Dadansoddi ar gyfer Polisi

YMCHWIL GYMDEITHASOL Y LLYWODRAETH

GSR

GOVERNMENT SOCIAL RESEARCH

Analysis for Policy

Llywodraeth Cymru
Welsh Government

www.cymru.gov.uk

Ymchwil gymdeithasol
Social research

Rhif/Number: 101/2014

Options for carrying out large-scale surveys in Wales

Options for carrying out large-scale surveys in Wales: A report for the Welsh Government

Patten Smith, Ipsos-MORI
Gerry Nicolaas, Ipsos-MORI
Patrick Sturgis, National Centre for Research Methods

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

For further information please contact:
Chris McGowan
Knowledge and Analytical Services
Welsh Government
Cathays Park
Cardiff
CF10 3NQ
Tel: 029 2080 1067
Email: chris.mcgowan@wales.gsi.gov.uk

Welsh Government Social Research, 27 November 2014
ISBN 978-1-4734-2551-4

Table of contents

Executive summary

The Welsh Government (WG) and WG-funded bodies are committed to the collection of high quality survey data to ensure that their decisions and actions take into account the views, experiences and circumstances of people across Wales. However, the cost of conducting high quality surveys of the general population continues to increase at a time that WG and WG-funded bodies are facing financial challenges, prompting consideration of alternative ways of collecting survey data at reduced cost while balancing potentially negative effects on data quality.

It is in this context that in May 2014, the WG appointed Ipsos-MORI, in collaboration with Professor Patrick Sturgis, to undertake a review of the way in which large scale surveys commissioned by WG and WG funded bodies could be conducted. In addition to examining ways of reducing the cost of individual surveys without compromising the quality of outputs, the review also explores the scope for efficiencies across surveys.

The surveys included in this review are the WG-funded National Survey for Wales and Welsh Health Survey, the Sport Wales funded Active Adults Survey, the Arts Council of Wales funded Arts in Wales Survey and the Natural Resources Wales funded Wales Outdoor Recreation Survey. Other in-scope surveys include some which are conducted periodically and are wholly or partly funded by WG, such as the Infant Feeding Survey, the National Diet & Nutrition Survey, the Dental survey, and the European Health Interview Survey. Additionally the review includes a possible future housing condition survey. Finally, there is interest in exploring whether it may be possible to develop a flexible survey vehicle for collecting ad hoc needs in an omnibus-style manner.

The project involved desk research on the design and focus of the Welsh survey portfolio, interviews with key stakeholders, and a review of similar initiatives that have been undertaken in other country contexts. Following this work, the project team developed three broad approaches to redesigning the Welsh survey portfolio: (i) use of cheaper data collection methods, (ii) a reduction in the total number of interviews, and (iii) survey integration. Each of these broad approaches contains between two and four more detailed variants, making a total of nine redesign options. These nine options do not represent an exhaustive list of possible survey redesigns but provide an overview of options in general terms which can aid further thinking and development of other options.

Detailed descriptions for each of the nine options have been provided in this report, along with examples of how they might be implemented and approximate estimates of cost savings. We have also assessed the advantages and disadvantages of each one according to a set of essential and desirable criteria that have been agreed with WG.

Looking at the first broad approach of using cheaper data collection modes, considerable cost savings could be made by eliminating the need for interviewer visits to households by moving some or all in-scope surveys from face-to-face contact and data collection to telephone, postal or web modes. However, we have demonstrated that these cost savings come at the expense of data quality: low response rates with an increased risk of non-response bias, self-selection bias when contacting households by post, difficulties of including mobile only households and low response rates for telephone surveys, a reduction in the usefulness of data due to shortening and simplifying of questionnaires, risk of measurement differences and notable breaks in established time series. Any change in data collection modes would therefore require substantial investment in the management of the transition, including the parallel administration of the old survey design and the redesigned survey for a sufficient period of time. For these reasons, we do not recommend proceeding with the option of using cheaper modes unless the required cost savings cannot be met in any other way.

Because the key cost driver in face-to-face interview surveys is interviewer time, substantial savings could be made by reducing the total number of interviews that are undertaken across the in-scope surveys while preserving all other survey design features. This could be achieved by either reducing the frequency of data collection on one or more surveys or by lowering annual achieved sample sizes, either for the survey as a whole, or for particular question blocks or topic areas. There would be little concern regarding breaks in time-series due to changes in measurement properties and/or sample composition, and the level of planning and resource required to manage the transition would be minimal. However, both of these options would considerably reduce the timeliness of key statistics (particularly at geographical levels below national) and new statistics needed for emerging policy needs. It is therefore unlikely that this will be an acceptable option for many stakeholders.

The final approach to reducing survey costs involves survey integration, ranging from the introduction of a core module of questions across two or more surveys to full integration of all in-scope surveys.

A core module of questions for two or more surveys will only deliver cost benefits when survey sample sizes are reduced and key questions requiring large sample sizes are included in the core module. In practice, it is likely that a significant proportion of the core module will comprise demographic variables and that there will be limited space for more policy-relevant outcomes. Survey stakeholders would have to make considerable sacrifices in their ability to produce local area estimates if this option was adopted.

Integrating the National Survey and WHS within a single survey vehicle is a workable option which has the potential to make very significant cost savings. However there are substantial methodological issues which would need to be considered in detail. In particular, the within-household sampling procedure would have to change,

certainly for WHS and possibly for the National Survey. The required changes in design are likely to have a larger impact on WHS than the National Survey, and there would be a clear need for pre-operational stages of pre-testing and piloting to test the impact of integration on response rates, question context effects, and other possible changes in measurement which could have an impact on time series. It would also be prudent to implement the integrated design in parallel with the old survey design in order to assess, and potentially correct for, breaks in time-series.

Integrating all in-scope surveys in a single survey vehicle using a modular questionnaire approach without cutting sample sizes and topic coverage is the most ambitious option discussed in this review. It would not only make significant cost savings but it would also improve the quality of survey data currently being collected using non-random samples and other modes. It allows for the inclusion of an ad hoc data collection module as a matter of design rather than as an add-on which has the advantage that it would enable ad hoc data to be collected to higher quality standards – mainly because it would be based on higher response rate interviews and a random probability sample.

We feel that a single survey vehicle has the greatest potential to reduce costs substantially without significantly compromising quality or the ability to disaggregate data geographically. However, a fully integrated survey with no cuts in sample sizes and/or topic coverage[1] may still not deliver the magnitude of cost savings required and it is also very likely that this large survey vehicle could only be delivered by a single consortium of two or more major survey houses resulting in a de facto monopoly for fieldwork delivery.

These difficulties can be avoided by allowing cuts to be made in the amount of questionnaire material fielded and / or the number of respondents to which the questions are fielded. Reductions in sample size will be compatible with continuing to provide precise LA level estimates if the Williams Commission recommendation to reduce the number of Local Authorities in Wales is implemented. A reduced integrated survey is developed which has the following characteristics:

1. A reduction in the number of respondents/frequency of data collection (relative to the current situation) for many current questions;
2. Face-to-face interviewing with some CASI sections as required;
3. 40-45 minute interviews;
4. 12,000 interviews annually distributed evenly across local authorities.

Survey questions would be divided into c*ore* questions asked of all respondents (demographic questions and other questions required for the whole sample), n*on-core trend* question modules asked regularly to allow trends to be monitored, *non-core time-limited* questions asked over a set time period in response to current policy needs and *omnibus* questions asked on a short-term as-needed basis. A high

[1] Topic coverage could be managed through a cut in questions, a cut in the frequency of asking questions and/or modular approach with sub-sampling.

degree of flexibility would be introduced by allowing non-core questions to be administered to random part-samples and in only some survey years.

We suggest that either option 3c or option 3d could be profitably explored in an empirically based feasibility study. Further discussion about cuts in sample sizes and topic coverage would be required among all survey stakeholders. And more thought would be given to how an integrated survey might be packaged for tendering purposes.

1. Background and context

As the cost of undertaking high quality, random sample surveys of the general population continues to rise at a time of fiscal austerity, governments around the world are seeking innovative yet cost-effective and robust ways of obtaining timely information about population demographics, attitudes, and behaviours. Although much promise lies in the more efficient use of administrative, transactional, and other forms of 'big' data, these data sources have their own shortcomings such as limited range and depth of data, measurement error, non-representativeness of study population, and lack of transparency in methods. Thus, it is likely that sample survey research will remain core to understanding the views and characteristics of general populations.

It is therefore essential that governments ensure that their substantial financial outlay on survey research is implemented as cost-effectively as possible. Cost-efficiencies in survey implementation can be obtained in a number of ways, including reducing the total survey effort, lowering the marginal cost of interviews by changing to cheaper modes of administration, and integrating what are currently independent surveys into a single survey vehicle.

It is in this context that in May 2014, the Welsh Government (WG) appointed Ipsos-MORI, in collaboration with Professor Patrick Sturgis, to undertake a review of the way in which large scale surveys commissioned by WG, or WG funded bodies, are conducted. The review explores the extent to which the design and implementation of these surveys could be changed to reduce the overall cost of the Welsh survey effort, without compromising the quality of its outputs.

Some of the options considered for redesigning the survey portfolio also aim to obtain additional flexibility in the way that ad hoc survey needs are met and to yield enhanced precision for sub-group and small area estimates for key policy-relevant variables.

The overall objective of this project is to offer advice on options for redesigning the regular large-scale social surveys commissioned by the WG, or WG funded bodies, as well as a possible future housing condition survey. The review considers a range of possible re-design options, ranging from simple reductions in the number and length of interviews to full integration into a new continuous, modular survey vehicle.

The changes considered cover not only how samples and questionnaires might be efficiently integrated but also whether there may be benefits in altering the survey administration mode, in full or in part. The major component of the current survey budget is taken up by the cost of paying interviewers to visit addresses and conduct interviews in person. Shifting some of these 'face-to-face' interviews to telephone interviews or to modes which enable self-completion of questionnaires by respondents can, in principle at least, result in substantial cost savings.

The report presents a range of options for changing the way that WG and WG funded bodies implement their survey-based information needs and assess the advantages and disadvantages of each one according to criteria agreed with WG. The report will provide an evidential basis for WG and WG funded bodies to make an informed decision about how best to proceed with commissioning their survey portfolios from 2016 onward.

The report proceeds in the following manner. First we describe how the project team went about producing the options for potential survey redesigns before providing an overview of the in-scope surveys and their key design features. We then set out the criteria against which the options are to be evaluated. The next section reviews similar exercises which have been undertaken in comparator countries and draws lessons from what has and has not been successful regarding survey integration. The options for survey-redesign are then presented in sequential order, the advantages and disadvantages of each are described, and concrete examples of how the changes might be implemented are provided, with approximate estimates of the likely degree of cost-savings. It must be stressed that these cost saving estimates are indicative only and intended purely for comparison of the different options. Furthermore, the cost savings only relate to fieldwork costs because it is not possible to produce estimates of overall cost savings without detailed information on the cost structures for each survey.[2] Next we consider how ad hoc survey needs (the 'omnibus' option) might be met under each of the identified options and consider some additional methodological issues which arise from some of the potential changes. We finish with a summary evaluation of the options against the criteria and present our overall conclusions and recommendations.

[2] Detailed information on the cost structures of the surveys is deemed to be commercially sensitive and therefore not available to the authors of this report.

2. The approach to developing re-design options

The work to produce the redesign options commenced with an inception meeting at WG on 13 May, at which the project objectives, in-scope surveys, and evaluation criteria were clarified and agreed. This meeting contained survey managers and senior analysts from WG and WG funded bodies. It enabled the project team to obtain early insight into the background and rationale of the in-scope surveys, as well as the key drivers of information needs and the logistical and practical constraints around carrying out large-scale surveys in Wales.

Following the inception meeting the project proceeded via two parallel and interrelated strands. The first strand involved desk research, in which the project team reviewed key documents relating to the in-scope surveys in order to establish the core survey objectives, data uses and the technical details of survey implementation. This strand also comprised reviewing documents relating to similar exercises planned or carried out in other countries and conducting phone interviews with key players in these overseas initiatives.

Alongside the desk research, interviews were carried out (both in-person and via telephone) with survey stakeholders. Stakeholders were representatives of: (i) WG and other Welsh public body research staff responsible for managing the surveys; and (ii) a small number of WG policy teams. The objectives of these interviews were to: (i) fill in and contextualise knowledge gaps remaining after the desk research has been completed (ii) identify uses/objectives of the survey that are not manifest in published outputs and (iii) identify whether there are non-logistical/statistical barriers to change, relating to how each survey is controlled, presented or administered.

Following the completion of the desk research and stakeholder interviews, the project team undertook a series of brain-storming meetings in which ideas for redesigns were suggested, worked up, critically evaluated and either dropped or retained for further consideration. This resulted in the development of three broad approaches to redesign, each containing two or three more detailed variants. This total of 8 options were presented (via teleconference) to a second meeting of stakeholders on 15 July. Suggested refinements and amendments made at this meeting were incorporated into a revised set of options as set out in this report.

3. In-scope surveys

The primary motivation for commissioning this report is to reduce the total cost of delivering the survey-based information needs of the WG and WG funded bodies, while preserving data quality and maintaining or improving flexibility of analysis. In order to focus these efforts, a number of important candidate surveys were identified as in-scope, while others (such as the Annual Population Survey, surveys of businesses, and surveys of under 16s) were deemed out of scope.

The in-scope surveys are the WG funded National Survey for Wales and Welsh Health Survey, the Sport Wales funded Active Adults Survey, the Arts Council of Wales funded Arts in Wales Survey and the Natural Resources Wales funded Wales Outdoor Recreation Survey. Other in-scope surveys include some which are conducted periodically and are wholly or partly funded by WG, such as the Infant Feeding Survey, the National Diet & Nutrition Survey, the Dental survey, and the European Health Interview Survey. Additionally the review includes a possible future housing condition survey. Finally, there is interest in exploring whether it may be possible to develop, as part of this exercise, a flexible survey vehicle for collecting fast turnaround ad hoc needs in an omnibus-style manner.

It is important to note that the list of surveys designated as in-scope is not fixed. The core objective of this initiative is to reduce expenditure on the survey portfolio without negatively affecting the quality of the data collected. It may, therefore, be desirable to add new surveys to the list of those that are in-scope, as well as to remove some from the list, as this work progresses.

The National Survey for Wales (National Survey) is a continuous survey involving 14,500 interviews spread evenly across local authorities. A single adult aged 16 or over is randomly selected for interview where more than one adult is resident in the household. The 25-minute CAPI interview is focused on policy relevant topics, notably well-being and assessments of key public services.

The Welsh Health Survey (WHS) is a continuous survey which collects in-home data from 10,000 households per year (16,000 adults and 3,000 children), again distributed so as to allow local authority level estimates to be made. A short interview is conducted at each household and self-completion questionnaires are then left for completion by eligible adults and up to two selected children in each household. The questionnaire collects data on health status, lifestyle and health service use.

The Sport Wales Active Adults Survey (AAS) is collected on a biennial basis, collecting data for all months of the years in which it is fielded. In 2014 it will collect data from 8,000 individuals spread evenly across local authorities.[3] Interviews are undertaken with one randomly selected household member aged 15 or over. The 25-minute CAPI interview is mainly focused on participation levels in a wide range of sports and other physical activities.

[3] The target sample size for the Active Adult Survey was 22,000 per survey wave from 1998/99 to 2012. A much smaller sample size was agreed in 2014 due to budgetary constraints.

The Arts in Wales survey (AIWS) investigates levels of engagement with and attitudes to the arts among adults in Wales. The survey runs every 5 years with the last wave carried out in 2010 when over 7,000 adult interviews were undertaken using an in-home CAPI methodology. An even distribution of interviews was achieved in each local authority and fieldwork was split into three periods to capture seasonal variation. The sample design is not random: sampling points are randomly selected with probability proportional to size and then in a second stage, interviews are undertaken with an adult household member aged 16 or over. Quota controls are set on age and social class within sex.

The **Wales Outdoor Recreation Survey** (WORS) investigates use and barriers to use of outdoor spaces in Wales. It is carried out every three years with 15-20 minute telephone interviews being conducted with a single adult in each contacted household; the total achieved sample size is 6,400. The technical report states that the sample is selected using random digit dialling (RDD) methods although insufficient detail is provided to enable to allow assessment of whether a truly random probability selection method has been used.

A possible Welsh Housing Condition Survey (WHCS) would provide information about the condition of the Welsh housing stock.[4] Data could be collected on a periodic basis, for example every five years, and an achieved sample size of about 3,000 is likely to be needed.[5] Information about the condition of the property could be collected from the Household Reference Person or any adult aged 18 or over. In addition to the collection of socio-economic characteristics, the survey would include about five to ten minutes of questions on the respondent's perception of the property condition, sufficiency of dwelling space, suitability of accommodation to meet access requirements, help with housing maintenance, thermal comfort, fuel expenditure, and the installation of energy efficiency measures. This information would then be used to target properties for the second stage of the study - a physical inspection of the property by qualified surveyors.

Further detail on the sample designs and topic focus of the in-scope surveys can be found in Table 6.5 in section 6.

4. Criteria for assessing the options

The Invitation to Tender for this project specified a number of essential and desirable criteria against which redesign options should be evaluated. These initial criteria

[4] Similar surveys have previously been carried out in Wales: A series of repeated Welsh House Condition Surveys was carried out from 1968-1998. This survey series was then replaced by the Living in Wales survey, a repeated cross-sectional face-to-face survey which ran from 2004-2008. The Living in Wales survey was also referred to as the Welsh Household and Dwelling Survey.

[5] The achieved sample size for the 2008 LIW survey was 2,741 properties.

were refined and augmented following discussion at the project inception meeting on 13 May.

The essential criteria are:

- Reduce overall survey costs substantially;
- Meet user needs;
    - Estimates must be provided at required geographic levels;
    - The survey vehicle should be flexible enough to meet ad hoc needs as they arise;
- Be statistically sound;
    - Estimates should be sufficiently precise and unbiased to meet policy needs and the requirements of official statistics;
    - High response rates should be maintained;
    - Changes should not lead to discontinuities in essential time series;
    - Robust sampling methods should be used;
- Be practicable;
    - The practical demands of any proposed survey option should be compatible with the available Welsh survey infrastructure;
    - Surveys should be engaging and should avoid excessive respondent burden;
    - The level of risk to all aspects of survey delivery should be acceptable;
    - The delivery of other surveys in Wales should not be adversely affected by implementation of the proposed survey the impact on competition;
- Provide value for money;
    - The survey vehicle should be deliverable by competitive tender by a range of survey contractors and within the normal timescales for government contracts;
    - Survey samples should only be as large as necessary for the required precision;
    - Unnecessarily frequent collection of data on slow-changing topics should be avoided;
    - Wherever possible, the linking of survey results to administrative data should be supported.

It will also be desirable for any proposed changes to:

- Meet user needs;
    - Minimise the time taken between identification of a data need and data delivery;
    - Allow for flexible analyses that can meet the requirements of a range of data users;
    - Offer some omnibus survey capacity;
    - Enable quantification of impacts of methodological changes;
- Provide value for money;
    - Minimise time (and costs) of survey set-up phases;
    - Exploit relevant technological developments;

- Maximise the efficient use of overall respondent contact time.

