



The National Survey for Wales: question comparability study

Caroline Bryson, Tom Huskinson

April 2012

Contact Details:

Caroline Bryson

Bryson Purdon Social Research

10 Etherow Street

London SE22 0JY

+447966 167192

caroline.bryson@bpsr.co.uk

Contents

1	Executive Summary	4
1.1	Introduction	4
1.1.1	The National Survey for Wales	4
1.1.2	Aims of our study	4
1.2	Principles of question comparability	5
1.2.1	Sample design and the eligible population	5
1.2.2	Survey administration	5
1.2.3	Question design	6
1.2.4	Summary of the principles	7
1.3	Concluding comments	7
2	Introduction	9
2.1	Background to the study	9
2.1.1	The National Survey for Wales	9
2.1.2	Aims of our study	9
2.2	Study methodology	10
2.3	Outline of the report structure	11
3	The National Survey questions and their comparators	12
3.1	Introduction	12
3.2	The National Survey questions	12
3.3	Comparator surveys	13
4	Principles of survey question comparability	17
4.1	Introduction	17
4.2	Sample design and the eligible population	17
4.2.1	Sample design and sampling frames	17
4.2.2	Response rates	19
4.2.3	Household versus individual surveys	19
4.2.4	Age range and other exclusions	20
4.2.5	Concluding comments	20
4.3	Survey administration	20
4.3.1	Survey mode	20
4.3.2	Focus of the interview	21
4.3.3	Timing of the survey	22
4.3.4	Concluding comments	22
4.4	Question design	22
4.4.1	Question order, and position in the interview	22
4.4.2	Question phrasing	23

4.4.3	Response scales.....	24
4.4.4	Concluding comments	25
4.5	Summary of the key considerations for the National Survey.....	26
5	A comparison of the 20 National Survey questions	27
5.1	Introduction	27
5.2	Health and well-being	27
5.2.1	Overview.....	27
5.2.2	Self-perceived health	28
5.2.3	Overall life satisfaction	30
5.3	Living in the community.....	31
5.3.1	Overview.....	31
5.3.2	Feelings of belonging	32
5.3.3	Experience of harassment and discrimination.....	33
5.4	Local authority services	35
5.4.1	Overview.....	35
5.4.2	Provision of high quality services.....	35
5.4.3	Information and involvement	36
5.5	Education and children’s development.....	37
5.5.1	Overview.....	37
5.5.2	Satisfaction with the education system	38
5.5.3	Satisfaction with secondary schools	39
5.5.4	Post compulsory school aspirations	40
5.5.5	Home learning activities	41
5.6	Health care.....	42
5.6.1	Overview.....	42
5.6.2	Satisfaction with health services.....	42
5.6.3	Satisfaction with hospital care	43
5.6.4	Satisfaction with GP care.....	44
5.6.5	Convenience of GP appointments	46
5.7	Views on other government policy issues.....	47
5.7.1	Overview.....	47
5.7.2	Satisfaction with the Welsh Government	47
5.7.3	Satisfaction with transport provision	48
5.8	Demographics	49
5.8.1	Overview.....	49
5.8.2	Welsh language skills	49
5.8.3	Educational qualifications	51
5.8.4	Financial stability.....	52
5.8.5	Use of the internet.....	53
6	Concluding thoughts	56

6.1	Introduction	56
6.2	Thoughts on future work	56
	References.....	58
	Appendix 1: comparator survey spreadsheet.....	59
	Appendix 2: comparator question spreadsheet	60

1 Executive Summary

1.1 Introduction

1.1.1 The National Survey for Wales

The National Survey for Wales was launched in early 2012 and is a key source of data on the attitudes, behaviours and circumstances of people living in Wales. It will be used by the Welsh Government in its assessment of public service provision, identifying areas for improvement, tracking change over time and helping to assess the effect of changes in national or local government policy or practice. Approximately 14,500 face-to-face interviews will be carried out each year, with fieldwork running continuously throughout the year.

1.1.2 Aims of our study

The purpose of this review, conducted by Bryson Purdon Social Research (BPSR) and Ipsos MORI, is to improve the Welsh Government's understanding of how the National Survey can be used in combination with other surveys to make comparisons – within Wales over time, across GB or the UK, or across a wider set of European countries. Although the full review has a particular focus on 20 National Survey questions, it also provides a more general overview of the issues around survey question comparability. Underlying our review are the following questions -

- When it is appropriate to use National Survey findings and survey findings in other UK or other European countries to make comparisons between the views and experiences of people in Wales and those of their UK and European counterparts?
- When it is appropriate to use National Survey findings and survey findings from previous Welsh surveys to produce time trends?
- Why, on occasions, might findings from the National Survey differ to those in other surveys fielding similar questions in Wales or elsewhere in UK?

