surveys, albeit with a cap on four participating household members for practical reasons (e.g. maximum number of unique login details that can be provided in the letter). This removes the need for any selection of adults in nearly all households, with only 1% of households having more than four adults resident. However, surveys relying on mail contact tend to use conditional incentives to achieve an acceptable response rate and this approach has the potential for fraud whereby additional adults are fabricated in order

to get the incentives. Some development work carried out in advance of the first year of the Active Lives Survey found that 4% of addresses filled in the survey for more adults than living at the address, and the average size of participating households was higher than would be expected; 2.19 compared to an average household size of around 1.8 (unpublished). This approach may also encourage more proxy responses when other household members are not immediately available or hesitant to take part.

With the “**up to two adults**” approach, respondent selection is only required among those households that contain more than two adults; i.e. about 15% of households compared with 65% of households if only one adult is selected per household. The next or previous birthday approach could be used to select the two adults. Or alternatively, any two adults in the household could complete the questionnaire. Although the “any two” approach will introduce some risk of self-selection bias, this is unlikely to have any notable impact on representativeness because 93% of the sample would be the adults that would have been selected using the next/previous birthday method and we know that about one in three would not have followed the next/previous birthday instruction correctly anyway.

There is still some scope for fraud when the “up to two” adults approach is coupled with a conditional incentive, but it is reduced because it is only possible for single adult households to fabricate an extra adult for an additional incentive²⁰ and the monetary reward is less tempting than when it is possible to fabricate up to three additional adults. There will be a small loss in precision due to clustering but any loss in precision is in theory likely to be more than outweighed by the gain in precision from having less-variable selection probabilities. Furthermore, the clustering effect within households will be almost negligible for those estimates reported by sex because the majority of households with more than one adult comprise of one male and one female adult.

Typically, the invitation letter will state how many adults can complete a questionnaire, and unique login details are provided for each potential respondent. There is some concern that this instruction and multiple login details are not easy to

²⁰ An extra adult can only be fabricated in single adult households which account for 35% of all households, compared to 95% of households which have fewer than four adults.

comprehend, can be off-putting, and add clutter to the letter which distracts from the motivation or reason for taking part.

An alternative approach is to provide a single set of login details and a request for any adult in the household to go online and complete the questionnaire. The questionnaire will include one or more questions to ascertain whether there are more eligible adults in the household and, if so, additional adult(s) are invited to also complete a questionnaire – either all, any adult(s) or randomly selected adult(s). The results from an experiment with the two-step approach on the Fundamental Rights Survey Pilot suggest that this may increase household-level response but the results were inconclusive for within household response rates (Cleary et al., 2017). Lynn (2020) found no difference in response rates but did find that the two-step approach reduced the proportion of households with four or more adults which is consistent with the hypothesis that the one-step approach to getting multiple adults to take part coupled with incentives may encourage fraudulent recording of extra household members.

Figure A1: A summary of within household selection methods

Method	Examples	Random selection	Ease of instruction	Compliance	Risk of selection bias	Clustering	Other issues
Kish	BSA	random selection	difficult to explain	1 in 4 non-compliant	Yes, some	No	Will a complicated instruction reduce response?
Last/Next bday	BSA, CLS, BES	quasi-random	easy instruction	1 in 4 non-compliant	Yes, some	No	Some recommend alternating last and next birthday
Any eligible adult	Flood Insurance	self-selection	easy instruction	n/a	Yes	No	May be acceptable for some household surveys
All eligible adults	CLS, SoL	n/a	easy instruction	n/a	No	Yes	Clustering. "Extra" adults fabricated when coupled with a conditional incentive.
Up to 2 eligible adults	FLS, ALS, Bike Life(?)	self-selection	not too difficult	n/a	Negligible	Yes	Small loss of precision due to clustering but likely to be more than outweighed by the gain in precision from having less variable selection weights.

Figure A2: A summary of two-step approach to selecting respondents

2-step approaches	Examples	Random selection	Ease of instruction	Compliance	Risk of selection bias	Clustering	Other issues
1. Any adult to go online or phone up 2. Random selection of one adult online or by phone		random selection	Easy instruction in letter but selection of other adult online or by phone could be problematic	If second stage is online, then there may be some non-compliance (no evidence)	If second stage is online, then there may be some self-selection bias (no evidence)	No	
1. Any adult to go online or phone up 2. Any adult plus one randomly selected adult to complete questionnaire	Fundamental Rights Survey Pilot	Self-selection with random selection of second adult	Easy instruction	If second stage is online, then there may be some non-compliance (no evidence)	If second stage is online, then there may be some self-selection bias (no evidence)	Yes	Small loss of precision due to clustering but likely to be more than outweighed by the gain in precision from having less variable selection weights.