At the project inception meeting the project team suggested the addition of two further criteria:

- market impact
- stakeholder acceptability

Market impact concerns the possible effect of each option on competition between survey agencies. For example, if no research agency has sufficient Welsh fieldwork capacity to deliver a large combined survey on its own, then delivery would in practice be restricted to consortia. Given the very limited number of agencies with both large field-forces in Wales and experience of delivering high quality random probability surveys, this might result in competition being either eliminated or severely curtailed.

Stakeholder acceptability needs to be considered explicitly rather than assuming it will be covered when survey options are assessed against stated survey needs. This is because there may be, for example, political/presentational reasons why stakeholders desire standalone surveys.

In order to make the assessment of the re-design options against this long list of criteria more manageable, the full set of essential and desirable criteria were mapped onto the following 6 broader headings:

- Cost
- Data Quality
- Flexibility
- Market Impact
- Stakeholder Acceptability
- Survey Applicability

A table setting out the performance of each of the options against these 6 criteria can be found in 0. We provide a narrative account of these evaluations in section 0.

5. Survey integration in other countries

Although there are a number of reasons why integration of the regular surveys in a country might be deemed desirable, they generally fall under one of three headings: cost reduction; quality improvement; and increased flexibility. Survey integration can reduce costs because it offers possibilities of resource sharing – particularly those relating to set-up and central management – and elimination of redundancy in questionnaire content. It can improve quality by enabling more statistically efficient sample designs to be used (e.g. moving from multi-stage to single stage sampling) and by allowing some questions to be asked of larger samples than would have been possible using discrete survey vehicles. It can increase flexibility, by using a

modular questionnaire structure such that different modules can be fielded amongst larger or smaller sub-samples and at longer and shorter intervals.

Given these possible advantages, it is unsurprising that possibilities for survey integration have been explored in a number of countries. Below we discuss four examples where the case for survey integration has been considered and/or implemented. These were the only cases that were identified as part of our desk research. However, because much of the reporting of this type of activity is either unpublished, or in the 'grey' literature, it is possible that additional examples exist.

5.1 UK: Integrated Household Survey

The Integrated Household Survey (IHS) was conceived as a composite survey based on constituent surveys run by the Office for National Statistics (ONS). The aim of running the IHS as stated in the survey documentation[6] is to "produce high level estimates for particular themes to a higher precision and lower geographic level than current Office for National Statistics social surveys".

As originally planned it would have been implemented as a fully integrated modular survey with an unclustered sample design. In practice, for budgetary and other reasons a more modest design was adopted in which a core set of around 100 questions were asked on a number of host surveys, each of which retained its original sample design.

The core questions cover the following broad topic areas:

- Economic activity;
- Education;
- Health;
- Smoking;
- Sexual identity;
- Subjective wellbeing.

Since IHS was set up in 2008-9 the number of IHS constituent surveys has decreased from six to two, the Annual Population Survey (APS) and the Living Cost and Food Survey (LCF) (see Table 5.1 below). Furthermore, 96% of the sample (327,000 interviews annually) derive from the APS with a mere four per cent deriving from the LCF (12,700 interviews).

Given this, the level of integration in the current IHS is rather modest, effectively amounting to fielding the same set of questions on two surveys, one of which is very considerably larger than the other. Although it technically delivers on its aim of delivering more precise estimates than single ONS surveys, the increases in

[6] Integrated Household Survey User Guide, volume 1: IHS background and methodology 2012. http://www.ons.gov.uk/ons/guide-method/method-quality/specific/social-and-welfare-methodology/integrated-household-survey/index.html

precision over those which would obtain if core questions were fielded on the APS alone are very small.

The reasons why the IHS did not ultimately live up to the original ambitions may be illuminating for any similar exercises embarked upon in Wales and elsewhere. One of the key barriers to successful integration in the IHS project proved to be the disproportionate size of the core question block relative to the topic-specific questions on each survey. What integration into the IHS meant for many existing UK surveys was a loss of questionnaire space for the policy areas to which these surveys were dedicated and by which they were (directly or indirectly) funded. The benefits, on the other hand, were less clear and not directly relevant to the survey sponsors, amounting essentially to increased precision for the variables included in the core which were often of marginal interest to stakeholders.

Table 5.1 History of the IHS

	2008-9	2009-10	2010-11	2011-2012	2012-
Additions and subtractions to the IHS core	*Jan 2008* + General Lifestyle Survey (GLF) + Living Cost and Food Survey (LCF) + Opinions Survey (OPN) *April 2008* + English Housing Survey (EHS)	*April 2009* + Labour Force Survey/ Annual Population Survey (LFS/APS) *June 2009* + Life Opportunities Survey (LOS)	*Jan 2010* – Opinions Survey (OPN)	*April 2011* – Life Opportunities Survey (LOS) – English Housing Survey (EHS)	*Jan 2012* – General Lifestyle Survey (GLF)
List of component surveys of the IHS	• General Lifestyle Survey (GLF) • Living Cost and Food Survey (LCF) • Opinions Survey (OPN) • English Housing Survey (EHS)	• General Lifestyle Survey (GLF) • Living Cost and Food Survey (LCF) • Opinions Survey (OPN) • English Housing Survey (EHS) • Labour Force Survey/ Annual Population Survey (LFS/APS) • Life Opportunities Survey (LOS)	• General Lifestyle Survey (GLF) • Living Cost and Food Survey (LCF) • English Housing Survey (EHS) • Labour Force Survey/ Annual Population Survey (LFS/APS) • Life Opportunities Survey (LOS)	• General Lifestyle Survey (GLF) • Living Cost and Food Survey (LCF) • Labour Force Survey/ Annual Population Survey (LFS/APS)	• Living Cost and Food Survey (LCF) • Labour Force Survey/ Annual Population Survey (LFS/APS)

5.2 Scotland: Long Term Strategy for Population Surveys

The Scottish Government developed a Long Term Strategy for Population Surveys in Scotland for the period 2009 - 2019 and tasked the Population Surveys Coordinating Committee (SPSCC) with its delivery[7]. The aims of the strategy are: (i) to ensure that the Scottish Government's population surveys meet key information needs whilst maximizing the quality and value of data generated and value for money; and (ii) to give full consideration to issues of survey participation, respondent burden, data quality, and data security, and to make survey practice recommendations for Government and other public bodies. As a result of the strategy, the three main Scottish Government cross-sectional surveys (the Scottish Health Survey, the Scottish Household Survey, and the Scottish Crime and Justice Survey) have been redesigned and the samples drawn in coordination with one another. Related work being undertaken involves, inter alia, developing centralized weighting approaches, dissemination and communication, and data linkage strategies.

As a direct result of the initiative two major changes to the conduct of Scottish surveys have been introduced. First, the three major Scottish surveys now use unclustered sample designs which are drawn centrally. This enables better control over the frequency with which individual addresses are selected for any of the surveys. Second, a common set of 20 question topics (some of which include more than one question) are fielded on the three surveys, covering[8]:

- Household Relationship Matrix (used to derive household type)
- Gender
- Age and Date of Birth
- Religion
- Country of Birth
- Ethnic Group
- Sexual Orientation
- Self-Perception of Health
- Tenure
- Economic Activity
- Banded Household Income
- Qualifications Held (Can be used to determine highest qualification held)
- Smoking
- Car Access
- Marital Status
- Disability or Long Term Health

[7] Scottish population surveys co-ordinating committee: Long term strategy for population surveys in Scotland 2009 – 2019. http://www.scotland.gov.uk/Topics/Statistics/About/SurvStrat

[8] The Scottish Government: Core and Harmonised Questions. http://www.scotland.gov.uk/Topics/Statistics/About/SurveyHarm/corealldownload

- Mental well-being
- Perception of local crime rate
- Perception of police performance
- Caring

As can be seen, a majority of these relate to demographic and other household classification variables and would therefore have been candidates for inclusion on the three surveys in any event (albeit not necessarily using harmonised wordings). Of greater relevance to this review are the seven topic areas covering self-perception of health, smoking, car access, mental well-being, perception of local crime rate, perception of local police performance, and caring which would not have been asked on all three surveys in the absence of the integration strategy. Therefore, as a result of the strategy (i) the precision of total and sub-group estimates based on these questions has been increased and (ii) estimates can be obtained for lower levels of geography.

In sum, the Scottish work on integration has been relatively modest in ambition but successful in meeting the objectives of improving sampling efficiency, controlling survey burden on households and increasing the precision of estimates for a core set of variables. The strategy has been successful in improving the quality of survey estimates in Scotland but has not served to reduce the cost of implementing the Scottish survey effort. However, cost reduction was not (so far as we have been able to determine) an objective of the strategy.

5.3 New Zealand: Survey integration at Statistics New Zealand and the Ministry of Health

Another country that has implemented a survey integration strategy is New Zealand. The main drivers towards survey integration in New Zealand were:

- minimising the development time and costs for new survey content; and
- maximising the information collected within the constraints of respondent burden and data collection costs.

Rather than opting for full integration of all household surveys using a single master sample, Statistics New Zealand opted for a hybrid solution with new integrated survey vehicles built on three pre-existing continuous surveys:

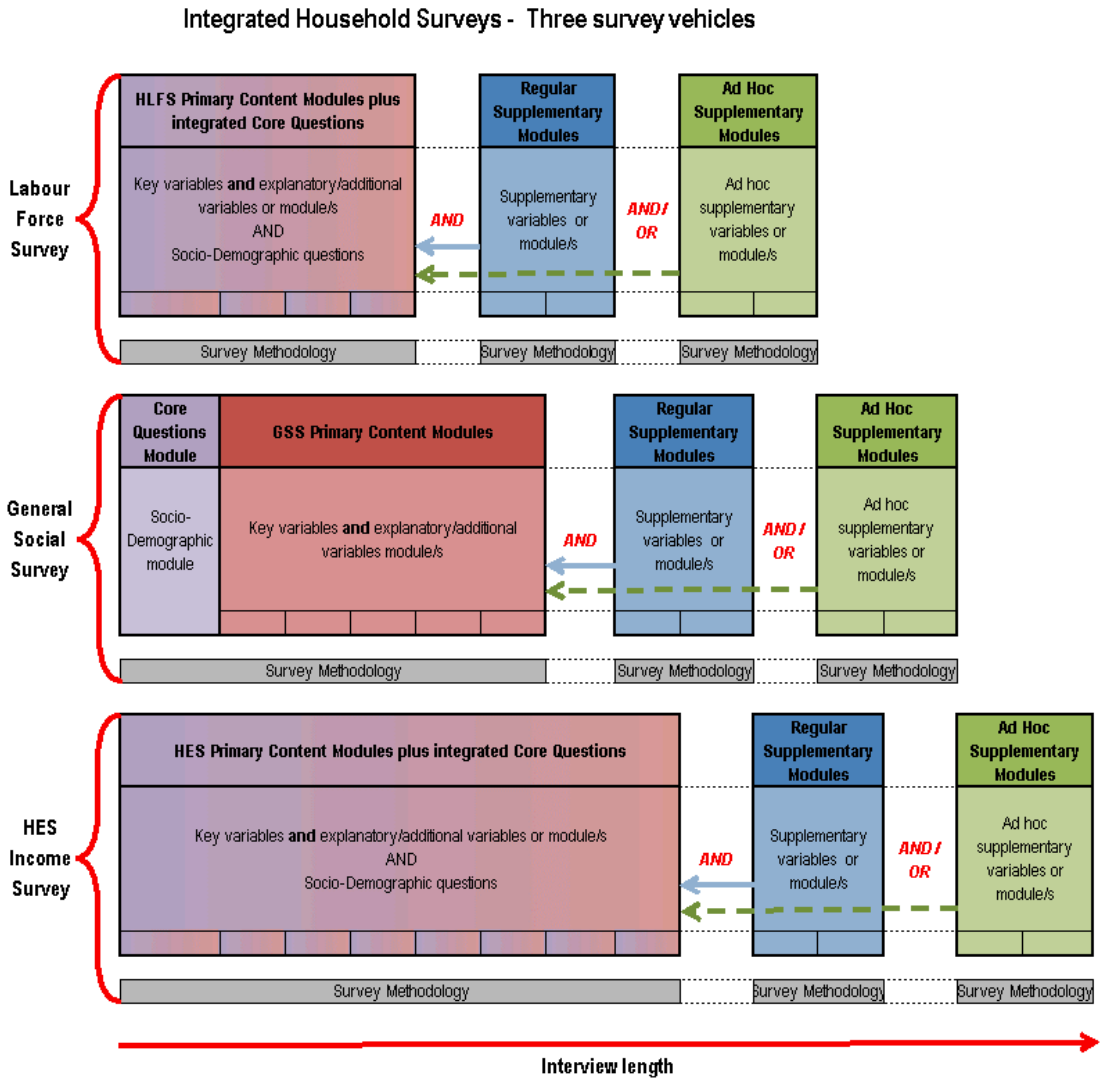
- the Household Labour Force Survey (HLFS),
- the Household Economic Survey (HES) and
- the General Social Survey (GSS)

The three new survey vehicles include the main questions from the original surveys but also include space for rotational modules and ad hoc topics which are organised around the specific theme of the main questions: labour market activity, expenditure

and income, and social attitudes. A core set of harmonised socio-demographic questions has also been developed and is included in all three survey vehicles.[9]

The average face-to-face interview length varies by vehicle: 12 minutes for the HLFS, 46 minutes for the GSS, and 2 hours for the full HES which is conducted every three years. This reduces to 45 minutes for the shorter HES (income) which is conducted in the two years between each full HES. Although New Zealand residents are obliged by law to cooperate under the Statistics Act 1975, response rates vary somewhat across the survey vehicles: about 86% for the HLFS, 78% for the GSS, and about 67% for the HES.

Figure 5.1 New Zealand household surveys integrated into 3 survey vehicles[10]



[9] Age, sex, geographic area, ethnicity, Maori descent, country of birth, years since arrival in New Zealand, social marital status, family type, dependent children, household composition, highest qualification and year of qualification, labour force status, income (personal and household), sources of income and household tenure.

[10] Bycroft, C (2010). Integrated household surveys: A survey vehicles approach. Wellington: Statistics New Zealand.

One of the benefits of this hybrid approach, building on pre-existing surveys, is that it makes for an easier transition compared to setting up a completely new infrastructure. New survey content is included within the new vehicles and the introduction of new stand-alone surveys has become the exception. A new standalone survey can only be justified if target populations are very different from those covered by the integrated vehicles, or if essential information needs cannot be met within the existing vehicles. Although this new system is still in its early days operationally, achieving a satisfactory balance between pre-planned rotating and allowing sufficient space for ad hoc information needs has proven to be challenging.

At present each survey vehicle has a separate sample design but Statistics New Zealand is exploring the potential benefits of switching to a master sample approach.

Alongside the survey integration efforts at Statistics New Zealand, the Ministry of Health in New Zealand has also developed a survey vehicle which carries a set of core health questions and space for topic modules that change every six or twelve months: The New Zealand Health Survey (NZHS). This continuous survey replaces the previous NZHS which was repeated every four to five years as well as a number of stand-alone surveys on various health topics.

The core health questions on NZHS cover population health, health risk and protective factors, and health service utilisation. The survey also includes measurement of height and weight. To minimise survey burden on respondents, the NZHS selects only one eligible adult and one eligible child per dwelling. Even though participation is voluntary, the weighted response rate is high at 80% for adults and 85% for children.[11]

5.4 Statistics Canada: New household survey strategy

Since its inception in 1945 the Canadian Labour Force Survey (LFS) has been used as a general purpose vehicle for Canadian household surveys, using either active or rotated-out LFS sample dwellings, or using fresh dwellings selected in LFS sample clusters. Growing demand for surveys, however, has led to increasing respondent burden using this approach, resulting in lower response rates.

Furthermore, it has become increasingly difficult to deliver surveys using the LFS infrastructure in a timely and cost-effective way. For this reason Statistics Canada undertook a review which explored different approaches to conducting household surveys featuring increased integration, common core content, a master sample and possible use of the Canadian Address Register as a sampling frame.

[11] Ministry of Health (2013). New Zealand Health Survey: Methodology Report 2012/13. Wellington: Ministry of Health.

Several alternatives were explored all of which involved the creation of a master sample and most of which were variations of the following two designs.[12, 13]

1) Pure two-phase design:
   - Phase 1 – A very large master sample of dwellings (e.g. 300,000) is selected which receive a basic short questionnaire covering only core content such as demographics, education, income and immigration;
   - Phase 2: Basic information from Phase 1 is used to select sub-samples for a variety of surveys including the LFS and surveys of targeted populations;

This alternative was ultimately rejected on the grounds of cost because even if the first phase is conducted by post, follow-ups with non-respondents by phone or face-to-face would still be needed to achieve an acceptable response rate.

2) The master sample as a union of surveys:
   - Major surveys such as the LFS and the Canadian Community Health Survey continue to be conducted on their own but include a common core set of questions (e.g. demographics, education, income and immigration);
   - The samples of these major surveys are combined to form a master sample which can be used to select sub-samples for other surveys;

This alternative was also rejected following the results of a pilot test which were disappointing, with lower response rates than anticipated and increased analytical complexity (e.g. weighting, efficient allocation of sample among second phase surveys) with little benefit (e.g. using first phase data for non-response adjustment and first phase estimates for calibration produced estimates that were not significantly different from their simple counterpart).

Statistics Canada is currently in the process of developing a “Household Survey Frame Service” (HSFS) which involves the integration of various frames used by social surveys under a single umbrella. It will include the Address Register, Clusters (the area frame), and telephone number databases including both landlines and some mobile phone providers. The overall objective of setting up the HSFS is to improve sample coordination and ultimately to consider simpler sample designs for major surveys such as the LFS.

[12] Gambino, J., Tambay, J-L and Laflamme, G. (2007). Statistics Canada’s New Household Survey Strategy. Paper presented at the SSC Annual Meeting, June 2007, Proceedings of the Survey Methods Section.

[13] Tambay, J-L., Laflamme, G. and Gambino, J. (2009). The Canadian Experience in Creating a Master Sample. Paper presented at the 57th session of the International Statistical Institute, 16-22 August 2009.

5.5 International experiences: lessons learned

The experiences of survey integration in these four countries are informative for the development and implementation of similar initiatives in Wales and elsewhere.

First, we can see that these exercises have been driven primarily by a desire to increase operational and statistical efficiency rather than to reduce costs. This is particularly true of the UK, Scottish and New Zealand initiatives, less so for the Canadian efforts which have been aimed at achieving both cost and statistical efficiencies.

Second, it is clear that successful implementation is not straightforward. The experience of the IHS in the UK is particularly illuminating in this regard; the failure of the IHS to deliver the desired level of integration in the long-term appears to have foundered on the lack of clear benefits of joining the IHS to the individual survey vehicles and their funders. By way of contrast, the Scottish initiative has been successful in its integration efforts, possibly as a result of the different policy context and the manner in which surveys are funded and delivered in Scotland compared to the UK as a whole.