In this summary, we focus on the key issues which arose out of the review – the main issues around survey comparability and what further work might be done, rather than the detail of the question by question comparison (which can be found in Chapter 5 of the full report).

While we found very few examples of questions which were directly comparable to the 20 National Survey questions selected for review, in the main the comparator surveys employ the same face-to-face mode and comparable random probability sampling. In the main, the comparator surveys are general-purpose, general-population surveys, although a number focus on a particular topic (e.g. health services) or a sub-group of the population (e.g. patients). Therefore, the main issues around comparability are (a) the focus of the interview; (b) the data collection period; (c) the position of the question in the interview; (d) the precise question wording; and (e) the response scales.

1.2 Principles of question comparability

There are a wide range of reasons why the same or similar question fielded on two different surveys might result in somewhat different findings. Unfortunately, identifying reasons, or potential reasons, for differences in responses to two seemingly similar questions can be easier than formally measuring the size of their effect. In most cases, the assessment of how far, and in what direction, the differences between the questions have affected the findings is a matter of judgement. The task is easier if we are comparing two identical questions fielded on two or more surveys. Comparability issues can be grouped into those concerning sampling and the eligible population; survey administration; and, question design.

1.2.1 Sample design and the eligible population

- Random probability sampling versus quota sampling: The use of random or stratified random sampling, as adopted in the National Survey, ensures that the population *approached* to be interviewed accurately represents the eligible population. Surveys adopting this approach produce comparable data provided the sampling frames cover the same population. It is appropriate to compare two random probability samples using different sampling frames provided that the analysis can be restricted to the same population. On the other hand, quota surveys ensure simply that respondents are broadly representative on a small number of (usually) socio-demographics; but those less easily available or more reluctant to take part are more likely to be excluded. We might expect different results from the National Survey and a quota survey.
- Response rates: Differences in response rates between random probability surveys can compromise comparability, particularly if non-responders are more likely to come from particular sub-groups of the population.
- Household versus individual surveys: Statistically, whether one adult (as in the National Survey) or all adults from a household are included in the survey does not affect the comparability of the surveys. However, interviewing all adults in the household is a bigger task and can lead to lower response rates, and thereby affect survey comparability.

1.2.2 Survey administration

- Survey mode: The mode in which the survey is administered is well-documented as having the potential to alter how respondents answer questions. Some of these issues are related to sampling frames available for different modes and the response rates achieved. Other mode effects are caused by the way in which questions are asked and presented in different modes. However, part of the variation in survey findings across different modes is caused by the relationship that the respondent has with the interviewer in an interviewer-administered survey versus completing self-completion questionnaires, with different modes giving rise to varying levels of acquiescence bias or satisficing.

- Survey focus: In a general survey like the National Survey, respondents are asked about a wide range of issues about which they are unlikely to be 'primed' in advance of the interview. This is not the case of other potential comparator surveys, which have a clear focus on a particular topic. We might expect some variation in the profile of people who take part in the survey and in their responses to particular questions.
- Fieldwork period: Any assessment of the comparability of surveys needs to take account of the timing of the data collection. If surveys have different fieldwork periods, differences in responses, particularly to attitudinal questions, may be a reflection of respondents' awareness of pertinent events happening locally or nationally, the position in the economic cycle, or seasonality. The extent to which this is an issue varies depending on the focus of the question, and is key for measures which might be expected to alter quickly over time or be affected by external factors.

1.2.3 Question design

- Question order and position in the interview: The way in which people respond to questions can be affected by previous questions they have been asked in the survey. For instance, an identical question about public satisfaction with the Government may be answered differently in one survey if it follows questions on political affiliation and the state of the economy compared to another where it is preceded by a set of questions less related to Government performance. This reduces the ability to make direct comparisons between two surveys.
- Question phrasing: Two similarly, but not identically phrased, questions in different surveys can elicit different responses. The impact of minor differences in wording needs to be assessed on a case by case basis. However, we can highlight two common issues which affect respondents' answers. Firstly, questions which introduce a response scale using only the start of the scale (e.g. 'how satisfied are you') are different to those which introduce both ends (e.g. 'how satisfied or dissatisfied are you'). The latter question format arguably implies that both views are equally valid, and thereby reduces acquiescence bias. Secondly, questions which are prefaced by introductory or explanatory text change the context of the questions (e.g. by prompting respondents to think about a particular range of issues) in comparison to questions which do not provide this type of introduction.
- Response scale: The precise wording and format of the response scales can affect how people respond to particular questions, and thus the comparability of questions using different response scales. Differences include -
 - The length of the scale
 - The use of a mid-point
 - Verbal or numeric scales
 - Extremity of scale ends
 - Scale presentation
 - Offering "don't know" as an option

1.2.4 Summary of the principles

We need to consider the relative weight that should be given to each of the comparability issues when deciding if and when a question provides an appropriate comparison for the National Survey. The importance of different issues will vary depending on the nature of the question. It appears that most comparable surveys for the National Survey are face-to-face interviewer administered surveys: therefore, in most cases mode effects will not be an issue. Likewise, the majority of surveys use comparable sampling methods. There are some differences in response rates, but these can be taken into account in analysis, at least to some extent.

So, the primary issues to take into account in the assessment of the comparability of questions with the National Survey are (a) survey focus; (b) fieldwork period; (c) position of the question in the interview; (d) question phrasing; and (e) response scales. While it is impossible to measure the effect of each of these, as a general rule, our view is that most attention should be paid to ensuring close comparability with the question wording and response scale. The focus of the survey and position of the question in the interview are important for questions which might arguably be affected by these issues. And the fieldwork period is especially important for variables for which we may expect fairly rapid changes over time.

1.3 Concluding comments

At the time of writing this report, no data are yet available from the National Survey. Our assessment of the questions was therefore made on the basis of a comparison of the methods and questions, rather than analysis of the results. We identified a number of questions among the 20 we were asked to review where comparisons can be made with a fair degree of confidence – either with other countries or within Wales over time. For other questions, we identified some questions which might provide relatively good comparisons, and some which provide nothing more than a broad comparison. It is almost impossible to have the ‘perfect’ comparison question, taking into account sample design, survey administration and question wordings. We illustrate how sometimes small differences (e.g. the position of the question, differences in wording) may have a large effect on how respondents react, and therefore on the ability to make robust comparisons. While we explain how some differences between surveys can be taken into account in data analysis (increasing comparability), many issues are not ‘solvable’ and must simply be recognised when assessing the findings across different surveys or over time.

Once some National Survey data are available, it will be possible to test out whether some of the *potential* comparator questions identified do in fact provide robust comparisons. Analysis of the National Survey data and data from comparator surveys will show whether, and how, the comparability issues have an effect on the distribution of responses in different surveys. The greatest opportunities for doing this are with the small number of surveys like the Integrated Household Survey which have sufficiently large Welsh sample sizes, collected over a similar timeframe to the National Survey. However, most of the queries we raise about comparability concern surveys with sample sizes in Wales too small for this kind of analysis. It would be possible to use other surveys, although in the main, this

will necessarily be a relatively unscientific exercise, comparing questions selected on the basis that we would not expect large differences between countries, and countries which we might expect to be similar to Wales. In doing so we could be *reasonably* confident that any differences identified in the analysis are *probably* due to issues of survey/question comparability.

We suggest an additional or alternative approach: to introduce split-run experiments to the National Survey to test – with a random sub-set of respondents – whether some of the comparator questions identified on other surveys would in fact elicit different responses. Fielding both the National Survey question and the comparator question on the National Survey would allow for a formal test of the difference between them.

2 Introduction

2.1 Background to the study

2.1.1 The National Survey for Wales

The National Survey for Wales was launched in early 2012 after a thorough development and piloting phase. Replacing the Living in Wales survey series, it is a key source of data on the attitudes, characteristics and behaviours of people in Wales – covering public services, where they live and how they feel about their lives. It will be used by the Welsh Government in its assessment of public service provision, identifying areas for improvement, tracking change over time and assessing the effect of changes in national or local government policy or practice.

Approximately 14,500 face-to-face interviews will be carried out each year, with fieldwork running continuously throughout the year. Households will be selected for participation from the small-user Postcode Address File, with one adult aged 16 or over per household randomly selected for interview. The current 25-minute interview covers, in broad terms: attitudes towards and experiences of the education system, health services, transport and local authority service provision; satisfaction with the performance of the Welsh Government; views of and experience of living in the local area; health and well-being; Welsh language skills, internet use and demographics.

2.1.2 Aims of our study

The Welsh Government commissioned BPSR and Ipsos MORI to conduct a review of 20 questions included in the current National Survey interview. The purpose of the review is to improve the Welsh Government's understanding of how the National Survey can be used in combination with other surveys to make comparisons – within Wales over time, across GB or the UK, or across a wider set of European countries.