Finally, it also seems clear that a substantial barrier to successful implementation of efforts to integrate surveys relates to the size and content of 'the core' question block. If the core is set too large then incentives to integrate may be too low because surveys and the policy-makers who drive them have insufficient space to field the questions of greatest interest to them. While, on the other hand, if the core is set too small then the benefits of integration are likely to be of insufficient magnitude to justify.

6. The options

In this section we describe the nine options that have been identified in our review as potentially meeting some or all of the evaluation criteria. For each option, we describe the general approach before setting out a specific example of how the redesign might actually be implemented in practice. We provide estimates of approximate cost savings for each option and enumerate the advantages and disadvantages against the remaining criteria.

It is important to note that these nine options are not intended to represent an exhaustive list of possible survey implementations. Each option presented could certainly be added to or amended to create a new sub-option. For example option 2 (reducing sample size) could be combined with option 1 (changing to a cheaper mode of data collection), to create a new option that is not currently listed here. Our intention in setting the options out in this way is to strike an appropriate balance between, on the one hand, describing options in very general terms and, on the other, providing a very long list of detailed design options, many of which would differ from one another by only small degrees.

Option 1 Introduction of cheaper modes of contact and data collection

All but one of the in-scope surveys[14] for this review use a combination of face-to-face interviews and self-completion questionnaires that are dropped-off and collected by interviewers. The involvement of face-to-face interviewers has a positive effect on sample representativeness and data quality. First of all, response rates for face-to-face surveys tend to be considerably higher than equivalent telephone, postal and online surveys, thus reducing the risk of non-response bias. Furthermore, face-to-face interviewers appear to be more effective at recruiting 'hard-to-get' population sub-groups compared to other modes, thus improving the match between the achieved sample and the target population. The interviewer also plays an important role in motivating and helping respondents to provide full and accurate answers. Longer and more complex questionnaires can be administered in face-to-face interviews than in telephone interviews and self-completion questionnaires.

However, the cost of face-to-face surveys is much higher than alternative modes, mainly due to the cost of interviewers travelling to addresses. Considerable savings could therefore be made by eliminating the need for interviewer visits to households by moving some or the entire existing survey portfolio from face-to-face contact and data collection to telephone, postal or web modes.

[14] The Welsh Outdoor Recreation Survey uses telephone contact and telephone interviews. The Welsh Health Survey uses face-to-face contact and a short face-to-face household interview, but all individual questionnaires are completed by the respondents using a paper and pencil questionnaire. All other in-scope surveys use face-to-face contact and face-to-face data collection.

Option 1a: Postal contact and self-completion questionnaires

The most complete coverage of the Welsh population can be achieved by sampling from the Postcode Address File (PAF), as is currently done on the National Survey, WHS and AAS. However, PAF is a list of postal addresses and does not include any further information that can be used to contact people by phone or email.[15] Consequently, when using PAF, the only mode of first contact other than a face-to-face visit is by post. Letters have to be addressed to "the resident" or "the householder" and include instructions for random respondent selection[16]. Following postal contact, data can be collected using a postal questionnaire, a web questionnaire or a combination of both.

Recent experiments carried out in the UK suggest that web data collection on its own is unlikely to achieve a response rate much above 25%, and only then with pecuniary incentives.[17, 18] About twelve percentage points can be added to the response rate by offering a postal questionnaire with the second postal reminder[17] but this will still be considerably lower than the response rate that can be achieved when using face-to-face contact[19]. A similar or possibly higher response rate can be achieved using postal questionnaires only.

A recent study carried out in the UK[17] found that, compared to an equivalent face-to-face survey, a web sample contacted by post was more likely to comprise high earners, people with higher education qualifications, native English speakers, and middle-aged people living with a partner and owning their own home. Not surprisingly, the web sample also included more regular internet users. Following-up web non-respondents with the offer of completing the questionnaire by post made

[15] Matching names and telephone numbers to sampled PAF addresses is not an option. Names would have to be found in the edited Electoral Register (ER) before looking up telephone numbers that are listed in Telephone Directories or other commercially available lists. Both the edited ER and telephone lists suffer from considerable under-coverage which will result in low matching rates (about 25%).

[16] The most commonly used method for respondent selection in postal surveys is to identify the eligible resident who celebrated his/her birthday most recently (Last Birthday) or who will be celebrating their birthday next (Next Birthday). In theory, it would be possible to instruct all eligible household members to complete a web questionnaire. This is, however, not possible when using postal questionnaires because it is unknown how many questionnaires need to be sent to the sampled address.

[17] TNS BMRB. (2013). Community Life Survey: Summary of web experiment findings. Last accessed on 14/08/2014 at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/325872/Annex_B_-_Summary_of_web_experiment_findings_2012-13.pdf

[18] Villar, A. (2013). Feasibility of Using Web to Survey at a Sample of Addresses: a UK ESS experiment. Last accessed on 11/07/2014 at http://www.natcenweb.co.uk/genpopweb/outputs.htm

[19] When using a £5 conditional incentive, the response rates for the Community Life Experiment (TNS BMRB, 2013)[17] were 19% for web only, 31% for web and postal, and 60% for face-to-face only.

the age profile worse, with more middle-aged and older people taking part but not more young people.

Postal and online modes also carry the risk of self-selection bias with research showing that approximately one in four questionnaires is completed by someone who is not the randomly selected respondent.[17] There will be a tendency for those who are particularly interested in the survey topic to self-select which is likely to result in biased data; e.g. a household member who is very sporty will be tempted to complete a postal version of the AAS instead of the randomly selected household member who does no sport at all.

Using postal contact rather than face-to-face visits has an additional complication for the WHS which currently includes all adults in the household and up to two children. The number of adults and children at a sampled PAF address is unknown and it is therefore not possible to send the appropriate number of questionnaires with the initial request for survey participation. The options are (1) send one household questionnaire, four adult questionnaires and two child questionnaires to all sampled addresses; (2) simplify the WHS questionnaires considerably and use a single questionnaire with a separate column for up to 6 household members[20]; (3) select only one adult at random in the household; (4) send one household questionnaire and one adult questionnaire to all sampled addresses to be completed by any adult in the household and follow-up other adults and up to two children by post; (5) combinations of options 1, 2 and 3 with 4. None of these options is without disadvantages, such as an increase in perceived burden and therefore a reduction in response rates, longer fieldwork periods for data collection, or loss of rich household level data.

Substantial cost savings could be achieved if the National Survey, WHS and AAS used postal contact methods rather than face-to-face interviewers, ranging from about 70% to 90% of the fieldwork costs. The cost savings would be lower for AWS which uses quota sampling methods and more modest for surveys that use telephone methods such as the Welsh Outdoor Recreation Survey (WORS; about 50% to 80% of fieldwork costs). There will be additional costs for data entry and data processing but these are marginal relative to the fieldwork cost savings.

[20] This is similar to the UK Census form approach.

Table 6.1 Example of using postal contact & self-completion survey design for the National Survey

Sampling frame	Postcode Address File
Sampling method	Unclustered with disproportionate stratification by LA
Within household selection	One adult aged 16+, Last/Next Birthday selection method
Contact strategy	1. Letter with URL address & password for web questionnaire 2. Reminder letter with URL address, password & postal questionnaire 3. Second reminder letter with URL address & password
Incentive strategy	e.g. £10 conditional on survey participation
Questionnaire length	Questionnaire may have to be shortened and simplified for the postal version
Issued sample size	46,000 addresses
Response rate	35% (compared to current response rate of about 70%)
Achieved sample size	14,000 adults aged 16+ (at least 600 per LA)
Estimated fieldwork cost	About 25% of current fieldwork cost

Relying solely on postal contact means that the data collection mode would have to change for the National Survey, AAS, AWS and WHS (household level) from face-to-face interviewing to self-completion questionnaires (postal and/or web). It should be noted that self-completion questionnaires are less accessible to those with visual disabilities and those with reading difficulties (including those who do not read English). Accessibility can be improved by using clear print[21], allowing respondents to adjust text and colour settings for web questionnaires, and adding audio to web questionnaires. Paper and online questionnaires can also be provided in Welsh and other languages.

Respondents may answer questions differently in self-completion modes compared to face-to-face interviews. For example, there is a tendency for respondents to give more socially desirable answers in face-to-face interviews compared to postal surveys, but this tends to be minimised by adding short self-completion questionnaires to the face-to-face interview for sensitive questions. More problematic are the potentially higher levels of missing data, non-differentiation[22], and middle category responses when an interviewer is not available to encourage the respondent to exert the cognitive effort required to provide full and accurate

[21] RNIB Cymru (2012). Make it Accessible: A guide to making information accessible for communities in Wales. Last accessed on 13/08/2014 at http://www.wlga.gov.uk/older-people-and-ageing1/make-it-accessible-a-guide-to-making-information-accessible-for-communities-in-wales

[22] Non-differentiation, also referred to as 'straight-lining', occurs when respondents select the same response category across several consecutive questions. Non-differentiation is most likely to occur with a battery of questions using rating scales.

answers[23]. Self-completion respondents are also more likely to make mistakes with complex or difficult tasks, such as complex routing in paper questionnaires and difficult questions required for coding industry, occupation and socio-economic classifications. And finally, the risk of measurement differences is increased when questions that have been optimally designed for face-to-face administration have to be adapted for self-completion questionnaires (e.g. detailed interviewer instructions that have to be simplified for self-completion questionnaires). The risk of all of these measurement differences can be minimised through good questionnaire design but it is unlikely to be completely removed. Consequently a change from face-to-face interviewing to self-completion is likely to result in discontinuities in time series.

Further information about mode effects can be found in a report prepared for the Welsh Government on mixing modes for the National Survey for Wales.[24]

Option 1b: Face-to-face contact and self-completion questionnaires

"Drop-off, pick-up"

An alternative strategy for the National Survey, AAS, AWS and a possible WHCS would be to maintain face-to-face contact but to collect the survey data using cheaper modes. This strategy has been used successfully for the WHS with short household interviews being conducted face-to-face and paper self-completion questionnaires left for all adults and up to two children in the household, which the interviewers collect at a later date. The amount of time interviewers spend at households (particularly large households) is thereby reduced, fewer return visits are needed, household members can complete the questionnaires at their own convenience, and interviewer collection of completed questionnaires minimises the decline in response compared to face-to-face interviews. Face-to-face contact rather than postal contact would also be beneficial to a possible WHCS because interviewers would be able to record relevant observations about the property for the entire sample, including non-responding households.

Using the WHS approach of delivering and collecting paper self-completion questionnaires, however, will not reduce costs for the National Survey, AAS and AWS using their current design of selecting only one adult in the household for a relatively short interview of about 25 minutes. Consequently the cost of travel to re-contact respondents and collect the completed questionnaire plus the additional cost of data entry and data processing will be greater than the cost of conducting the

[23] The shortcutting of cognitive effort to answer a question is known as 'survey satisficing'.

[24] For further information, see Nicolaas, G. (2012). Mixing Modes within a Social Survey: Opportunities and constraints for the National Survey for Wales. A NatCen Social Research report prepared for the Welsh Government. Last accessed on 19/06/2014 at http://wales.gov.uk/docs/caecd/research/120706nswMeth3Mixedmodesen.pdf

interview face-to-face. Furthermore, the response rate is likely to be lower. Although we don't have a proper empirical basis for making a comparison, we would envisage a decline of about 10% in response when using this methodology compared to face-to-face interviews. This decline in response rate could to some extent be counterbalanced by using incentives but this will further increase costs. Self-completion questionnaires are also less accessible to those with limited functional literacy. And finally, as discussed above under Option 1a, there is some risk of introducing measurement differences when switching from face-to-face interviewing to self-completion questionnaires. While the risk of measurement differences can be reduced through good questionnaire design, there is no guarantee that this can be avoided.

Table 6.2 Example of face-to-face contact and self-completion survey design for the AAS

Sampling frame	Postcode Address File
Sampling method	Unclustered with disproportionate stratification by LA
Within household selection	One adult aged 15+, Kish method
Contact strategy	1. Face-to-face contact at address 2. Short doorstep questionnaire & drop-off paper self-completion 3. Interviewer collects self-completion questionnaire
Incentive strategy	A conditional incentive could be used to minimise the decline in response rate due to self-completion rather than face-to-face administration but this will increase costs.
Questionnaire length	Questionnaire may have to be shortened and simplified for self-completion version
Issued sample size	20,000 addresses
Response rate	45%-50% (compared to current response rate of 52%)
Achieved sample size	8,000 adults aged 15+ (at least 360 per LA)
Estimated fieldwork cost	About 105% of current fieldwork cost

"Drop-off, postal return"

Rather than getting interviewers to re-visit addresses, costs can be reduced by asking respondents to return their completed questionnaires by post. Response rates will be lower which can be minimised to some extent by using incentives and telephone reminders (having collected telephone numbers at the initial face-to-face contact). Nonetheless, we envisage a significant decline in response which will be unacceptable to most if not all stakeholders. Using AAS as an example, we estimate a response rate of about 45% when using the "drop-off, pick-up" methodology and

about 25% when using the “drop-off, postal return” methodology compared to a response rate of 52% with face-to-face interviews.

Offering an online option

Alternatively, households could be persuaded to complete the questionnaire online which would not only save on data entry costs but also the cost of interviewers returning to the households to collect completed paper questionnaires. In the Welsh context, shifting from paper to online completion could yield even greater returns due to the requirement for postal questionnaires to be printed in both English and Welsh. In addition to the National Survey, AAS, AWS and a possible WHCS, this online strategy could also be applied to WHS which currently uses paper questionnaires that are completed by respondents at their own convenience and collected by interviewers at a later date.

However, not everyone will complete questionnaires online. About one in four households in Wales does not have access to the internet[25] and not everyone living in households with internet access will be able or willing to complete a web questionnaire. Furthermore, not everyone who initially agrees to complete the web questionnaires will actually do so. Consequently, paper questionnaires will be needed to boost population coverage and response rates, and interviewers will still be required to make multiple visits to the area in order to collect paper questionnaires as well as following up web non-respondents.

Mixing web data collection and paper self-completion with interviewer collection of completed paper questionnaires will add complexity to sample management, fieldwork operations, interactions with potential respondents and data management. Web completions would have to be monitored frequently in order to identify and assign web non-respondents to face-to-face interviewers for follow-up and conversion. Time lags could result in wasted visits when web questionnaires are completed after these cases have been assigned for follow-up. Interactions between interviewers and web non-respondents would be complicated with the interviewer attempting to convert web non-respondents to either go online or complete a paper questionnaire. Interviewers have more control over paper self-completions than web completion but this would necessitate a further visit (boosting response but increasing cost) whereas sample members may prefer web completion, particularly if they had previously been offered a conditional incentive for web completion.

Evidence from various experiments with mixed mode designs suggest that mixing online questionnaires with paper self-completions will reduce the response rate

[25] Welsh Government (2014). National Survey for Wales: Headline results, April 2013 – March 2014. Last accessed on 12/08/2014 at http://wales.gov.uk/docs/statistics/2014/140530-national-survey-wales-2013-14-headline-results-revised1-en.pdf

compared to only using paper self-completions.[26] Possible explanations for this decline in response rate include: (1) the additional effort required of the respondent to go online deters final response, (2) respondents being unable to complete the web questionnaire due to technical difficulties, (3) the paradox of choice[27], and (4) operational complexities (as discussed above). Although it may be possible to boost online completion by offering incentives and using telephone reminders, we would still envisage a significant decline in response for the National Survey, AAS and AWS when using this mixed mode design compared to the single mode design. Using AAS as an example, we estimate a response rate of about 40% when using the "drop-off, pick-up" *mixed mode* design (paper and online) compared to 45% for the "drop-off, pick-up" *single mode* design (paper only).

The impact of offering a web option on response rates and costs is more complex for WHS which attempts to collect self-completion questionnaires from all adults in the household and up to two children. Consequently, significant cost savings can only be achieved when a worthwhile proportion of *whole* households participate by web thus avoiding the need to send an interviewer to the household to collect paper questionnaires. There is no relevant survey experience to draw on but the 5th wave of the Innovation Panel of Understanding Society only managed to achieve a whole household web response rate of 25% among an established panel, using unconditional incentives as well as bonuses conditional on all eligible household members completing their questionnaires online.[28] The whole household web response rate is likely to be much lower for a fresh PAF sample. Therefore, interviewers will still be required to make multiple visits to the area in order to collect paper questionnaires as well as following up web non-respondents.

Allowing WHS respondents to complete the questionnaire online will not result in any real cost saving. The reduction in travel costs will be negligible because most travel is between the interviewer's home and the area where he/she works and not between the addresses within an area. Interviewer fees for productive cases on

[26] Several studies have found that mixing online questionnaires with traditional modes of data collection (either concurrently or sequentially) will decrease response rates compared to using the traditional mode on its own. See for example:

- Medway, R. and Fulton, J. (2012). When more gets you less: A meta-analysis of the effect of concurrent web options on mail survey response rates. Public Opinion Quarterly, 76, 733-746.
- Messer, B. and Dillman, D. (2011). Surveying the general public over the Internet using address-based sampling and mail contact procedures, Public Opinion Quarterly, 75, 429-457.
- Jäckle, A., Lynn, P. & Burton, J. (2013) Going online with a face-to-face household panel: initial results from an experiment on the Understanding Society Innovation Panel. Understanding Society Working Paper 2013-03, Institute for Social and Economic Research, University of Essex.

[27] Being given a choice entails an additional cognitive burden that functions to deter response. See: Schwartz, Barry. 2004. The Paradox of Choice: Why More Is Less. New York: Harper Perennial.

[28] Cullinane, C. & Nicolaas, G. (2013) Weekday of Advance Mailing and Conditional Incentives Experiment. Last accessed on 13/08/2014 at https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2013-06.pdf

WHS are relatively low so little saving can be made for not having to pay these fees for online questionnaires. And any saving in travel costs and interviewer fees will be offset by the additional costs of programming, sample management, helpline for web respondents, data management, etc.[29]

Table 6.3 Face-to-face contact & self-completion survey design with web option for WHS

Sampling frame	Postcode Address File
Sampling method	Unclustered with disproportionate stratification by LA
Within household selection	All adults aged 16+ and up to two children aged 0-15, Kish method
Contact strategy	1. Face-to-face contact at address 2. Short household questionnaire & online completion or paper questionnaires for all eligible household members 3. Interviewers collect completed paper questionnaires from those who did not agree to complete web questionnaires 4. Interviewers follow up web non-respondents and leave paper questionnaires for all eligible household members 5. Interviewers collect completed paper questionnaires for web non-respondents
Incentive strategy	e.g. conditional incentive for online completion
Questionnaire length	Unchanged
Issued sample size	62,500 addresses
Response rate	Whole household web response rate will be very low Final household response rate=74% (compared to current household response rate of about 78%)
Achieved sample size	14,000 adults aged 16+ (at least 600 per LA),
Estimated fieldwork cost	Similar or higher than current fieldwork costs

Option 1c: Telephone contact and telephone interviews

As discussed above, the options for using cheaper data collection modes without sacrificing quality is limited when using PAF as a sampling frame. Another possible method for selecting a true random probability sample of the Welsh population is

[29] There is very limited published evidence on the impact of using web in mixed mode designs for cutting survey costs. For a brief overview of the cost of introducing web in a face-to-face panel survey, see Toomse-Smith, M. (2013). Does mixing face-to-face and web lead to cost savings? A presentation given at the 5th Conference of the European Survey Research Association, Ljubljana, Slovenia, 15-19 July 2013. Last accessed on 14/08/2014 at http://www.europeansurveyresearch.org/conference/programme

Random Digit Dialling (RDD) using the Ofcom database of allocated blocks of 10,000 landline numbers (e.g. 029 2080 XXXX)[30] as seeds for number generation.[31] The random generation of numbers from the Ofcom database ensures that the sample will include ex-directory numbers as well as listed numbers, and all telephone numbers have an equal probability of being selected. However, RDD samples will also include very high proportions of non-working and non-residential numbers.[32] The cost of handling such high proportions of ineligible numbers can be reduced by automated screening of the sample for non-working numbers and removing numbers that are listed as businesses in advance of telephone fieldwork.[33]

This RDD method is rarely used in the UK. Instead "RDD" in the UK tends to refer to the random generation of the last N digits of telephone numbers sampled from other sources such as telephone directories and commercial databases. It is claimed that this method is more cost-efficient with negligible loss in coverage of households with landlines. However, it is not a true random probability sample because it is unknown whether all telephone numbers have a probability of selection. The extent to which this will result in coverage bias in the survey estimates is unknown.

Typically "RDD" surveys in the UK, including WORS, tend to include only landline phone numbers. However, the proportion of mobile-only households is increasing and is particularly high for Wales; 22% of Welsh households are mobile-only compared to the UK average of 16%.[34] Furthermore, mobile-only households are more likely to belong to lower socio-economic groups and to contain younger adults.[35] Consequently the exclusion of mobile-only households could bias survey estimates.

It is possible, but both complex and expensive, to include mobile phone numbers in RDD surveys using a dual-frame approach. Working blocks of '07' numbers can be selected from the Ofcom database and the last five digits randomly generated to

[30] It is also possible to include allocated blocks of 100,000 mobile numbers (e.g. 07769 1XX XXX).

[31] For more information on using the Ofcom database for RDD in the UK, see Nicolaas, G. & Lynn, P. (2002). Random-digit dialling in the UK: viability revisited. Journal of the Royal Statistical Society Series A, 165(2): 297-316.

[32] The proportion of non-working numbers in a RDD sample will vary depending on the specifics of the RDD design (e.g. source of area codes and exchange numbers) and the allocation and availability of Welsh telephone numbers (e.g. level of saturation of telephone number blocks allocated to Wales).

[33] Various companies that provide samples of telephone numbers will screen for non-working and business numbers but there is a lack of transparency in how this is done and its impact on the quality of the sample (e.g. false positives).

[34] Ofcom (2014), The Communications Market 2014: Wales. Last accessed on 07/08/2014 at http://stakeholders.ofcom.org.uk/market-data-research/market-data/communications-market-reports/cmr14/wales/

[35] Ofcom (2014). The Communications Market 2014: UK. Last accessed on 14/07/2014 at http://stakeholders.ofcom.org.uk/market-data-research/market-data/communications-market-reports/cmr14/uk/. Results are not available for Wales but it is assumed that a similar pattern would apply.

produce a separate sample of mobile numbers to be included alongside the landline RDD sample. An RDD sample of mobile numbers for Welsh surveys would then require extensive and costly screening because it would be UK-wide (the prefix '07' is used for all UK mobile numbers and allocation of blocks is not based on geography).

The two RDD samples will overlap in their coverage of households that have both landlines and mobile phones. There are two main approaches for dealing with this overlap in coverage. The first approach screens the mobile RDD sample for individuals in mobile-only households. The second approach attempts to interview all individuals contacted by mobile phone, regardless of their access to a landline, and then to deal with the over representation of individuals with access to both type of phone in weighting.[36]

Once telephone contact has been established with an eligible household, the interviewer would have to identify and select household members for interview. Selection is relatively straightforward for the WHS which currently includes all adults in the household and up to two children but making contact and interviewing all eligible household members by phone will be challenging. For the other in-scope surveys which involve random selection of one adult in the household, the Rizzo method [37] would produce a true probability sample of eligible persons in the household that is simple and minimally intrusive.

In addition to the usual weighting for non-response and unequal selection probabilities of individuals within households, RDD surveys may also require additional weighting for non-coverage of mobile-only households and higher selection probabilities of households with more than one telephone number. The latter requires additional questions to be asked about telephone ownership and usage. Weighting for RDD samples, particularly those that include mobile phone numbers, is very complicated.[38]

---

[36] For further information on dual-frame RDD designs, see:

- Brick, J.M., Brick, P., Dipko, S., Presser, S., Tucker, C., and Yuan, Y. (2007). Cell phone survey feasibility in the U.S.: Sampling and calling cell numbers versus landline numbers. Public Opinion Quarterly, 71, 23-39.
- Benford, R., T. Tompson, C. Fleury, G. Feinberg, B. Feinberg, N. Speulda and A. Weber. 2009. Cell phone and landline – considerations for sample design, estimates, weighting, and costs. Paper presented at the 64th annual conference of the American Association for Public Opinion Research; Hollywood FL; May 14.
- Boyle, J.M., Lewis, F. & Tefft, B. (2010). Segmented or Overlapping Dual Frame Samples in Telephone Surveys, Survey Practice, 3(6), last accessed on 14/08/2014 at http://www.surveypractice.org/index.php/SurveyPractice/article/view/149/html

[37] Rizzo, L. Brick, JM and Park, I (2004) "A minimally intrusive method of for sampling persons in random digit dial surveys" in *Public Opinion Quarterly* 68:2

[38] For further information on RDD weighting, see:

- Battaglia, M., Eisenhower, D., Immerwahr, S., Konty, K. 2010. "Dual-Frame Weighting of RDD+Cell Phone Interviews at the Local Level." Paper presented at 65th annual conference of the American Association for Public Opinion Research, Chicago.

Table 6.4 Telephone contact & telephone interviews for the National Survey

Sampling method	Random Digit Dialling, inclusion of mobile numbers
Within household selection	One adult aged 16+, Rizzo selection method
Contact strategy	Telephone, multiple calls on different days, at different times of the day, throughout the fieldwork period
Questionnaire length	Unchanged
Issued sample size	40,000 addresses
Response rate	20%-30%
Achieved sample size	14,000 adults aged 16+ (at least 600 per LA)
Estimated fieldwork cost	About 50% of current fieldwork cost

In addition to the complicated sampling and weighting, RDD surveys also suffer from low response rates and measurement issues which are exacerbated when including mobile-only households.

Response rates to RDD surveys tend to be low. For example, in 2011/2012 the "RDD" (landline only) response rate for the Active People Survey commissioned by Sport England achieved a response rate of about 28% which was considerably lower than the 53% response rate for the parallel face-to-face survey[39]. The "RDD" (landline only) response rate for WORS is similarly low at 23%. The inclusion of mobile phone numbers is likely to reduce response rates even further because it is more difficult to reach people on their mobile phones due to screening and the inconvenience of taking calls and being interviewed when not at home (e.g. on the go, in public spaces, at work). Such low response rates carry the risk of increasing non-response bias and damaging confidence in statistical quality amongst stakeholders, compared to the current face-to-face methodology.

As well as lowering the response rate, the improved coverage of including mobile phones in RDD surveys is associated with additional costs to the survey. Results

- Brick, J.M., Edwards, W.S., Cervantes, I.F., and Lee, S. 2008. "Sampling and Weighting Cell Phone Surveys to Supplement RDD Surveys." Paper presented at the 63rd annual conference of the American Association for Public Opinion Research, New Orleans.
- ZuWallack, R. and Conrey, F. 2010. "Weighting Landline and Cell Phone Dual-Frame Survey Samples." Paper presented at the 65th annual conference of the American Association for Public Opinion Research, Chicago.

[39] TNS BMRB (2013). Active People Survey 5-7 Technical Report. Updated to include APS6 (July 2013). Sport England, JN 260105344, Controlled document - Issue 6. Last accessed on 21/07/2014 at http://www.sportengland.org/media/312344/Active-People-Survey-6_Technical-Report_final.pdf

from a recent study carried out by TNS BMRB[40] suggest that mobile telephone interviews will costs about 2-2.5 times as much as landline telephone interviews, mainly due to the greater number of calls needed to achieve an interview.

The questionnaires for the National Survey, AAS and AWS have been designed to be optimal for face-to-face interviews and make use of aural and visual presentation of questions and response lists (i.e. show cards in face-to-face interview), whereas only aural presentation can be used in telephone interviews. Similarly, the questionnaires for WHS have been optimised for self-completion and rely solely on visual presentation. Question formats that rely on visual presentation may have to be changed for telephone interviews, which could cause differences in measurement. For example, face-to-face and self-completion questionnaires tend to include multi-coded questions with long response lists. For telephone interviews, it is common practice to change these questions into a series of yes/no questions for each item on the showcard. However, there is ample evidence to show that this will result in higher endorsements of each item as well as a larger number of items being selected.[41]

Similarly, questions with scales are often split into two or more questions for telephone interviews, with respondents first being asked to report the direction of their attitude (e.g. positive, neutral, negative) before asked for the strength of their attitude (e.g. very positive or just positive). However, there is some evidence showing that this will not produce comparable data with face-to-face interviews and self-completion questionnaires.[41]

Generally telephone interviews are more burdensome for respondents compared to equivalent face-to-face interviews and, consequently, questionnaires for telephone interviews tend to be shorter. However, the average questionnaire length for the National Survey, WHS, AAS and AWS is relatively short at 20 to 25 minutes and it is probably not necessary to reduce the length of these questionnaires for telephone administration.

Telephone surveys will under-represent those with hearing problems; about one in six in the population. About six percent of those with some hearing loss will not be able to hear well enough to use a voice telephone, even with amplifiers.[42]

---

[40] Sport England (2013). Active People Survey: Mobile phone data collection. Consultation on Proposed Changes. November 2013. Last accessed on 21/07/2014 at http://www.sportengland.org/media/228207/mobile-survey-consultation-document.pdf

[41] For a summary of the literature, see section 1.1 in: Nicolaas, G., Campanelli, P., Hope, S. Jäckle, A. and Lynn, P. (2011). Is it a good idea to optimise question format for mode of data collection? Results from a mixed modes experiment. ISER Working Paper No. 2011-31. University of Essex. Last accessed on 21/07/2014 at https://www.iser.essex.ac.uk/publications/working-papers/iser/2011-31

[42] Action on Hearing Loss: Facts and figures on hearing loss and tinnitus. Last accessed on 18/08/2014 at http://www.actiononhearingloss.org.uk/your-hearing/about-deafness-and-hearing-loss/statistics.aspx .

To sum up, it is possible to select a true random probability sample using dual-frame RDD that will cover almost all of the Welsh population, but the complexity of sampling, data collection and weighting (particularly when including mobile-only households) make this an unattractive option for general population surveys in Wales. Changing the data collection mode from face-to-face to telephone interviewing will also break established time series.[43]

Option 2 Reduce the total number of interviews

Because the key cost driver in face-to-face interview surveys is interviewer time, substantial savings could be made by reducing the total number of interviews that are undertaken across the in-scope surveys on an annual basis. There are two primary ways of achieving this cut in the total number of interviews. The first is to reduce the frequency of data collection on one or more surveys. The second is to lower annual achieved sample sizes, either for the survey as a whole, or for particular question blocks/topic areas. Some combination of these two strategies would also be effective in reducing costs.

Option 2a: Reduce frequency of data collection

The total number of interviews could be reduced by decreasing the frequency of data collection for one or more surveys, while preserving all other survey design features. This is more likely to be an option for continuous surveys such as the National Survey and WHS rather than the other surveys that are conducted less frequently. The stakeholder interviews confirmed that many questions currently asked on the WHS do not need to be collected every year with such large sample sizes.

Therefore, it may be possible to move either or both of these surveys to biennial data collection. If both surveys moved to biennial data collection, the two surveys could be conducted in alternate years to ensure a more even spread of fieldwork, as well as reducing overall fieldwork in Wales.

If sample sizes remained the same as they are now, survey estimates would be produced to the same level of precision as under the current design but less frequently. Estimates that currently require one year of data would be produced every two years. Local area estimates that are currently based on two or three years of aggregated data would be produced every four or six years. This would of course considerably reduce the timeliness of the statistics. This may not be an issue for those estimates that are quite static and small area estimation may be a viable

[43] For an overview of issues when including mobile numbers in RDD surveys, see AAPOR Cell Phone Task Force (2010). New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers. Last accessed on 14/08/2014 at http://www.aapor.org/Cell_Phone_Task_Force_Report.htm#.U-zrb01wa9I

solution for other statistics at local authority level.[44] Nonetheless, this option may not be acceptable to many stakeholders who will not only require annual statistics but also need to be confident in the precision of these statistics. Another disadvantage for many stakeholders is that this option will significantly reduce the timeliness of meeting changing data needs of analysts and policy makers.

Local area statistics would be particularly affected by a reduction in sample size. WHS already aggregates data over two years to produce Local Authority level estimates and so collecting data every other year, for example, would mean that local area statistics could only be produced every four years at the same level of precision. The timeliness of these local area estimates could be improved by increasing the sample size somewhat for the WHS. The implication for the National Survey would be less severe as it currently produces local area statistics every year. Appendix A provides a comparison of timeliness for a Welsh Health Survey estimate under the current design and using simple aggregates every two years. A more detailed discussion about the use of aggregate statistics for surveys can be found in Purdon (2012).[45]

It should also be noted that the National Survey has a modular questionnaire design, with many questions being administered only to random subsets of the full sample throughout the year. If the National Survey were not carried out every year, then these modules of questions could be asked of the full sample. However, this would serve to increase the interview length.

As well as reducing the timeliness of statistics, this option would also be detrimental to operations and fieldwork processes. Decreasing the frequency of data collection would reduce efficiencies that can be achieved when running surveys continuously, such as survey-specific operational logistics and interviewer performance monitoring. Furthermore, sustaining an experienced pool of interviewers in Wales from one year to the next may not be possible for survey agencies unless they are involved in the fieldwork for both surveys. This would not only have a potential impact on response rates but might also jeopardise the capacity of survey agencies to provide adequate face-to-face fieldwork in Wales.

And finally, we understand that under current WG procedures securing budgets for biennial surveys would prove to be more difficult than is the case for annual surveys.

[44] Whitworth, A., Jones, B. & Thomas, B. (2014). National Survey for Wales, 2012-13: Understanding Wales at the neighbourhood level. Welsh Government Social Research, Number: 38/2014. Last accessed on 14/08/2014 at http://wales.gov.uk/docs/caecd/research/2014/140626-national-survey-wales-2012-13-neighbourhood-level-en.pdf

[45] Purdon, S. (2012) Generating aggregate statistics from National Survey for Wales data. A report to the Welsh Government. Last accessed on 12/06/2014 at http://wales.gov.uk/docs/caecd/research/120706nswMeth1Aggregatingsurveyyearsen.pdf

Option 2b: Reduce annual sample sizes

A straightforward way of reducing the total number of interviews undertaken and, therefore, overall survey costs would be to reduce the sample size of each existing survey, while preserving all other survey design features. For example, reducing the achieved sample for the National Survey from 14,500 to 7,250 would reduce costs from £1.4 million to approximately £0.9 million per year. Note, however, that this 50% reduction in sample size translates into only a 35% reduction in cost.

The rationale for the annual sample size of approximately 14,500 is driven by the need to produce local authority level estimates for some variables. This calculation flows in a 'bottom-up' manner such that a minimum level of precision (for example +/- 4.5 percentage points around a 50% estimate) is specified for local authority level estimates. This level of precision will imply a minimum sample size at the LA level annually which is then multiplied by the number of LAs to give the total annual sample size. While this approach to determining total sample size ensures key LA level estimates are sufficiently precise, it also means that sample sizes are considerably larger than is necessary for variables for which estimates are required only at the national level.

As for option 2(a), data would have to be aggregated over two or more years under option 2(b) to produce estimates with the same level of precision as the current surveys. Consequently, this option would also reduce the timeliness of statistics, particularly local area statistics based on WHS data. However, it is possible to produce more frequent and more up-to-date statistics by calculating moving averages rather than simple aggregates.[45] The impact of using simple aggregates and moving averages on timeliness is illustrated in Appendix A along with the current design and Option 2(a).

To conclude, considerable cost savings could be made by reducing the total number of interviews per year across all surveys, either by reducing the frequency of data collection (option 2a) or reducing the size of the samples (option 2b). Although option 2(b) is to be preferred over option 2(a), both options would reduce timeliness of key statistics (including new statistics) and there could be some loss of representativeness of the underlying single years when using aggregated data.

Option 3 Integration of surveys

The first two options effectively retain the existing survey identities and infrastructures but make changes to the mode of interview or the number of interviews undertaken. Option 3 involves a more radical realignment of the existing survey portfolio: the integration of two or more surveys, ranging from the introduction of a common core set of questions within each separate survey to complete integration of surveys into one or more survey vehicles.

Other things being equal a number of advantages should accrue when surveys are integrated:

- Improved efficiencies and reduced costs by centralising data collection, processing and management;
- Improved implementation of best practices and solutions across all surveys;
- Improved coherence and comparability of data as well as reducing duplication;
- Improved opportunities to analyse across topics;
- Improved opportunities for innovation, development and research by pooling resources.

However, there are also potential disadvantages of survey integration, such as:

- Not having survey designs and processes that are tailored to the specific needs of each survey and its users;
- A break in time series that are based on questions that need to be harmonised or are affected by other changes in survey design;
- Communication, decision making (for example on funding and other practical issues) and agreeing content within a large survey program can be more challenging;
- Consortia of survey agencies may be required to deliver fieldwork for a large integrated survey which would reduce competition between survey agencies;
- Longer questionnaires which will increase respondent burden and potentially reduce response rates.

Whether advantages outweigh the disadvantages of integration will depend on which surveys are integrated, and the nature and extent of integration.

There is of course a near limitless number of ways in which integration might take place and it is impossible to discuss them all. In the following discussion we have therefore focused on four possible approaches, ranging from minimal to maximal integration, which we feel to be especially promising.

In developing the options we have worked on the assumption that initial consideration should be given to five existing surveys but that if possible allowance should also be made for the inclusion of additional question modules[46] when needed. The five surveys explicitly included in our thinking are:

1 The National Survey for Wales (National Survey)
2 The Welsh Health Survey (WHS)
3 The Active Adults Survey (AAS)

[46] The five to ten minutes of questions required for a possible WHCS would also be included as a separate module.

4 The Arts in Wales Survey (AIWS)
5 The Wales Outdoor Recreation Survey (WORS)

Table 6.5 summarises the key characteristics of these surveys.