In developing the National Survey, the Welsh Government drew on questions fielded on other surveys. More generally, a number of the topics included in the survey are similar to those covered within other surveys - so, while the question wordings are sometimes not identical, there are data from other surveys to which the National Survey may, or could usefully, be compared. Decisions on when such comparisons are appropriate in the face of methodological differences can be complex, and are often a matter of judgement rather than science. This report aims to add to the Welsh Government's evidence base on question comparability and to aid its work in this area – particularly in relation to the 20 National Survey questions forming the focus of this report, but also more generally about the surveys it commissions or uses.

Underlying our discussions in subsequent chapters are the questions –

- When it is appropriate to use National Survey findings and survey findings in other UK or other European countries to make comparisons between the views and experiences of people in Wales and those of their UK and European counterparts?

- When it is appropriate to use National Survey findings and survey findings from previous Welsh or cross-country surveys to produce time trends of the attitudes of people in Wales?
- Why, on occasions, might findings from the National Survey differ to those in other surveys fielding similar questions in Wales?

2.2 Study methodology

We have assessed the comparability of the 20 National Survey questions identified as of interest by the Welsh Government against a set of identical or similar questions identified from a wide range of other surveys.

In the identification of comparator questions, we focused on general population surveys fielded in one or more of the UK countries or Europe-wide, using our prior knowledge of these other surveys in conjunction with a wide-reaching internet search. The internet search process included: searching the Survey Question Bank section of the Survey Resources Network site¹ for questions on other surveys with the same or similar wording; searching questionnaires of studies identified as potentially relevant from the list of major studies (both national and international) included on the Economic and Social Data Service site², and performing more basic internet searches to identify any other questions with the same or similar wording not covered by these sites.

In the main, we restricted ourselves to other random probability surveys³. However, in cases where we found no or few comparator questions for a particular National Survey question, we have, on occasion included quota surveys. We have attempted to identify comparator questions from the major publically-available general population surveys as well as a number of specific-population surveys (e.g. surveys of parents). However, inevitably, we may have missed some potential comparator questions (for example from surveys commissioned with a particular purpose which are less readily available, as well surveys where the results or surveys materials have not (yet) been published). Moreover, with the exception of time series comparisons, the Welsh Government will want to compare the National Survey findings with data collected in a comparable time frame. However, while we know about the current frequency of data collection for each of the comparator surveys, and the latest data collection for each particular question, we cannot be sure that the comparator questions will continue to be fielded, and in what frequency. For all these reasons, the appended spreadsheets might be updated over time.

We have assessed the comparability of each of the 20 National Survey questions drawing on the body of evidence about the effects of methodology and question design on respondent responses and, therefore, on the comparability of survey questions. At the time of writing this report, no data are yet available from the National Survey. Our assessment of the questions is therefore made on the basis of a comparison of the methods and questions,

¹ <http://survey.net.ac.uk/sqb/qb/questions.asp>

² <http://www.esds.ac.uk/findingData/majorStudies.asp>

³ Quota surveys both provide less robust data and, by dint of their sample design, are immediately less directly comparable to the National Survey (see Section 4.2).

rather than analysis of the results. This further work may be useful once data are available, and is something we discuss in Chapter 6.

2.3 Outline of the report structure

In the following chapter (Chapter 3), we talk in outline about the 20 National Survey questions and the surveys which can be used for comparative purposes.

Chapter 4 details the principles of survey question comparability focusing on sample design (Section 4.2), survey administration (Section 4.3) and question design (Section 4.4). We illustrate these principles using examples drawn from the 20 National Survey questions and their comparator surveys introduced in Chapter 3.

Chapter 5 is a question-by-question discussion of each of the 20 National Survey questions identified as of interest by the Welsh Government, and their comparator questions.

We conclude in Chapter 6 focusing, in particular, on further work that could be done to test or improve the comparability of questions fielded in the National Survey.

Appendices 1 and 2 provide the full detail of the survey methodology for each comparator survey, and question structure and wording for each comparator question.

3 The National Survey questions and their comparators

3.1 Introduction

Central to this report is a review of 20 questions selected by the Welsh Government from the current National Survey interview. In this chapter, we set out the parameters of the data we have reviewed by –

- introducing the 20 National Survey questions, organised in terms of their substantive focus (Section 3.2);
- listing the surveys in which we have identified appropriate comparator questions, and discussing some of the methodological issues pertinent to comparing them to the National Survey (Section 3.3).