Table 6.5 Main surveys under consideration

	National Survey	WHS	AAS	AIWS	WORS
Achieved n / frequency	c. 14,500 annually	c. 15,000 adult q'naires annually c. 3.000 child q'naires annually	c. 22,000 biennially from 1998 to 2012 c. 8,000 in 2014	c. 7,000 every 5 years	c. 6,400 every 3 years
Topics	Various incl. well-being, community cohesion, public services, health & education	Health and lifestyle	Sports and activities	Attendance and participation in arts event	Outdoor leisure activities
Random, Quota or other non-probability	Random	Random	Random	Quota	Non-probability "RDD"
Number of interviews per household	1 adult	All adults 2 children	1 adult	1 adult	1 adult
Mode	Face-to-face	Face-to-face (household) and paper self-completion (individual)	Face-to-face	Face-to-face	Telephone
Interview length	25 mins	c. 20 mins for adult q'naire and c. 10 mins for hhld interview	25 mins	15 mins	15-20 mins
Mean no. interviews per year	14,500	15,000 adults 3,000 children	4,000	1,400	2,100

Option 3a: Core set of questions across all surveys and reduced sample sizes

One form of integration would involve introducing a common set of questions in two or more in-scope surveys. This would provide precise small area estimates for the variables in the core set. As discussed above, this approach has been successfully implemented in Scotland. On its own, however, this would not yield cost savings. In order to reduce costs, sample sizes on some or all surveys on which the core question set is included would need to be reduced. The advantage of this, compared to option 2, would be that the core question set could be combined across surveys, yielding large sample sizes for precise sub-group and small area estimates for these variables.

To some extent a core set of questions has already been introduced with harmonised questions for government social surveys being used wherever possible on the National Survey, WHS and AAS.[47] Most of the harmonised questions included in these three surveys collect socio-demographic information (e.g. age, sex, ethnicity, housing tenure economic status, education) but they also include some harmonised health variables (e.g. general health, longstanding illness).

The feasibility of this approach hinges on whether all key estimates requiring large sample sizes can be included in the core set. This will only be possible if the number of key statistics that require the larger sample size is small. In practice, it is likely that a significant proportion of the core questions will comprise demographic variables and that extending them to include more policy-relevant outcomes may result in the balance between the core and the survey specific content being unpalatable to stakeholders.

Furthermore, the flow of the questionnaire could be affected if the core set includes a diverse range of topics, particularly if these topics are also covered (perhaps in more detail) elsewhere in the survey questionnaires. Moving questions out of the core and placing them elsewhere in the questionnaire is possible but consideration should be given to the risk of context effects[48] which would reduce comparability.

Option 3b: Integrate two or three surveys

A second form of integration would involve collapsing two or more existing surveys into a single survey vehicle. Below we suggest four possibilities.

[47] Further information about the harmonised concepts and questions for government social surveys which have been developed by the Office for National Statistics (ONS) can be found in the document A-Z of Harmonised Standards.

[48] Previous questions can influence the answers respondents choose to give to survey questions. More information about context effects can be found in section 8.3.

Integrate WHS and AAS

The questionnaire content for the Welsh Health and Active Adults surveys cover related content and there should therefore be no difficulty in constructing a coherent, respondent-friendly questionnaire. The Active Adults questions might be added to the WHS questionnaire for 8,000 cases every two years[49] or for 4,000 cases every year. In order to achieve this:

1 the WHS adult self-completion questionnaire would have to be extended considerably for AAS module cases; or
2 the content of both surveys would have to be reduced substantially so that the length of the WHS adult self-completion questionnaire would remain the same; or
3 extensive sub-sampling of question modules would have to take place so that the length of the WHS adult self-completion questionnaire would remain the same;
4 some topics would need to be included less frequently than at present; and/or
5 the survey would have to be administered face-to-face, possibly with a self-completion module for WHS questions that are prone to mode effects.

The first of these would (i) prove detrimental to response rates and (ii) require considerable change to the AAS questions which are currently designed for interviewer administration. The second would involve substantial cuts in content which would require further discussion with a wide range of stakeholders. The third would require different versions of the adult questionnaire which would complicate administration to some extent, and both the third and the fourth would reduce the scope for analysis and the data could be affected by context effects[50]. And finally the fifth would raise overall costs quite considerably.

The two surveys also define adults differently which adds a further complication to integration; an adult is defined as someone who is aged 16 or over for the WHS and 15 or over for the AAS. Stakeholders of each survey would be reluctant to change their definition because these are (i) linked to key performance indicators and (ii) this would result in trend discontinuities. The AAS questions could be added to the WHS child questionnaire for 15 year olds but a solution would have to be found for the increase in content. As already shown for the adult questionnaire, any of the proposed solutions is likely to have an impact on response rates, measurement, analysis and/or costs.

For these reasons, although combining these two surveys appears, on the face of it, to be attractive, we believe that the approach would not be viable in practice.

[49] The 8,000 sample size for the AAS component is used here for illustration. This was the sample size for the AAS in 2014, but previous waves of that survey were substantially larger. Sport Wales is due to discuss this issue further with its stakeholders to establish exactly what sample size will be required to meet their information needs.

[50] Context effects are discussed in section 8.3

Integrate the National Survey, AIWS and WORS

The questionnaire content for these three surveys could be readily integrated in a survey branded as being about living in Wales. The AIWS module could be added to the National Survey questionnaire content to form a 40 minute[51] questionnaire for 7,000 cases every 5 years, for 1,400 cases annually or in some intermediate way. Similarly, the WORS questions could be added to the National Survey questionnaire to form a 40 minute questionnaire for 6,400 cases every 3 years or for 2,100 cases every year or in an intermediate way.

It is questionable, however, whether integrating these surveys would lead to net cost savings as both AIWS and WORS already use cheaper data collection methods than the National Survey (quota sampling for the former and telephone interviewing for the latter). There would however be undoubted quality improvements to AIWS and WORS data resulting from the move to a relatively high response rate random probability survey[52].

A Welsh Participation Survey

It is easy to see how question material from AAS, AIWS and WORS could be combined into a coherent unified questionnaire branded as being about participation in leisure activities in Wales. Essentially, this would involve adding questionnaire content from either AIWS or WORS to the AAS questionnaire to form two questionnaire variants:

1. A 35-minute sports and arts questionnaire;
2. A 40 minute sports and outdoors questionnaire.

Such a combined Welsh Participation Survey (WPS) might for example be fielded with 8,000 respondents every two years or with 4,000 annually, with half the questionnaires including AIWS material and half including WORS material. This would deliver an annual average of 2,000 AIWS interviews and 2,000 WORS interviews, representing an increase in annual average of 600 interviews a year for the former and a decrease of 100 for the latter.

Differences in the definition of an adult are not a problem because key performance indicators and time series for AIWS and WORS can be estimated using only data from those aged 16 and over.

For the reasons discussed in relation to the previous option, this approach would result in quality improvements but probably few cost savings.

[51] In estimating interview lengths for combined questionnaire we have assumed that 5 minutes worth of demographic material has been fielded in both parent questionnaires.

[52] Although this improvement would inevitably be attended by trend discontinuities for both AIW and WORS.

Integrate the National Survey and WHS

Although there are few similarities in questionnaire content between the National Survey and WHS there are compelling logistical and cost grounds for seriously exploring possibilities for their integration. Both surveys are continuous surveys funded by the Welsh Government, have a similar strategy for selecting addresses, approximately the same sample size and have relatively short interview lengths. By rationalising areas of similar content and dropping a small number of less policy-relevant questions, it would be possible to move to an integrated survey with an interview length of approximately 45 minutes. While this would represent a substantial increase in the current interview length for both the National Survey and the WHS, for many surveys interviews are this long or longer. Although there is a negative relationship between questionnaire length and response rates for self-completion surveys, the limited evidence is less compelling for face-to-face interviews. Indeed, given the effort and expense required to obtain a face-to-face interview, there are good grounds for arguing that it is sensible to collect as much information from respondents as possible without placing too much burden on respondents.

However, despite these similarities between the National Survey and WHS, there are of course important differences that would require careful consideration prior to integration. Of these, the most important are (i) that that the National Survey seeks to interview one randomly selected adult within each eligible household, while the WHS interviews all adults and up to two randomly selected children and (ii) that the two surveys use different data collection modes. As we discuss below, we do not regard these differences as presenting insuperable barriers to integration.

First, if we are to integrate both surveys, the adult selection procedure will need to change for at least one of the two surveys - a decision will have to be made on whether to select only one adult per household or two or more adults in the household. Although, survey stakeholders would clearly need to explore the full implications of making a change to selection procedures, our understanding is that the National Survey and WHS research teams are willing to consider the different options. We discuss the question of respondent selection further in section 8.1.

Second, data collection mode differences should be relatively easy to deal with by administering WHS questions (either all or just those which look to be vulnerable to mode effects) by means of CASI (Computer aided self-administration) during the interview. Most available evidence indicates that respondents give similar answers to questions asked whether using CASI or a paper questionnaire.[53]

[53] For further information about the use of CASI including user acceptability and data quality, please see De Leeuw, E., Hox, J. & Kef, S. (2003). Computer-Assisted Self-Interviewing Tailored for Special Populations and Topics. Field Methods, 15(3): 223-251.

In order to indicate what a combined sample might look like, we develop here two illustrative designs. In so doing, we have assumed that:

1. It will prove feasible to combine the key questions currently fielded in the National Survey and WHS into a single CAPI questionnaire taking an average of 45 minutes or less[54] to administer;
2. That achieved sample sizes of around 15,700 adults would be acceptable for both surveys.

In respect of 1, key concerns are context effects and whether respondents will accept a questionnaire containing both National Survey and WHS content. In relation to the first of these, as we discuss in section 8.3, available evidence relating to context effects indicates that these will be manageable. We are, however, less confident about the second. Although we believe that with good questionnaire design and survey branding most respondents will find the combined questionnaire acceptable, it is possible that response rates may be somewhat lower that they would for a more focused questionnaire of similar length. Preliminary pilot work would need to be undertaken to explore the extent to which this would be a significant problem.

We justify our sample size assumption on the grounds that recent implementations of the National Survey and WHS have both delivered around 15,000-16,000 adult interviews (see Table 6.6 below).

Table 6.6 Sample sizes in the National Survey and WHS as documented in most recently available technical reports

	Issued addresses	**Achieved adult interviews**
WHS (2012)	14,775	15,687
National Survey (2013-14)	23,898	14,771
Total	38,673	30,458

Combining the two surveys would bring a considerable reduction in the total number of *achieved interviews*. Taking the figures in Table 6.6 as illustrative of current burden we see that with a combined questionnaire fielded to 15,700 respondents, the total number of interviews per year would reduce by 48% (from 30,458 to 15,700). Although under the integrated survey option, interviews would be substantially longer (up to 45 minutes instead of 20-25 minutes), the impact of this on costs would be considerably less than that of reducing the sample size.

Furthermore, a considerable reduction in the number of *issued addresses* would also be achieved by this integration, although the extent of the reduction would depend on how many adults were selected per sampled household. Currently the National

[54] Because there is some content overlap in the two questionnaires.

Survey samples a single adult (aged 16+) per household whereas the WHS samples all (eligible) adults in each household. As discussed in Section 8.1 we believe that both statistical and cost advantages are likely to accrue to sampling more than one adult per household, although taking such an approach would also lead to risks of reduced response rates (because of increased household burden), intra-household correlation, and intra-household interview contamination.

In the following, we have outlined two possible designs: one in which one adult is selected in each household and one in which all adults, to a maximum of two, are selected (we have set the limit at 2 in this illustration because we consider more than two 45 minute interviews with adults would place excessive burden on households). The two tables below summarise sample numbers for the two designs on the assumptions that response rates would be 65% with a single adult selected and 60% with (up to) two adults selected[55].

Table 6.7 Illustrative sample numbers with single adult selection

	n	**%**
Issued addresses	27,280	100%
Deadwood	3,137	12%
In-scope addresses	24,143	100%
Responding adults	15,693	65%

Table 6.8 Illustrative sample numbers with selection of up to two adults

	n	**%**
Issued addresses	18,260	100%
Deadwood	2,100	12%
In-scope addresses	16,160	
of which **Contain single adult**	6,092	
Contain 2+ adults	10,068	
Total selected adults	26,228	100%
Responding adults	15,737	60%
Responding HHs – min	9,696	
Responding HHs – max	15,737	

[55] These response rate assumptions have had to be based on the authors' combined professional judgment in the absence of surveys closely paralleling the proposed survey in content, survey population and sampling mechanisms.

Reductions in numbers of issued addresses differ for the two designs relative to the current 38,673: with a single adult selected the issued sample reduces by some 29% with the single adult design and by 53% with the up to two adults selected design.

Fieldwork cost savings arising from selecting two adults in each household, rather than just one, would be perhaps in the order of 10-15% for a given sample size. The effective sample size for each design would, we judge, be fairly similar, although for the two-adults design it would vary by survey variable intra-household homogeneity.

Option 3c: Integrate all surveys retaining current questions and survey sample sizes

A more ambitious redesign, which has the potential to yield both cost-savings and a more flexible and efficient means of meeting WG and WG funded bodies' survey needs would involve integrating the majority of their surveys into a single, modular continuous survey. This could readily be achieved by building on Option 3b, using the National Survey and WHS as a joint vehicle for an integrated survey.

Combining National Survey and WHS interviews as discussed in Option 3b offers no scope for integrating any other Welsh surveys without increasing the interview length to more than 45 minutes (which we consider to be undesirable with regard to response rates). Although additional surveys *could* be included by the simple expedient of considerably reducing the number of National Survey / WHS questions included in the questionnaire (see discussion of option 3d below), integration can also be achieved *without* removing questions by increasing the overall sample size. We explore the latter approach here.

We start, by way of illustration, by exploring how an integrated survey might accommodate the Active Adults Survey (AAS) *without* removing National Survey / WHS questions before outlining wider possibilities for integration.

Including the Active Adults Survey
Inclusion of AAS in an integrated survey based on a combined National Survey-WHS vehicle would be possible if three assumptions were met:

(i) that the AAS questions could be reasonably asked in the same questionnaire as either National Survey questions or WHS ones;
(ii) that both National Survey+AAS and WHS+AAS combinations produce interviews of no longer than 45 minutes;
(iii) that all questions in the combined questionnaires are relatively insensitive to context effects.

As noted above, the AAS questions can be readily combined with the WHS questions to form a coherent questionnaire. They fit less well with the National Survey questions but could probably be combined with them with good questionnaire design and suitable survey branding. The second assumption is readily met even

with little or no question pruning, and as discussed in section 8.3 we consider context effects to be manageable.

If we wished to maintain the AAS at its current 8,000 interviews every two years[56] we could integrate it with an expanded National Survey-WHS vehicle in one of two ways:

(i) by expanding the option 3b National Survey/WHS sample by 4,000 interviews biennially;
(ii) by expanding the option 3b National Survey/WHS sample by 2,000 interviews annually.

With the first approach the annual sample size would alternate between 19,700 and 15,700 interviews on a two-yearly cycle. With the second approach it would remain constant from year to year. Possible approaches to structuring the samples under the two approaches are shown in tables 6.9 and 6.10 below. It should be noted however that there would be some flexibility within the two overall designs. For example, they could easily be amended so that AAS questions were fielded alongside WHS questions but not National Survey ones, or vice-versa.

Table 6.9 Including the AAS interviews every two years

Interview content	Year 1, 3, etc	Year 2, 4, etc
National Survey+WHS	11,700	15,700
National Survey+AAS	4,000	0
WHS+AAS	4,000	0
TOTAL	19,700	15,700

Table 6.10 Including the AAS interviews annually

Interview content	Annual
National Survey+WHS	13,700
National Survey+AAS	2,000
WHS+AAS	2,000
TOTAL	17,700

Both designs are economically advantageous in that they allow 8,000 AAS interviews to be delivered every two years whilst increasing the net sample size by 4,000 only. Although the 4,000 additional interviews would be 45 minutes long and therefore more expensive than would be 4,000 25-minute AAS interviews they would still cost considerably less that the currently fielded 8,000 AAS interviews

[56] As noted previously, the exact sample size needed for the Active Adults component is still to be finalised; 8,000, as per the 2014 wave, is used here for illustration.

(other things being equal we estimate that the cost of a 45 minute interview is about 1.4 times that of a 25 minute one).

Adding questionnaire modules

The sample size for the National Survey+WHS vehicle could be further expanded either to increase the AAS sample size, to include additional survey modules or both.

The maximum sample size that could be attained with the National Survey+WHS vehicle would be 31,400 interviews annually, although in practice it is unlikely that such a large sample would ever be required.

Under the maximal National Survey+WHS vehicle design, and assuming that 4,000 AAS interviews are conducted annually, additional 20-minute questionnaire modules could be fielded to 27,400 respondents annually. New and old questionnaire modules could be combined in many ways. Table 6.11 below shows how one might add three survey modules (X, Y and X), each fielded to over 9,000 respondents annually.

Table 6.11 Illustrative example of a maximal National Survey-WHS vehicle design

Interview content	Annual
WHS+AAS	4,000
WHS+ survey module X	9,133
WHS+ survey module Y	2,567
National Survey+ survey module Y	6,567
National Survey+ survey module Z	9,133
TOTAL	31,400

In terms of individual survey modules this would deliver the following interview numbers:

WHS:	15,700
National Survey:	15,700
AAS:	4,000
survey module X:	9,133
survey module Y:	9,134
survey module Z:	9,133

Illustrative example

To give a better sense of how this approach might look in practice, in Table 6.12 below we present an integrated design which would deliver annually:

- 15,700 WHS interviews
- 15,700 National Survey interviews
- 4,000 AAS interviews
- 2,000 AIWS interviews
- 2,000 WORS interviews
- 4,000 omnibus interviews

Table 6.12 Illustrative example of integrated survey which includes the five major Welsh surveys plus an omnibus module in an annual survey

Interview content	Annual
WHS+National Survey	9,700
WHS+AAS	4,000
WHS+ WORS	2,000
National Survey+ AIWS	2,000
National Survey+ omnibus	4,000
TOTAL	21,700

Of course, in reality there may well be a preference to include different survey modules on different cycles, meaning that, each year, survey combinations would differ. In developing such a design it would be prudent to make efforts to balance modules across survey years in order to ensure that the overall survey sample size does not vary excessively from year to year.

Methods for combining questionnaire modules

In the example just discussed, we have deliberately kept matters simple and assumed (i) that questionnaire modules each comprised simple question sets and (ii) that the module combinations would remain unchanged throughout a survey year.

In practice it would be possible to combine questionnaire modules in a more flexible manner, thereby allowing larger numbers of questions to be fielded overall. The various ways in which this might be done can be seen in the current implementation of the National Survey. The National Survey comprises a core set of questions which are administered to all respondents in every quarter. Allied to this core set are rotating modular question blocks which are either administered to the full sample but not in every quarter, or administered to random sub-samples also on a periodic basis. By analogy, in our illustrative design, one or the other of the National Survey-WHS question set would form the questionnaire core and the AAS questions would form one rotating module. What we have designated modules X, Y and Z in Table

6.11 could be either simple survey modules or 'place markers' for more complex arrangements.

As examples of what we mean by this, what we have designated survey module Z might in fact represent two related question sets each asked of around 1,140 respondents every quarter (4,570 annually), or four different question modules each asked of 2,280 respondents for a single quarter only, or some other complex method of fielding different question sets within and across quarters.

Carrying capacity

As stated above under our current assumptions, the National Survey+WHS vehicle can range in size between 15,700 and 31,400 interviews annually. However, if survey modules other than those derived from the National Survey/WHS are to be included in the integrated survey it would be necessary to increase the sample size from the minimum 15,700. As can be seen below, expanding the National Survey+WHS vehicle sample size by *n/2* respondents allows additional 20-minute questionnaire modules to be fielded amongst *n* respondents.

National Survey-WHS vehicle sample size	Additional capacity (20 min. modules)
15700:	0
20000:	8600
25000:	18600
31400:	31400

Thus with a vehicle sample size of 25,000 it would be possible to field 4,000 AAS and 14,600 other modules annually. The additional modules might be used in a number of ways: ranging from a single large survey requiring LA level estimates to five or so surveys providing national estimates.

Impact on need for local area estimates

The overall sample sizes for the National Survey, WHS and AAS are all currently determined by the need to deliver LA level estimates. The WHS also has to deliver estimates for Local Health Boards (LHBs) but because each LHB consists of one or more LAs any design capable of delivering LA estimates is, ipso facto, capable also of delivering LHB estimates.

More generally, the sample size of an integrated survey will in the main be determined by the size of finest grained geography for which estimates are required. As a result of this, if the Williams Commission recommendation to reduce number of LAs from 22 to twelve is implemented, this is likely to have a major impact on desired sample size. If the design continues to deliver around 710 interviews per LA the overall minimum sample size for an integrated survey would be reduced from about 15,700 interviews to 8,600. Alternatively, the LA level sample size could be

increased whilst still reducing the size of the all-Wales sample; for example a national sample of 12,000 interviews nationally would now deliver 1,000 interviews per local authority.

We note that under the proposed changes a design which delivered a minimum sample in each local authority would also deliver sufficient interviews for separate LHB estimates.

A second important implication of the need for local area estimates concerns statistical efficiency. Both the National Survey and WHS are currently designed to deliver LA level estimates. The most statistically efficient design for achieving this, one which delivers equal numbers of interviews in each LA, is however not the most efficient design for delivering national estimates – because data need to be weighted by LA level selection weights in order to deliver unbiased national estimates. Under the current sample designs for the National Survey, WHS and AAS, this weighting increases design effects for all-Wales estimates by a factor of around 1.2. Although the consequent increase in the width of confidence intervals for national estimates is of relatively little concern for modules asked of large samples, this does become an issue when question modules are asked of relatively small numbers of respondents with national estimates in mind. This concern can be easily dealt with by varying the number of cases to whom relevant questionnaire modules are administered across LAs such that the number of respondents of whom new modules are asked in each LA is inversely related to the LA selection probabilities applying to the National Survey/WHS modules thereby reducing design effects for national estimates. In principle this approach could be used to deliver an equal probability address sample for additional questionnaire modules whilst retaining the LA estimate optimised design for National Survey and WHS modules and for any other modules for which this was required.

Option 3d: Integrate all surveys with reduced number of questions and reduced sample sizes

In developing options 3b and 3c we assumed that any integrated design would need to allow all (or almost all) current survey questions to be asked of roughly the same numbers of respondents as they are currently asked of. For example, we assumed that both the National Survey and the WHS would require retention of current questions and no reduction of sample size. Even with these requirements we identified considerable scope for cost savings.

However, to provide as wide a range of options as possible, for option 3d we have relaxed these assumptions and developed a design in which previously-included questions are asked of fewer respondents or none at all. As there are many possible ways to do this, we agreed with WG and WGSBs a broad set of characteristics of an approach which would be potentially acceptable and affordable. We develop this below.

This approach would have the following characteristics:

1. A reduction in the number of respondents/frequency of data collection (relative to the current situation) for a number of question modules[57];
2. Face-to-face interviewing with some CASI sections as required (for some of the WHS questions for example);
3. 40-45 minute interviews;
4. 12,000 interviews annually distributed evenly across twelve local authorities (this number is taken from the Williams Commission report; the actual number of authorities post-reorganisation is not yet known).

Characteristic 1 represents a major change from the approaches developed in options 3b and 3c where it was assumed (i) that much of the structure of parent survey questionnaires would be maintained and (ii) that there would be no (or at least very little) loss in sample size or administration frequency for any questions administered in the current surveys.

Characteristics 2 and 3 are in line with the approaches developed in options 3b and 3c, as is characteristic 4 if the Williams Commission recommendations are implemented.

In order to implement the approach, mechanisms need to be developed for categorising question modules and allocating them to questionnaires over time. On the basis of our discussions we propose that question modules should be categorised into one of the following four groups:

1. *Core* question modules to be asked of all respondents; these will include demographic questions such as age, gender, ethnic group, economic activity and education, and other questions required for the whole sample;
2. *Non-core trend* question modules to be asked regularly, but not necessarily annually, in order to allow trends to be monitored; the modules might be administered to the whole sample or to random sub-samples depending on sub-group analysis requirements; questions in such modules should remain unchanged from implementation to implementation;
3. *Non-core time-limited* modules to be included for a set time period in response to current policy needs; these modules might be administered to the whole sample or to random sub-samples depending on sub-group analysis requirements;
4. An *omnibus* module enabling questions to be included on a short-term as-needed basis; the module would allow maximum flexibility in question administration enabling questions to be, for example, administered for single

[57] We use the term question "module" to refer to an individual questions or a group of linked questions which are asked together.

quarters or to random sub-samples; for some quarters there may be no requirement for an omnibus module at all.

Core question modules will be administered to all respondents in all survey quarters in the integrated survey. If demand is sufficient the omnibus module will also be included in all survey quarters, although not necessarily administered to all respondents in the quarter; individual questions included in the omnibus module will, of course, vary from quarter to quarter.

Non-core trend and time-limited modules will be allocated to questionnaires on the basis of their position in a two dimensional matrix crossing required implementation frequency with required annual sample size[58]. Modules might, for example, be fielded annually to half the respondents, every two years, to all respondents, every three years, to a quarter of respondents every three years and so on.
It is important to note that this design will hold both overall sample size and interview length constant and, as such, will closer in structure to the combined survey designs discussed for option 2b than to the more flexible option 3c designs (which accommodated increases in questionnaire material by increasing sample size). Furthermore, because *both* interview length and overall sample size are to be held constant in this way, the number of modules that can be allocated to the various cells in the frequency by sample-size matrix will be mutually interdependent. For example, if module X containing 5 minutes of questions is administered to 6,000 respondents in years Y1, Y3, Y5, etc., it will need to be balanced by allocations of one or more other 5-minute modules (i) to the 6,000 non-module X respondents in Y1, Y3, etc. and (ii) to all 12,000 respondents in years Y2, Y4, etc. Of course, if it is possible to allow the overall interview length to vary a little (say between 40 and 45 minutes), some minor imbalances will be allowable.

We note that the design will allow some sample optimisation for supra-LA (e.g. national) estimates for a limited number of modules, but this will almost certainly come at the price of variability in interview length. By way of illustration, if module X is allocated to a sub-sample of 2,000 optimised for national estimates, it will be allocated to more than one in six respondents in LAs with larger than average populations and to fewer than one in six respondents in LAs with smaller than average populations. Given this variability, it will almost certainly be impossible to allocate precisely balancing modules across the remaining 10,000 cases. This difficulty was avoided for option 3c by allowing sample sizes to vary across LAs.

In the light of the constraints outlined in the last two paragraphs, the allocation of modules to sub-samples and survey years will require careful long-term advance planning, and will also almost certainly require compromises to be made in the allocation of some modules.

[58] We are assuming here that core and non-core questions modules will remain unchanged *within* survey years.

Allocation of questionnaire space to stakeholders

Because interview length will be largely fixed and because the overall achieved sample size is to be limited to 12,000 respondents, it will generally be the case that survey stakeholders will have either to reduce the number of questions they field or reduce the number of respondents they field them to (or both). In the face of these reductions a fair questionnaire space allocation system will be required. The simplest approach to this will probably be to allocate an agreed number of interview minutes to each stakeholder on the basis of an agreed formula (presumably largely based on funding input). The calculation of number of minutes taken by a given stakeholder will, of course, have to take into account both sub-sample size and implementation frequency for each question module fielded by that stakeholder. For example, a stakeholder's 10-minute allocation of questionnaire space might be used to field: two five-minute modules to the whole sample every year; four five-minute modules to the whole sample every second year; four five-minute modules to a half-sample every year; or something more complex such as:

Frequency	Proportion of sample receiving 5-minute module	Mean questionnaire minutes per year
Annual	1	5
Annual	0.25	1.25
Biennial	1	2.5
Biennial	0.25	0.625
Biennial	0.25	0.625
		10

Of course, from time to time, additional modules may need to be added to the integrated survey (a Welsh Housing Condition Survey module for example[59]). Such needs can be accommodated in one of three ways:

1. by reallocating interview time without increasing interview length or sample size;
2. by increasing overall questionnaire length either by lengthening the interview or by adding a module to be administered using a different mode (as when a self-completion module is added to a CAPI interview;
3. by increasing sample size.

Of these, approach 1 would be the cheapest to implement, but would also require other stakeholders to reduce their questionnaire allocation, something which may prove unpalatable. Approach 2 would be possible and would be less expensive than approach 3, but would almost certainly be associated with either reductions in

[59] See section 8.4.

response rate (possibly to unacceptably low levels) or poor add-on module completion rates (depending on the approach taken to increasing questionnaire length). For this reason we suspect that approach 3 may be the best way of meeting additional demand for questionnaire space. In order to field new question modules for *n* respondents the overall integrated survey sample size would need to be increased by *n/2* respondents (see discussion of option 3c above), and this would almost certainly prove to be substantially cheaper than the alternative approach of setting up a separate stand-alone survey.

7 Meeting ad hoc survey data needs: the 'omnibus' option

In addition to offering advice on the redesign of regular large-scale social surveys, the project team was also asked to consider how the Welsh survey portfolio could be enhanced to meet ad hoc survey needs as well as to collect additional data to maintain time series. These supplemental data are currently met using Welsh omnibus surveys which use quota sampling and are therefore at risk of unquantifiable survey error.

Of all the options described in this report, only options 3c and 3d have built-in flexibility for fielding rotating and ad hoc question modules without the need for setting up separate data collection exercises. Rotating modules for maintaining time series can be planned in advance while some questionnaire space can be reserved for immediate policy information needs.

For the other options, these additional data needs cannot be met within the survey vehicles without further increasing questionnaire length, substantially reducing existing content or frequency, or extensive sub-sampling of the questionnaire. An alternative strategy is to piggy-back on the survey vehicles, either following up respondents with an extra questionnaire or recruiting them to take part in a panel survey. Piggy-backing is more practical and efficient than selecting fresh samples because it eliminates the extra cost and effort spent on ineligible cases and hard refusals. It provides the opportunity of covering all household members including children. It is a particularly cost effective method for sub-sampling based on survey responses or respondent characteristics. Specific survey respondents, for example Welsh speakers, can be identified and followed-up to explore some of their survey responses in more depth.[60] Further cost savings can be achieved by using cheaper data collection modes at the follow-up stage, taking advantage of the more detailed contact information collected in the original survey such as names, telephone numbers and email addresses. Email contact and web data collection are particularly attractive for meeting immediate policy information needs because of speed as well as low cost.

In order to have the option of carrying out a follow-up study, respondents would have to be asked for their permission to be recontacted in the original survey. Typically up to 80% of respondents to face-to-face surveys will agree to do so.[61] If permission is given, further contact details should be collected including email addresses which will facilitate web data collection. However, only about two in five respondents who

[60] For example, self-completion questions on Welsh language use were left for Welsh speakers living in households covered by the National Survey in 2013-14 and 2014-15.

[61] This proportion will vary depending on survey topic, survey mode, respondent burden, and the wording of the recontact question. For example, the proportion agreeing to recontact is 68% for the National Survey.

agreed to be recontacted would be able and willing to give an email address so other modes of contact are likely to be needed.[62]

7.1 Follow-up studies

Routinely asking permission for recontact on all in-scope surveys would provide WG and WG funded bodies with a pool of named individuals with postal addresses, telephone numbers and email addresses that could be followed-up using a variety of modes when needed – either for pre-planned data collection to fill information gaps and maintain time series, or ad hoc data collections to meet immediate policy information needs. However, not everyone who agrees to be recontacted will take part in a follow-up study.[63] Consequently the cumulative response rate would likely be low, particularly if the response rate to the first study is low, and analyses of smaller sub-groups of interest will be problematic.

For these reasons, options 1a and 1c are unlikely to be suitable vehicles for follow-up studies. Agreement to recontact will be very low when asked in a postal survey (option 1a) but those who agree are likely to be highly compliant with the request to take part in the follow-up study. We therefore envisage a cumulative response rate of about 5% to a study that follows up postal survey respondents.[64] The response rate to a RDD survey (option 1b) will be lower than for a postal survey but agreement to recontact is likely to be higher. Depending on the mode of follow-up, we estimate a cumulative response rate of about 10% to a study that follows up RDD survey respondents. [65]

Although higher, options 1b, 2a, 2b, 3a and 3b will also produce low cumulative response rates. If piggybacking on the National Survey, we would envisage cumulative response rates in the region of 15%-20% for web completion, 25% for postal questionnaires, 30% for telephone interviews and 35% for face-to-face interviews.[66]

---

[62] D'Ardenne, J. & Blake, M. (2012). Contacting respondents for survey research: Is email a useful method? Presentation given at the SRA conference in November 2012. Last accessed on 24/07/2014 at http://the-sra.org.uk/wp-content/uploads/ardenne_blake.pdf

[63] Actual response rate will depend on saliency of the topic, mode of contact, mode of data collection, time lag between first survey and follow-up study, etc.

[64] This estimate is based on the assumptions that the response rate to the original postal survey is 35%, agreement to recontact with contact details is about 10%-20% and completion of the follow-up questionnaire is about 70%-90%.

[65] This estimate is based on the assumptions that the response rate to the original RDD survey is 25%, agreement to recontact with contact details is about 60% and completion of the follow-up questionnaire is about 60%-80%.

[66] These are very rough estimates. For example, the cumulative response rate for the web follow-up assumes a response rate of about 60%-70% to the National Survey (depending on which option), about 60%-70% of the respondents will agree to recontact of which about 75%-80% will have access to the internet and about 50%-60% will complete the follow-up questionnaire online.

Losing a large portion of eligible respondents at the follow-up stage can add to the non-response bias in survey estimates, but only if those who don't take part in the follow-up study differ from those who do in systematic ways that can affect survey estimates. It has been demonstrated that low response rates do not necessarily translate into observable non-response bias (Groves and Peytcheva, 2008).[67] Nonetheless, there is an increased risk of non-response bias which can be assessed using data from the original survey.

7.2 Web panel

Alternatively, those who take part in Welsh surveys could be asked to join a panel which would involve their being surveyed at regular intervals by telephone or online. The high costs of setting up a panel can be reduced by attaching the recruitment stage to one of the Welsh surveys. The most obvious choice of mode would be web because of its low cost and speed. However not everyone has access to the internet (about 25% of Welsh households)[68] and those without internet access are known to be different from those who do. If population coverage is more of an issue than speed, then offline households need to be included.

A number of web panels in Europe[69] and the United States[70] have addressed the problem of web non-coverage by providing offline households with internet access (i.e. internet connection and device) but this requires significant resource investment. Another strategy is to offer an alternative mode to offline households such as a postal questionnaire, as is done with the GESIS web panel in Germany, or a telephone interview. Mixing postal questionnaires or telephone interviews with web questionnaires will have implications for questionnaire design and measurement (e.g. short and simple questionnaires). Offering postal questionnaires alongside web questionnaires will increase the length of the fieldwork period.

Panel surveys will suffer from cumulative non-response but this can be minimised by using face-to-face recruitment at the end of a regular survey interview. Drop-out between the recruitment stage and the first questionnaire can be reduced by using incentives and following-up soon after recruitment. Regular contact and further incentives would help to maintain the commitment of panellists. Attrition can be addressed through ongoing recruitment to maintain and refresh the panel.

[67] Groves, R.M. & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. Public Opinion Quarterly, 72 (2): 167-189.

[68] Welsh Government (2014). National Survey for Wales: Headline results, April 2013 – March 2014. Last accessed on 12/08/2014 at http://wales.gov.uk/docs/statistics/2014/140530-national-survey-wales-2013-14-headline-results-revised1-en.pdf

[69] The LISS panel in the Netherlands, the GIP in Germany and ELIPSS in France.

[70] For example, the Knowledge Networks panel in the USA run by GfK.

Nonetheless, we have estimated that the cumulative response rate at the first wave of a web panel would be about 20% and this will only be achieved if piggy-backing on a survey with a high response rate (e.g. the National Survey), panellists are recruited face-to-face at the end of the original survey interview, and incentives are used for recruitment and survey participation.

To conclude, our view is that the piggy-back approach, either a follow-up study or web panel, is very much a second-best solution compared to option 3(c) with its built-in flexibility for rotating and ad hoc modules.

8. Further methodological considerations

8.1 Respondent selection

The WHS selects all adults in each sampled household, whereas the remaining in-scope surveys select a single adult only. A third approach which is sometimes adopted is to select all adults to a maximum of *n* (usually two).

There are advantages and disadvantages associated with selecting one as opposed to multiple adults. These are summarised below in Table 8.1.

Table 8.1 Advantages and disadvantages of different adult selection regimes

	Select one adult	**Allow more than one adult interview**
Survey burden on household	Less	More
Inter-respondent contamination (respondents discuss survey content before all interviews completed)	None	Possible but unlikely to be significant
Statistical efficiency	Design effects arising from inverse probability weighting: c. 1.2	Design effects from clustering within households: depends on number interviewed and homogeneity of variables used. Roughly equivalent statistical efficiency to one-adult design for variables with household intra-class correlations of about 0.2
Accuracy of data on household adults and household-level data	Less	More
Cost for a given *n*	Higher	Lower perhaps by 10-15% for a given sample size

With a short interview covering an uncontentious topic[71] that is not expected to be heavily clustered within households, the advantage lies with interviewing more than one adult per household: household burden will not be excessive, contamination is unlikely to be a problem, overall design effects will be lower, household-level data and data on other adults in the household are likely to be more accurate, and it will be possible to afford a larger net sample for a fixed budget. Although these conditions apply to the WHS as currently designed, this may not be the case for an integrated survey.

One consequence of integrating surveys would be to increase interview length to around 45 minutes which is more burdensome to administer than the current short WHS adult self-completion questionnaire. Despite this we note that there are

[71] The feasibility of including new topics that are contentious would have to be carefully considered. One option is to include these questions in a self-completion questionnaire that is completed during the visit.

successful precedents for sampling two or more adults per household using interviews of this length[72] and therefore judge that this burden would be acceptable (bearing in mind that it may result in slightly lower response rate than the single adult design).

We also do not consider that respondent contamination would be an issue given most of the survey topics to be covered.

There remains of course the question of relative statistical efficiency. With a design which samples a maximum of two adults in a household, similar design effects (taking account of both selection weighting and household level clustering) will hold for variables for which intra-household correlations are around 0.2. This is a relatively high value for social surveys and for many variables its value will be lower thereby giving these variables a statistical efficiency advantage to the two-adults design. However, we acknowledge that for some variables there will be relatively high levels of homogeneity within households (views of services or participation in cultural events, for example) and for these, statistical efficiency will be lower than for a single adult design. Ultimately the question of how many variables will fare better under which design can only be determined on the basis of empirical test.