3.2 The National Survey questions

The 20 National Survey questions group into the following substantive topics⁴ –

Health and well-being

Self-perceived health: using 5-point verbalised scale (i.e. read out by the interviewer) from ‘very good’ to ‘very bad’;

Overall life satisfaction: using 11-point numeric scale where 0 is ‘not at all satisfied’ and 10 is ‘completely satisfied’.

Living in the community

Feelings of belonging: using 5-point verbalised scale from ‘strongly agree’ to ‘strongly disagree’;

Experience of harassment and discrimination: prevalence in past year, including reasons.

Local authority services

Provision of high quality services: using 5-point verbalised scale from ‘strongly agree’ to ‘strongly disagree’;

Information and involvement: battery of four questions using 5-point verbalised scale from ‘strongly agree’ to ‘strongly disagree’ on: the council keeping people informed on performance; wanting more information on performance; ability to influence decisions; wanting to influence decisions.

Education and children’s development

Views of the education system: using 11-point numeric scale where 0 is ‘extremely bad’ and 10 is ‘extremely good’;

⁴ Questions are asked of all respondents unless otherwise specified. (A small number are asked of only a random sub-set of respondents – but for the purposes of comparison these count as ‘all respondents’.)

Satisfaction with secondary schools: asked of parents with children aged 11 to 19 at a state secondary school using 5-point verbalised scale from ‘very satisfied’ to ‘very dissatisfied’;

Aspirations for children post compulsory school age: asked of parents with children aged 11 to 15, selecting from a precoded list;

Home learning activities: asked of parents with children aged 3 to 7 – five frequency measures of reading to child; helping to recognise letter, numbers, shapes; help with reading or writing; help with school work.

Health care

Views of health services: using 11-point numeric scale where 0 is ‘extremely bad’ and 10 is ‘extremely good’;

Satisfaction with hospital care: asked of people with a hospital appointment in the past year using 5-point verbalised scale from ‘very satisfied’ to ‘very dissatisfied’;

Satisfaction with GP care: asked of people seeing their GP in the past year using 5-point verbalised scale from ‘very satisfied’ to ‘very dissatisfied’;

Convenience of GP appointments: asked of people seeing their GP in the past year 4-point verbalised scale from ‘very easy’ to ‘very difficult’.

Views on other government policy issues

Satisfaction with the Welsh Government: using 11-point numeric scale where 0 is ‘extremely dissatisfied’ and 10 is ‘extremely satisfied’;

Views of the transport system: using 11-point numeric scale where 0 is ‘extremely bad’ and 10 is ‘extremely good’;

Demographics

Welsh language skills: prevalence of understanding, speaking, reading, writing Welsh;

Educational qualifications: prevalence of different qualification levels, from detailed precoded list;

Financial stability: perception of how well doing financially from a precoded description;

Use of the internet: prevalence of internet usage.

A good number of these questions have provenance, having been developed and tested for other surveys either within Wales or elsewhere. This accounts for the variation in the structure of the response scales used, where the Welsh Government has opted for comparability with other surveys over internal consistency within the National Survey questionnaire. Other questions, from our searches, have no direct comparisons. In these cases, we have usually found questions covering similar topics. However, the differences in context, question structure and wording will mean that, for some questions, only the very broadest comparisons can be made. These issues are discussed in detail in Chapter 5. The full question wording for each question can be found in Chapter 5 and in Appendix 2.

3.3 Comparator surveys

We found potential comparator questions in the following surveys (see Appendix 1 for full details). Those highlighted in bold include one or more questions which are identical (or

sufficiently close to identical) to be able to make direct comparisons with the National Survey⁵ –

Survey ⁶	Approx annual sample size ⁷	Point of comparison		
		Cross-country comparisons	Welsh time series/ comp'ns	Approx sample size in Wales
Annual Population Survey (Labour Force Survey)⁸	333,000	√ UK	√	36,000
British Social Attitudes	3,000	√ GB		
Census	62m	√ England	√	3m
Childcare and out of school activities Survey	6,350	√ England		
Citizenship Survey⁹	10,000	√ England	√	640
Continuous Household Survey	4,000	√ Northern Ireland		
Eurobarometer	1,000	√ Europe		
European Quality of Life Survey	1,500	√ Europe		
European Social Survey	2,400	√ Europe		
GP Access Survey (Wales)	10,000		√	10,000
GP Patient Survey (England)	2m	√ England		
Health Survey for England	4,500	√ England		
Integrated Household Survey	355,000	√ GB/UK	√	18,000
Living in Wales Survey	7,500		√	7,500
Millennium Cohort Study	19,000	√ UK	√	2,500
National Survey for Wales 2009-10 Pilot Survey¹⁰	4,600		√	4,600
ONS Opinions Survey	1,100	√ GB		

⁵ Whether any caveats should be included with these comparisons – because of other methodological differences or the position of the question in the interview – is discussed on a question-by-question basis in Chapter 5.