If it were decided to consider seriously using a multiple-adult per household design for an integrated survey with interview of 45 minutes or so as discussed above, we would recommend limiting the number of interview per household to two in order to cap maximum household burden and to control the statistical impact of household level clustering. Although we acknowledge that all-adults designs are workable (the WHS being a good example of this), we counsel against their adoption here unless (i) interviews could be reduced in length considerably from the 45 minutes or so considered in our earlier discussions and (ii) we could be confident that included variables would not be excessively homogeneous.

The questions needed for a possible WHCS could be answered by any adult aged 18 or over in the household. This is not an issue as long as one of the selected and responding adults is aged 18 or over. Alternative arrangements would have to be made when this is not the case, such as the interviewer identifying an adult aged 18 or over who is prepared to answer five to ten minutes of questions about the property.

We note that the WHS selects up to two children per household for questionnaire administration. We see no reason why this should be changed in an integrated survey. In households where the WHS module was administered to adults up to two children would be selected and administered paper self-completion questionnaires

[72] For example the Health Survey for England, the National Travel Survey and the National Adult Learning Survey.

as is currently done, with interviewers either revisiting the household to collect the questionnaires or the questionnaires being returned by post.[73]

8.2 Longitudinal data collection

Another possible way of redesigning one or more of the existing surveys would be to introduce a longitudinal dimension to the design, in which the same respondents are interviewed on two or more occasions. Some cost efficiencies can be obtained by re-interviewing the same households because rates of contact and cooperation are generally higher following an initial successful interview compared to a completely fresh cross-sectional sample. Additionally, a longitudinal design provides information on within-person/household change over time which can be advantageous analytically when attempting to make causal inferences.

However, the potential cost savings from moving to a longitudinal or partially longitudinal design for the in-scope surveys would in practice be likely to be, at best, small unless different data collection modes were adopted for follow-up interviews, thereby possibly introducing differential measurement effects. A longitudinal design would also introduce a higher level of analytical complexity for data users.

However, to our minds the most important limitation of a longitudinal design is that it would largely remove the possibility of improving the precision of sub-group or small area estimates by aggregating over years (see Purdon 2012). Given (i) that there will be a continuing need for such cross-year aggregation and (ii) that the stakeholder interviews undertaken in the first strand of the project revealed no strong demand for (within-person) longitudinal data, we do not consider a longitudinal design to be appropriate.

8.3 Context effects

Although the questionnaire design literature[74] advises researchers to be wary of possible context effects, surprisingly little has been written on exactly when and how often these are likely to be encountered. There are numerous examples of experiments in which question order has been manipulated and found to have no impact on respondent answers[75], and in practice it is generally safe to work on the assumption that there will be no context effects outside particular recognisable circumstances.

[73] As noted under Option 1(b) in section 0, response rates for the child questionnaire will be much lower with postal return than interviewer pick-up.

[74] For example, Converse, J. and Presser, S. (1986) Survey Questions: handcrafting the standardised questionnaire. Sage

[75] For the classic example see Schuman, H. and Presser, S. (1981) Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context. Academic Press

It is noteworthy that survey series using modular questionnaires such as the British Social Attitudes Survey series and survey integration approaches such as used in Scotland in which core questions are fielded across several surveys would not provide acceptable data if this assumption was not held to be true.

The general principles outlining the circumstances in which context effects are likely to occur are as listed below.[76]

- Bigger effects are to be expected if questions are placed after blocks of questions on a similar topic;
- Effects (or larger effects) are to be expected when questions are closer together in the questionnaire;
- Larger effects are to be expected when question subject matter is more closely related or where the answer to one question has an obvious implication for how another should be answered;
- Answers to general questions can be susceptible to the context provided by preceding related specific questions, whereas the obverse is less likely;
- Context effects are more likely to be found in questions which use ambiguous terms.

We note that the above circumstances are readily identifiable, and hence avoidable, during questionnaire design. Furthermore, careful cognitive testing in which respondents are asked to explain how they came up with their answers should identify cases where inappropriate links are being made to previous items.

In summary, we are confident that the risk that context effects will lead to significant measurement error or will impact on time-series trends as a result of integration is both low and manageable. For this reason we are sanguine about the likely impact of context effects in any integrated survey design. This does not of course mean that we can afford to ignore the possibility of context effects. Rather, it is to say that with careful questionnaire design and appropriate cognitive testing and pre-testing, they can be minimized to an extent that their impact will be negligible.

8.4 Inclusion of a housing condition module

As discussed earlier, there is a possibility that a five-yearly house condition survey would be included in a future integrated survey, with an aim of achieving around 3,000 successful physical property inspections. This would include two elements:

[76] Converse and Presser (1986) Op. cit.; Tourangeau, R., Rips, L. and Raskinski, K. (2000) The Psychology of Survey Response. Cambridge; Campanelli, P. (2014) personal communication.

1. A five to ten minute questionnaire module asked at around 6,000-6,500 addresses collecting data on the property condition, dwelling space, access requirements, housing maintenance, thermal comfort, fuel expenditure, and energy efficiency measures;
2. A physical inspection of the property by a qualified surveyor.

This overall design is similar to that employed in the Scottish Household Survey SHS) in which a house condition questionnaire module, including a request for a follow-up surveyor inspection, is administered in about a third of responding households[77]. The physical survey is successfully completed at around three quarters (for example, 73% in 2012) of households where it is attempted (although in the English Housing Survey, the physical survey completion rate is considerably lower at less than 50%).

Although the Scottish experience indicates that it is perfectly possible to include a full house condition module in an integrated survey, there is one critical difference between the Scottish approach and that proposed for Wales. The Scottish household interview is split into two parts, one of which is administered to a householder, defined as the highest income householder[78] or his/her spouse/partner, and one of which is administered to a randomly selected adult, and the property questions in the household interview are asked of the former rather than of the latter. In contrast as currently proposed, the Welsh integrated survey would only be administered to a random adult and there will be occasions where this adult is not able to answer the property questions (for example if a teenage child is sampled in their parental home).

There are two ways one might deal with this situation:

1. To administer the property questions only when the random adult is also a householder.
2. To administer the property questions to a householder (who should be able to answer the relevant questions) who is not also the selected random adult meaning that two individuals will be interviewed in such households;

The first approach has the great advantage over the second that it keeps the interviewing process in each household essentially unchanged and is also likely to be associated with higher housing module completion rates.

However, it has the disadvantage that it will no longer deliver an equal probability sample of households for housing module administration. The housing module will be administered to a sample of households in which each household is selected with

[77] See for example Scottish Household Survey technical report for 2012 at http://www.scotland.gov.uk/Topics/Statistics/16002/PublicationMethodology

[78] A householder is a person in whose name the property is owned or rented. In many households couples there will be more than one householder.

probability proportional to the number of householders divided by the number of eligible adults it contains. As a result household module estimates will have to be weighted and additional design effects will be generated. We have modelled the likely values of these design effects using Scottish Household Survey data and estimate these to be modest – at around 1.09. As a result we would expect a sample of 6,000 to generate an effective sample size of around 5,500 and an effective sample size of 6,000 to require around 6,540 interviews. Alternatively, the these design effects could be eliminated by further subsampling of households for housing module administration at rates proportional to number of eligible adults divided by number of householders.

With the second approach outlined above we estimate from Scottish Household Survey data that in around a fifth of households a second interview will be required. Despite this, the approach does, however, have one potential advantage over the first in that it involves a less strict selection protocol for determining who is asked the housing questions. The interviewer will be encouraged to seek an interview with whichever householder has the greatest knowledge of housing matters irrespective of whether or not they were also selected as the random adult; with the first approach the householder selected as a random adult may not turn out be the person best placed to answer the housing questions. We doubt, however that this advantage will be significant as in practice with both approaches household members will be free to consult with another when answering housing questions. From our experience in Scotland we would expect, with this approach, to administer successfully the housing module for around 95% of the achieved sample.

With either approach the interviewed householder (whether or not also the random adult) will also be asked to give permission for the surveyor to visit.

Whichever approach is adopted, the procedures for selecting and administering housing module, and for arranging and administering the physical surveys, will require careful pilot testing both to iron out problems and establish likely response rates. That said, on the basis of our experience in Scotland we would expect to obtain physical surveys on 70-75% of housing module respondents.

9. Management of change in methodology

Changing from the existing ways of commissioning and implementing the Welsh survey portfolio to any of the 8 options set out here will require careful planning and testing prior to any permanent changeover. The need to implement transition arrangements would need to be factored in to any plan for change. However, while all options would require pre-operational stages of pre-testing and piloting, some options would require more substantial planning and transitional arrangements than others. In particular, it would be prudent to include – for some options – an overlap stage, during which the new procedures are implemented in parallel alongside the old ones in order to assess, and potentially correct for, breaks in time-series and the effect on aggregated statistics. The need to resource transitional arrangements will impact on cost-savings which accrue to the implementation of new options and should be factored in to decisions regarding which re-design is optimal.

Design changes for which an overlap between the old and the new arrangements are most necessary are those which implement changes in the mode of administration, which is to say options 1a to 1c. Because changes to the mode of administration would likely lead to differences in sample composition (selection effects) and to the measurement properties of survey questions (measurement effects), potentially substantial changes in distributions on some variables should be expected. For this reason (and to rule out the possibility that observed changes are, in whole or in part, 'real'), the new mode should be implemented alongside the existing one for the first, and possibly the first and second quarters of the new system.

For options 2a and 2b, which reduce survey frequency and sample sizes respectively, the transitional costs would be considerably less. This is because there would be no need to run parallel designs or, indeed, to implement any significant additional testing relative to the current setup.

The integration options 3b to 3d would also require a significant amount of pre-implementation planning and testing. That said, for those surveys which would maintain their existing modes of administration (face-to-face interview for the National Survey and AAS; self-completion questionnaire for WHS), the likelihood of substantial changes in sample composition and measurement effects is relatively low. This, however, is not the case for WORS which would change data collection mode or for AIWS which would change its sampling approach.

There is also a possibility of change in measurement properties of some questions because the questions preceding then will change. However, as we have argued in section 8.3, the risk of substantial context effects can be mitigated to a tolerable level through careful questionnaire design.

Given the importance of trend analysis and combined-year statistics for the surveys considered here, if one of options 3b, 3c or 3d were adopted, it would undoubtedly be highly desirable to include a period of parallel running perhaps covering two quarters. As discussed above for options 1a to 1c above, without this it will not be possible unequivocally to ascribe trend changes observed over the survey transition period to real population changes. It may also be difficult to generate combined-year statistics that span the changes in methodology.

We acknowledge however that in practice the cost of doing this may exceed the available budget. If this is the case the interpretation of observed trend discontinuities will necessarily be assumption-based and, as such, will be to at least some degree contestable. We note, however, that sophisticated statistical techniques have been developed which can help us in our efforts to quantify the extent to which observed changes coincident upon methodological changes result from real population changes and the extent to which they are artifactual[79].

Similarly it will be possible to form multi-year datasets from data collected using old and new methods without parallel runs, although we would be more secure in doing this if little change is observed across the relevant years. However, even where changes are observed we may be able to use the kinds of modelling techniques discussed in the last paragraph to improve the accuracy of such combined-year statistics. Inevitably the quality of such model-adjusted estimates will be entirely dependent on the accuracy of the assumptions underlying them.

[79] See for example van den Brakel, J. and Roels, J. (2009) A time series approach to estimate discontinuities due to a survey redesign. *Statistics Netherlands Discussion Paper (09003)*.

10. Conclusions and recommendations

In this report we have described the processes and procedures through which the project team developed a set of nine options for changing the way that Welsh Government, and Welsh Government funded bodies, surveys are designed and implemented. The core objective of any redesign is to reduce the overall costs of the total survey portfolio funded by WG and WG funded bodies, However, any redesign option must achieve this objective without inducing significant adverse effects on the quality and timeliness of statistical estimates.

In drawing up the nine options, we have undertaken desk research on the design and focus of the Welsh survey portfolio, undertaken interviews with key stakeholders, and reviewed commensurate initiatives that have been undertaken in other country contexts. In section 0 of the report, we have provided detailed descriptions of and rationales for each of the nine options, along with examples of how they might be implemented in practice and approximate estimates of cost savings. We also set out the costs and benefits of each potential change.

In this concluding section of the report, we draw this material together to provide our overall evaluations of the options relative to the agreed criteria. We also make recommendations about which options are likely to merit further consideration. Appendix B provides a summary assessment of all nine options against the agreed evaluation criteria.

Options 1a to 1c would produce substantial cost savings, in the long term probably the greatest of any of the redesign options considered in this review. However, while cost-saving is the primary objective of any redesign option, all of these options would be accompanied by significant negative impacts on data quality.

First, we would anticipate response rates in the region of 15-40% for online, postal and telephone interviews. Although it is increasingly well-established that response rate is a poor indicator of nonresponse bias, it remains the case that low response rates increase the risk of bias due to differences between the sample composition and the target population.

Second, change to self-completion modes, whether online, postal or a combination thereof, would necessitate shortening and simplifying the existing questionnaires for the in-scope surveys. Further discussion with stakeholders would be needed to test whether this is acceptable.

Third, for self-completion modes, it is very difficult to control within household selection of respondents, meaning that the survey would tend to be completed by the person in the household most willing to complete it, rather than the one who is randomly selected, particularly when incentives for completion of the questionnaire are provided. This would introduce an additional source of error to survey estimates.

Fourth, although switching to telephone interviewing would allow complex questionnaires to be administered and would enable greater control over respondent selection, it would introduce additional problems relating to population under-coverage and the difficulty (and expense) of including mobile only households in the sample.

All three options would require substantial investment in management of the transition from face-to-face to the new mode, including periods of parallel administration. This would offset, initially at least, whatever cost savings are made by switching to the less expensive mode(s). Even with careful management of the transition, it is likely that options 1a to 1c would all produce notable breaks in time-series trends. None of the options under 1 provide any new flexibility for meeting ad hoc survey needs. These would need to be met under the existing arrangements, or possibly by the introduction of a bespoke survey vehicle using postal/online methods. This would suffer from the same issues outlined above and would, again, reduce apparent cost-savings.

For these reasons, we do not recommend proceeding with options 1a to 1c, unless the required cost savings cannot be met in any other way.

Options 2a and 2b have the benefit of simplicity, in that they essentially maintain the existing survey infrastructure but reduce the number of costly face-to-face interviews that are undertaken each year. There would, therefore, be little concern regarding breaks in time-series due to changes in measurement properties and/or sample composition and the level of planning and resource required to manage the transition would be minimal.

However, the reduction in the number of interviews comes at the cost of timeliness of statistics that are currently produced using data from WHS and NSW, particularly at geographical levels below national. For local authority estimates, it would be necessary to aggregate over extended time periods using moving averages. The acceptability of infrequent LA estimates (and other lower level estimates) would need to be explored further with stakeholders. Under option 2a, the move to biennial data collection would pose problems for the maintenance of field operations of data collection agencies. Because options 2a and 2b represent essentially no change beyond the frequency of data collection/size of the issued samples for NSW and WHS, they do not offer any additional flexibility with regard to ad hoc survey needs.

Options 2a and 2b would yield significant cost-savings without introducing risks to data quality through reduced response rates or changes to the measurement properties of survey questions. The key limitation of these options relate to their reduced ability to produce precise sub-group and small area estimates in a timely manner and the constraint that would be placed on questionnaire space. These limitations would be potentially unacceptable to data users and policy-makers.

Option 2b is, in our assessment, preferable to 2a because of the implications for fieldwork sustainability and market competition between agencies if the major surveys are fielded biennially.

Option 3a would be relatively straightforward to introduce and clearly workable logistically, but would deliver only relatively modest gains and would also carry with it potential disadvantages. The approach would only deliver cost benefits to the extent that the net sample size over all surveys carrying the core question module was reduced. This would inevitably mean that both precision and ability to produce local area estimates would be substantially reduced for a large number of current survey variables. Therefore the option would only be workable if stakeholders for each current survey were able to identify a very small number of questions for which local area estimates are required. On the basis of our stakeholder interviews we are not sanguine about this possibility.

Related to this, questionnaire length would probably have to be increased in affected surveys in order to make room for additional core questions, although if this increase was not excessive, it would be unlikely to have a major impact on response rates.

We note however that there would be some need to change wordings of questions to ensure that they (i) are harmonized and (ii) are capable of being administered across different data collection modes. This would inevitably introduce trend discontinuities for affected questions.

In itself this option does not enhance possibilities for collection of ad hoc data although included surveys could be used to recruit follow-up samples as needed or even to form the basis for web panel recruitment (see section 0).

The option would have little impact on the ability of agencies to deliver fieldwork in Wales and nor would it require significant changes to how surveys are packaged.

In sum, we regard option 3a as workable and as having potential to make significant cost-savings. But in order to do so stakeholders would have to make considerable sacrifices in their ability to produce local area estimates.

The variant of **Option 3b** in which the National Survey and WHS interviews are combined offers very considerable cost savings. In order to achieve these, however, a number of important issues would need to be satisfactorily addressed. First we would need to reassure ourselves that both increasing interview length and bringing diverse questionnaire topics together into a single questionnaire would not damage response rates excessively. Although we are reasonably confident that this would not be the case, there would be a clear need for advance pilot testing to provide full reassurance.

Second, possible question context effects would need to be fully addressed, although as we discuss in section 8.3, we do not feel that this will be problematic.

Third, it is likely that WHS would have to change its data collection mode from paper and pen questionnaire to CASI or CAPI and CASI in combination and this would render the WHS trend lines vulnerable to discontinuities. In the light of the evidence on mode effects and assuming significant use of CASI in the new survey we feel that in practice this is not likely to be a major issue.

Fourth, the household sampling mechanism would change, certainly for WHS and possibly for the National Survey. For analyses based around the collection of point estimates this would not cause a problem; indeed there may be some statistical advantages for some questions to moving the National Survey to a two-adults per household design. There may, however, be other forms of analysis which, relying on WHS's current all-adults approach, would no longer be possible under a new within-household selection regime.

Fifth, as discussed for option 3a it is likely that some question wordings would have to change and this would introduce trend discontinuities.

As with the previous option, option 3b does not enhance possibilities for collection of ad hoc data although the combined survey could be used as a platform to recruit follow-up samples as needed or even to form the basis for web panel recruitment (section 0).

The option should have little impact on the ability of agencies to deliver fieldwork in Wales.

The National Survey / WHS variant of option 3b is workable and has the potential to make very significant cost-savings. However there are substantial methodological issues which would need to be considered in detail by current survey stakeholders and addressed in careful pilot work before this design was adopted.

Option 3c is the most ambitious of those considered here. It offers a mechanism for integrating a large number of Welsh population surveys, including a possible WHCS, in a manner which both controls costs and delivers to a good standard of quality (notably by ensuring that random probability samples are used).

It carries all the cost advantages discussed for option 3b but additionally offers a cost effective mechanism for collecting additional survey data. Through its modular structure it enables short to medium length survey modules to be administered to *n* respondents whilst only increasing the overall sample size by a factor of *n/2*. Given the greater sensitivity of fieldwork costs to sample size than interview length, this allows (i) survey data currently collected through face-to-face random probability samples to be collected more cheaply and (ii) survey data collected using other

modes or by means of non-random samples to be collected to higher quality standards.