⁶ Other surveys we looked at which may be useful comparators for National Survey – for questions other than the 20 selected for this review – include the British Crime Survey, Family Resources Study, Life Opportunities Survey, European Union Statistics on income and living conditions, National Travel Survey, National Adult Learning Survey, Welsh Childcare and Early Years Survey and the World Values Survey. The Place Survey would have been a useful source of comparative data in England (e.g. on local authority provision). However, the survey ceased in 2010. Likewise, the Families and Children Study, which had comparable questions, ceased in 2008.

⁷ Or sample size per wave if run less frequently. Sample size per country (with UK counting as one country) in the European surveys.

⁸ The Annual Population Survey is currently part of the Integrated Household Survey. However, given fluctuations in the inclusion of various surveys in the Integrated Household Survey, we include it separately in this list.

⁹ The Citizenship Survey ceased to run in 2011. However, there are plans to launch a new survey which will retain elements from the Citizenship Survey. The exact methodology and interview content are not known, but we include it here – on the basis of the design and content of the Citizenship Survey – as it may be a key source of comparative data for the National Survey.

¹⁰ We include this in the list because it provides a short time series comparison for some questions.

Survey ⁶	Approx annual sample size ⁷	Point of comparison		
		Cross-country comparisons	Welsh time series/comp'ns	Approx sample size in Wales
Oxford Internet Survey (OXIS)	2,000	√ GB		
Public Perceptions of the NHS	1,000	√ England		
Scottish Social Attitudes Survey	1,600	√ Scotland		
Scottish Health Survey	7,200	√ Scotland		
Scottish Household Survey	35,000	√ Scotland		
Understanding Society	45,000	√ UK		2,000
Welsh Health Survey	15,000		√	15,000
Welsh Life and Times Study	880		√	880

In terms of making comparisons between these surveys and the National Survey, the following points are key –

- **Cross-country comparisons:** With the exception of the Annual Population Survey, the surveys carried out in England and Wales; in Britain; in the UK; or in wider Europe do not boost the sample sizes of Wales, Scotland or Northern Ireland. Rather, the number of interviews per country is proportionate to the populations in each respective GB or UK country. As such, only those with very large sample sizes (i.e. the Census, the Integrated Household Survey/Annual Population Survey and the Millennium Cohort Study), can be used alongside the National Survey to compare Wales to Scotland or Northern Ireland. These surveys can be used to compare people living in Wales to the whole GB or UK population (or, sometimes, to other European countries) – or to compare to the English population¹¹. The country-specific surveys in Scotland and Northern Ireland provide the best Welsh:Scottish and Welsh:Northern Irish comparisons.
- **Comparisons within Wales:** All of the surveys identified in the final column as having large enough sample sizes in Wales for 'Wales-only' analysis began prior to the National Survey. All but the National Survey 2009-10 Pilot and the Living in Wales Survey are still on-going studies (some continuous, others fielded every few years) and they may be used alongside the National Survey to look at trends over time. Equally, data collected in the same time period as the National Survey can be compared. In particular, it may be important to understand why findings may differ between the surveys fielding the same or very similar questions.
- **Mode:** All the surveys, with the exception of the Census and the GP Patient Survey, are comparable in terms of mode, conducted face-to-face using computer assisted interviewing¹².

¹¹ It may be appropriate on occasions to aggregate data over a period of time to increase the Welsh sample size. This is the focus of another review (Purdon, 2012) commissioned by the Welsh Government.

¹² A number also contain a short self-completion element.

- **Sample design:** All the Welsh, GB and UK surveys, with the exception of the NHS Perceptions Survey, use random probability sampling. Among the European studies, the European Social Survey has strict rules about random probability sampling. However, for the other Europe-wide surveys, the sample design can vary by country and we are less assured that the sample design is comparable with the National Survey. There are variations in the eligible age range (usually 16+ or 18+) which we highlight in our comparisons. Some surveys focus on a particular sub-group of the population (e.g. parents). However, by selecting the same population among the National Survey respondents, the samples can be made directly comparable¹³. This is further discussed in Section 4.2.
- **Fieldwork period:** When making comparisons across countries (or within Wales at a particular point in time), it is important to compare data collected as closely together as possible, especially for measures which we might expect to change relatively quickly or be affected by external events (see Section 4.3 for further discussion). The ability to do this will vary between surveys: fieldwork for some of the surveys runs continuously, like the National Survey, but others run for a fixed number of weeks per year, or every two years. What we do not know is the precise interview content of forthcoming waves of these surveys (especially those which vary more wave on wave). We therefore provide recommendations on the surveys providing the best comparators as far as we are aware of the fieldwork periods and interview content.

All these issues are further discussed in Chapter 4, on the principles of survey question comparability, and are raised in relation to particular survey questions in Chapter 5.

¹³ There are particular issues with cohort studies such as the Millennium Cohort Study which are discussed in later chapter.

4 Principles of survey question comparability

4.1 Introduction

There are a wide range of reasons why the same or similar question fielded on two different surveys might result in somewhat different findings. There is a body of work which shows that survey findings are sensitive to the context in which they are asked, the population involved and the precise wording of the questions. Unfortunately, identifying reasons, or potential reasons, for differences in responses to two seemingly similar questions can be easier than formally measuring the size of their effect. In some instances, especially where differences are due to sample composition, this measurement is possible. However in most cases, the assessment of how far, and in what direction, the differences between the questions has affected the findings is a matter of judgement¹⁴ – based on what we know from the wider literature about how people respond to different survey situations or question wordings. Of particular use are experimental studies which test different approaches on randomly allocated sets of respondents (e.g. Pudney, 2010).

Of course, the task is easier if we are comparing two identical questions fielded on two or more surveys. The number of potential factors which can impact on comparability are far fewer than when the two questions are similar, but not identical. In the sub-sections below, we therefore begin with issues which can affect the comparability of all survey questions, be they identical or similar. We then turn to the additional considerations required when comparing similar questions. We split the issues into sampling and the eligible population (Section 4.2); survey administration (Section 4.3); and question design (Section 4.4). In Section 4.5 we discuss the relative weight that might be put on these different factors, for different types of questions or comparison requirements.

4.2 Sample design and the eligible population

4.2.1 Sample design and sampling frames

The use of random or stratified random sampling, as adopted in the National Survey, ensures that the population *approached* to be interviewed¹⁵ accurately represents the eligible population (in the case of the National Survey, adults aged 16 and over living in Wales). All surveys adopting a random probability approach produce comparable data in this respect, provided that the sampling frames cover the same eligible population¹⁶. For instance, the Land and Property Services Agency's list of domestic addresses used in the Continuous Household Survey in Northern Ireland is comparable to the small-user Postcode Address File used for the National Survey. In reality, all of the general population surveys

¹⁴ What is more, given we are often assessing a *combination* of immeasurable comparability issues, we cannot assume that finding little or no difference between estimates in different surveys indicates that they are comparable.

¹⁵ Although representativeness can be compromised by non-response (given that differential response rates across subgroups (rather than low response rates in themselves) that lead to bias) (see Section 4.2.2).

¹⁶ The use of different stratification variables should not affect comparability between surveys.

we identified in GB to which the National Survey may get compared to sample from the small-user Postcode Address File, as does the GB fieldwork for the European Social Survey¹⁷.

It is also entirely appropriate to compare two random probability samples using different sampling frames provided that the analysis can be restricted to the same population. For instance, on occasions, it may be useful to compare findings from the National Survey with a survey of parents drawn from Child Benefit records¹⁸ such as the English Childcare and Out of Schools Activities Survey. Provided that the National Survey analysis is restricted to parents with children under 15, the two surveys will be comparable in terms of sample design.

Comparisons between cross-sectional surveys (like the National Survey, in which different respondents are interviewed at each wave) and longitudinal surveys (like Understanding Society, in which the same respondents are interviewed at each wave) are appropriate, provided that cross-sectional weights are applied to the longitudinal survey. However, comparability may be compromised by two factors: attrition in the longitudinal survey (where certain sub-groups of the population drop out of the survey reducing the representativeness of the sample); and the risk of panel conditioning, where the responses of longitudinal survey respondents are affected either by having been asked the question before or by being more used to answering survey questions of this nature.