Because our proposed option 3c vehicle is a development of the option 3b National Survey / WHS design discussed above, the same five methodological issues discussed for option 3b would also need to be addressed for option 3c.

Unlike the previously discussed designs, option 3c can include an ad hoc data collection module as a matter of design (i.e. as an integral part of the survey rather than as an add-on). This has the considerable advantage that would enable ad hoc data to be collected to higher quality standards – mainly because it would be based on higher response rate interviews.

Overall, we feel that this option has more potential than the others discussed so far to reduce costs substantially without reducing the overall number of questions fielded and without significantly compromising quality or ability to disaggregate data geographically, and suggest that it would be worthy of exploration in an empirically based feasibility study.

There are however two areas of significant concern with option 3c. First, in order to be fully workable a fully integrated survey would have to be delivered by a single contactor (or consortium). If the survey was relatively modest in scale it should be readily deliverable by *more than one* contractor/consortium in Wales and as such could be tendered effectively. If on the other hand it grew substantially – e.g. through the inclusion of a possible WHCS plus the surveys discussed above, it is possible that its fieldwork could only be delivered by a single consortium in Wales, comprising two or more of the major survey houses. If this were to be the case, there would be a de facto monopoly for fieldwork delivery and tendering would be ineffective. We suggest then that thought would need to be given to how a large integrated survey might be packaged for tendering purposes.

The second concern with option 3c is that it may be too expensive, and for this reason a further option was developed.

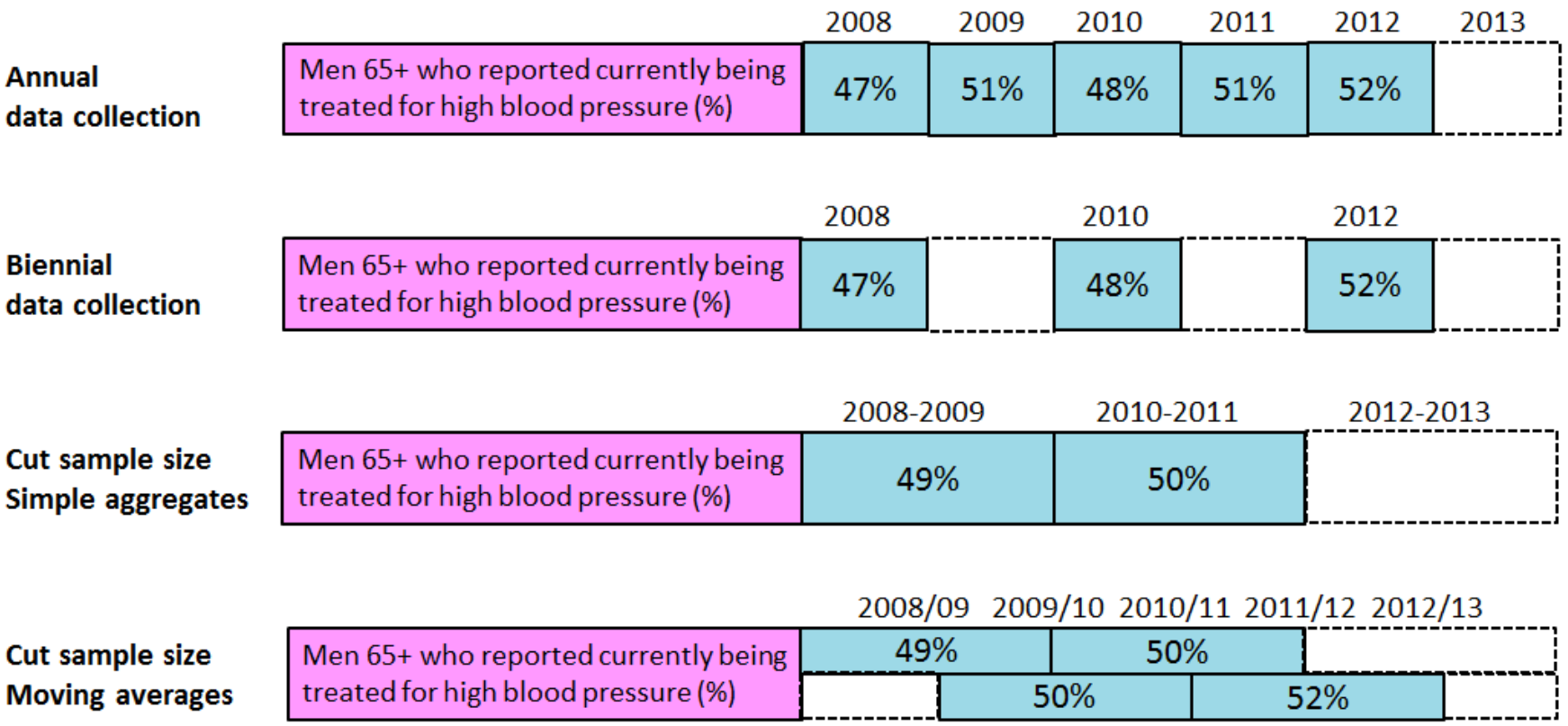
Like option 3c **option 3d** provides a mechanism for full survey integration based on 40-45 minute CAPI interviews. It differs from option 3c in that it fixes overall sample size (at 12,000 interviews) as well as interview length and in so doing requires a reduction in the overall number of questions fielded and / or the number of respondents of whom they are asked. Its modular structure retains much of the questionnaire flexibility associated with option 3c, and also like option 3c it offers a cost-effective mechanism for fielding new modules (e.g. a housing condition module) as required through sample size increases. Like option 3c, option 3d has the considerable advantage that it can include an ad hoc data collection module as a matter of design.

The five methodological issues discussed for options 3b and 3c would also need to be addressed for option 3d.

Although option 3d involves some data losses (either the loss of whole questions or reduced ability to deliver some subgroup estimates) relative to option 3c, if substantial cost reductions are required we believe it to offer the most effective available mechanism for collecting high quality data for a substantial number of variables. As with option 3c, we recommend that option 3d should be explored in an empirically based feasibility study before being finalised.

Appendices

Appendix A. Timeliness of aggregated statistics



Appendix B. Summary evaluation of the options

SELECTED OPTIONS	COST	DATA QUALITY[80]	FLEXIBILITY	MARKET IMPACT	STAKEHOLDER ACCEPTABILITY	SURVEY APPLICABILITY
Option 1: Use cheaper data collection modes						
1(a) Postal contact & self-completion questionnaire	Significant cost savings can be achieved by switching from face-to-face contact & interviews to postal contact & self-completion questionnaires.	• Almost complete coverage of the Welsh residential population. • Risk of increasing non-response bias due to low response rates. • Risk of self-selection bias. • Some risk of differences in measurement when switching from face-to-face interviews to self-completion. • Reduced comparability over time, spatial domains (e.g. other UK countries), and sub-population groups.	• Covering letters sent in English & Welsh with option of requesting Welsh questionnaire. • Less suitable for household surveys that include more than one randomly selected adult. • Ad hoc needs for data can be collected in follow-up surveys among respondents who have agreed to be recontacted, including web surveys among those who have provided email addresses when speed is of the essence. • Email addresses of respondents can be	Greater competition between fieldwork agencies; i.e. not dependent on face-to-face fieldwork capacity in Wales.	Unlikely to be an acceptable option for most if not all stakeholders because of the impact on representativity of the achieved sample, data quality in general and time series in particular.	• This design can be applied to all surveys currently using face-to-face interviewers (National Survey, WHS, AAS, AWS) and new surveys. • This design could also be applied to WORS which is a telephone survey but cost savings would be less compared to the face-to-face surveys and it will involve a change in sampling method.

[80] Relevance, accuracy & reliability, timeliness & punctuality, accessibility & clarity, coherence & comparability.

SELECTED OPTIONS	COST	DATA QUALITY[80]	FLEXIBILITY	MARKET IMPACT	STAKEHOLDER ACCEPTABILITY	SURVEY APPLICABILITY
			used to recruit web panel respondents for the collection of Omnibus data. • Representativity of follow-up samples and web panels are likely to have been seriously compromised due to cumulative non-response and web non-coverage.			
1(b) Face-to-face contact & self-completion questionnaire	Unlikely to produce meaningful cost savings because interviewer visits are needed to make contact and to collect completed questionnaires from one randomly selected adult in the household.	• Almost complete coverage of the Welsh residential population. • Impact on response rates is minimal and is unlikely to increase non-response bias. • Some risk of differences in measurement when switching from face-to-face interviews to self-completion which can be minimised through good questionnaire	• Suitable for household surveys that select one or more household members at random and surveys that screen for specific groups. • Ad hoc needs for data can be collected in follow-up surveys among respondents who have agreed to be recontacted, including web surveys among those who have provided an email address when	Unlikely to have any impact on competition between survey agencies.	This could be an acceptable option for many stakeholders, provided that changes in measurement for key estimates are either negligible or can be quantified. However, the lack of meaningful cost savings associated with this change in measurement makes this an undesirable option.	• WHS already uses this design. It could change to a mixed mode design with a web questionnaire alongside the paper questionnaire but this will complicate sample management and fieldwork without making any worthwhile cost savings. • This design could be applied to the National Survey, AAS, AWS and new surveys. • This design could also be applied to

SELECTED OPTIONS	COST	DATA QUALITY[80]	FLEXIBILITY	MARKET IMPACT	STAKEHOLDER ACCEPTABILITY	SURVEY APPLICABILITY
		design. • Some risk to data comparability but less than option 1(a).	speed is of the essence. • Representativity of follow-up surveys will be greater than under option 1(a) due to interviewer encouragement. Representativity of the web panel is compromised by non-coverage of offline households and low composite response rate.			WORS but this would involve not only a change in mode but also a change to the sampling method.
1(c) Telephone contact & telephone interview	Worthwhile cost savings can be achieved by switching from face-to-face contact & interviews to telephone contact & interviews (but less than option 1a).	• Significant risk of non-coverage bias when only landline numbers are included in the RDD sample. Almost complete coverage when mobile phone numbers are included. • Response rates will be much lower than equivalent face-to-face surveys (less than 30%). Inclusion	• Ideally telephone interviewers would be fluent in English & Welsh. If not, respondents who choose to complete the survey in Welsh must be re-assigned to Welsh speakers. • Suitable for household surveys that select one or more household members at random and surveys that	Survey agencies that can provide a sufficiently large panel of Welsh telephone interviewers, including Welsh speakers, will be at an advantage. To the best of our knowledge, such a panel does not exist at the moment.	Unlikely to be an acceptable option for many stakeholders because of the impact on representativity of the achieved sample, data quality in general and time series in particular.	• WORS already uses this design. • Population coverage could be improved by including mobile phone numbers in the sample, although this would increase cost, reduce response rates and will making weighting more complex. • This design can be applied to all surveys currently using face-to-

SELECTED OPTIONS	COST	DATA QUALITY[80]	FLEXIBILITY	MARKET IMPACT	STAKEHOLDER ACCEPTABILITY	SURVEY APPLICABILITY
		of mobile phone numbers will reduce the response rate further. • Some risk of differences in measurement when switching from face-to-face to telephone interviews. • Reduced comparability over time, spatial domains (e.g. other UK countries), and sub-population groups.	screen for specific groups. • Ad hoc needs for data can be collected in follow-up surveys among respondents who have agreed to be recontacted, including web surveys among those who have provided an email address when speed is of the essence. • Representativity of follow-up surveys will be greater than under option 1(a) but less than option 1(b). Representativity of the web panel is compromised by non-coverage of offline households and low composite response rate.			face interviewers (National Survey, WHS, AAS, AWS) and new surveys.

<u>Option 2</u>: Reduce total number of interviews						
2(a): Reduce frequency of surveys	Potential for significant cost savings - e.g. almost half of fieldwork costs in the case of biennial data collection while preserving all other design features.	• A reduction in timeliness of local estimates • Loss of annual data.	• Potentially more time to plan and set-up survey. • Off-survey years can be used for follow-up surveys using cheaper modes among respondents who have agreed to be recontacted in order to meet additional needs for data, including web surveys among those who have provided email addresses when speed is of the essence. • Respondents can be asked to join a web panel for the collection of Omnibus data. • Representativity of follow-up samples and web panels are likely to have been compromised by cumulative non-response and web non-coverage.	Ability of survey agencies to build and maintain an experienced panel of face-to-face interviewers in Wales is compromised when individual agencies are not running a large CAPI survey from one year to the next. This could reduce competitiveness.	Further discussion required with users of local estimates and those needing annual data. However, it is an acceptable option for the production of national-level statistics which do not require current large sample sizes.	More likely to be an option for continuous surveys such as the National Survey and WHS rather than the other surveys that are conducted less frequently.

2(b): Reduce sample sizes	Potential for significant cost savings - e.g. almost half of current fieldwork costs when halving the annual sample size while preserving all other design features.	• Precision of estimates reduced. • Precision of estimates can be increased through aggregation over two or more years but this will reduce timeliness of statistics. • Annual reporting with moving averages provides more timely data but these can be harder to interpret and to work with.	• Ad hoc needs for data can be collected in follow-up surveys using cheaper modes among respondents who have agreed to be recontacted, including web data collection among those who have provided email addresses where speed is of the essence. • Respondents can be asked to join a web panel for the collection of Omnibus data. • Representativity of follow-up samples and web panels are likely to have been compromised due to cumulative non-response and web non-coverage.	A reduction in the total number of face-to-face interviews being carried out will reduce the demand for an overstretched face-to-face fieldforce in Wales. This should make it easier for survey agencies to bid for contracts.	Further discussion required with users of local estimates. However, it is an acceptable option for the production of national-level statistics which do not require large sample sizes.	More likely to be an option for those surveys that currently have very large sample sizes in order to produce local estimates (e.g. the National Survey and WHS).

Option 3: Integration of surveys						
3(a): Harmonise core module of questions across existing surveys, and reduce sample sizes	Potential for significant cost savings by reducing sample sizes.	• Length of questionnaires will have to be increased to accommodate a core module which could have a small impact on response rates (evidence suggests that relatively small increases in questionnaire length have no significant impact on response rates in face-to-face surveys). • Some change may be required to questions included in the core module which may reduce user relevance. • Some change may be required to questions included in the core module which may break time series. • Precision of non-	• Ad hoc needs for data can be collected in follow-up surveys among respondents who have agreed to be recontacted, including web surveys among those who have provided email addresses when speed is of the essence. • Respondents can be asked to join a web panel for the collection of Omnibus data. • Representativity of follow-up samples and web panels are likely to have been compromised due to cumulative non-response and web non-coverage.	A reduction in the total number of face-to-face interviews being carried out will reduce the demand for an overstretched face-to-face fieldforce in Wales. This should make it easier for survey agencies to bid for contracts.	This could be an acceptable option for stakeholders, provided that key estimates can be included in the core module.	A core module of questions could be introduced on any of the existing Welsh surveys and new surveys. Many of these questions will have been designed for face-to-face interviews (including National Statistics Harmonised Concepts) and would have to be adapted if included on WORS (telephone survey).

		key estimates is reduced which can be improved through aggregation over two or more years.				
3(b): Integrate two or more surveys and reduce combined sample size	Potential for significant cost savings by reducing sample sizes. For example, reducing the combined sample sizes for the National Survey and WHS to 15,000 would introduce a cost saving of about £1m.	• Length of questionnaires will have to be increased to accommodate questions from integrated surveys which could have a small impact on response rates. • Integrating WHS with other surveys will require changes to its within-household sampling method which could reduce user relevance and introduce trend discontinuities. • Integrating WORS with other surveys will require a change to its sampling method and data collection mode which could affect time series.	• Ad hoc needs for data can be collected in follow-up surveys among respondents who have agreed to be recontacted, including web surveys among those who have provided an email address when speed is of the essence. • Respondents can be asked to join a web panel for the collection of Omnibus data. • Representativity of follow-up samples and web panels are likely to have been compromised due to cumulative non-response and web non-coverage.	A reduction in the total number of face-to-face interviews being carried out will reduce the demand for an overstretched face-to-face fieldforce in Wales. This should make it easier for survey agencies to bid for contracts.	Strong concerns among WHS stakeholders regarding the potential for discontinuities affecting trends and sub-group analysis.	All surveys could be integrated but some combinations require changes to sampling methods and data collection modes. For example, integrating the National Survey & WHS would require either one or both surveys to adopt a new procedure for within household selection. Integrating WORS with one or more of the other surveys would require a change to the sampling method and data collection mode for WORS. Integrating the National Survey, AAS, AWS and new surveys would appear to be the least problematic with the added benefit of a more robust sampling

		• Some change may be required to question wording and format to minimise repetition of similar questions which could reduce user relevance and affect time series. • Combining questionnaires could introduce context effects which will have an impact on time series.				method for AWS.
3(c): Integrate all surveys and reduce combined sample size	Potential for significant cost savings by reducing sample sizes.	• Length of questionnaires will have to be increased to accommodate questions from integrated surveys and scope for rotating/ad hoc modules which could have a small impact on response rates. • Integrating WHS with other surveys will require changes to its within-household sampling method which could	• This option is highly flexible with built-in capacity for rotating and ad hoc modules.	A reduction in the total number of face-to-face interviews being carried out will reduce the demand for an overstretched face-to-face fieldforce in Wales. How this impacts on competition will depend on the chosen size of the integrated survey, and how the contract is designed.	Strong concerns among WHS stakeholders regarding the potential for discontinuities affecting trends and sub-group analysis.	Integrating all surveys will require changes to sampling methods and data collection modes for at least two surveys: the within-household selection procedures for WHS, sampling method and data collection mode for WORS.

		reduce user relevance and introduce trend discontinuities. • Integrating WORS with other surveys will require a change to its sampling method and data collection mode which could affect time series. • Some change may be required to question wording and format to minimise repetition of similar questions which could reduce user relevance and affect time series. • Combining questionnaires could introduce context effects which will have an impact on time series.				

3(d): Integrate all surveys and limit overall sample size to 12,000	Significant cost savings by reducing sample sizes.	• Length of current questionnaires will have to be increased to accommodate questions from integrated surveys and scope for rotating/ad hoc modules which could have a small impact on response rates. • Integrating WHS with other surveys will require changes to its within-household sampling method which could reduce user relevance and introduce trend discontinuities. • Integrating WORS with other surveys will require a change to its sampling method and data collection mode which could affect time series. • Some change may be required to question wording	• This option is highly flexible with built-in capacity for rotating and ad hoc modules.	A reduction in the total number of face-to-face interviews being carried out will reduce the demand for an overstretched face-to-face fieldforce in Wales. Fully compatible with open competition.	Strong concerns among WHS stakeholders regarding the potential for discontinuities affecting trends and sub-group analysis.	Integrating all surveys will require changes to sampling methods and data collection modes for at least two surveys: the within-household selection procedures for WHS, sampling method and data collection mode for WORS.

		and format to minimise repetition of similar questions which could reduce user relevance and affect time series. • Combining questionnaires could introduce context effects which will have an impact on time series.				