Comparisons with the National Survey will be compromised if the other survey selects respondents using a quota design. Quota surveys ensure that respondents are broadly representative according to a number of key (usually) socio-demographics, such as age, gender and working status. However, those less easily available and those more reluctant to take part are more likely to be excluded under this design¹⁹. Quota sampling is thought to translate to around one in 10 people approached taking part (although records on non-contacts and refusals are not usually kept), so the chances of ensuring that a quota survey accurately represents the views and experience of the population are low. In contrast, we expect around seven in 10 people sampled for the National Survey to take part in an interview. We would be unsurprised to find differences in responses between the National Survey and a quota survey. It is not possible to measure – or correct in analysis – the ways in which quota surveys under-represent certain sections of the population – because it may often be on less observable factors such as holding particular attitudes. Rather, we recommend avoiding comparisons with quota surveys. In the absence of random probability alternatives (for instance, we suggest in Section 5.6 that some questions on health services could be compared to the NHS Perceptions Survey), any comparative analysis should include a caveat about the differences in sample design.

¹⁷ In the case of both PAF and the Land and Property Services Agency, as the sampling frame is one of addresses, there is further selection stage on the doorstep, when one household member is randomly selected for interview.

¹⁸ Child Benefit records are currently the most complete sampling frame for parents with dependent children. However, from 2013 higher earning parents will no longer be eligible for Child Benefit.

¹⁹ In comparison, with a random probability design, interviewers make several attempts to contact people and to encourage their participation in the survey.

In reality, we identified only two GB quota surveys which included comparator questions for the National Survey questions we were reviewing: the NHS Perceptions Survey and the Ipsos MORI Political Monitor. The sample designs for the pan-European surveys are more complicated to assess, as there can be variation between participating countries. Certainly all countries in the European Social Survey strictly adhere to random probability sampling, and we therefore recommend its use over other European studies wherever possible. To our knowledge, it is the only European-wide survey to use random probability sampling consistently across all countries. Certainly Eurobarometer adopts random route (i.e. quota) designs.

4.2.2 Response rates

While quota surveys provide one extreme in terms of the effect of low ‘response rates’ on sample representativeness – and therefore (lack of) comparability with random probability surveys with higher response rates – we may equally expect differences in response rates between random probability surveys to compromise their comparability, particularly if non-responders are more likely to come from particular sub-groups of the population. So, we might expect the National Survey, if it achieves a 70 per cent response rate similar to the Living in Wales Survey, to better represent the Welsh population than, say, the Welsh Life and Times Survey which achieved a response rate of around 56 per cent.

Where there is a choice of comparator surveys fielding the same question, prioritising the survey with a response rate similar to the National Survey would be the best way of ensuring comparability in terms of sample representativeness. However, the choice of surveys is often very limited (particularly taking account of the issues around survey administration and question design below). So, an alternative approach is to weight the comparative survey data to match the National Survey sample on a range of key variables (present in both surveys). In that way, they will be as closely comparable as possible in terms of sample composition.

Many surveys employ design weights to take account of the sample design (for instance, in PAF samples, correcting for the fact that individuals in single person households have a greater chance of selection than individuals in households with more than one person). Provided these weights are employed during analysis, these should not affect the comparability of the surveys. Some, but not all surveys, also employ non-response weights to re-balance the sample to accurately reflect the demographic profile of the eligible population. Different non-response models between surveys may have a marginal effect on comparability.

4.2.3 Household versus individual surveys

For the National Survey, one adult aged 16 and over per household is randomly selected for interview. Other surveys – such as the Continuous Household Survey, the Health Surveys in each country, the Annual Population Survey, and the first tranche of fieldwork for the National Survey 2009-10 Pilot – attempt to interview all adults in the household.

Statistically, this does not affect the comparability of the surveys²⁰. However, there are two reasons why the ‘all adults in the household’ approach may reduce comparability with surveys like the National Survey –

- Interviewing all adults (and sometimes older children) in the household is a bigger task for families, who may be less likely to agree to take part in the survey as a result. This can lead to lower response rates – an issue which we discuss in Section 4.2.2;
- If household members overhear each others’ interview and decide to answer questions in the same way, responses will be clustered around a household in a way that is not the case where only one person per household is selected for interview. However, the evidence that this happens is limited, and therefore not a major concern in terms of survey comparability.)

4.2.4 Age range and other exclusions

Even among general population surveys, there are differences in the eligible population. The National Survey includes all adults aged 16 and over with no upper age limit. Other surveys include respondents aged 15 and over (e.g. the European Social Survey) and others adults aged 18 and over (e.g. the British Social Attitudes survey). There can be variation in whether people are included if they are not living in private households or if they live in hard to reach areas. Again, provided that the same age range (or population profile) is selected in the analysis, there should be no effect on the comparability of the two surveys²¹.

4.2.5 Concluding comments

A number of the sample design issues which can affect comparability between surveys can be overcome during analysis. Comparability is most compromised between random probability and quota surveys.

4.3 Survey administration

4.3.1 Survey mode

The mode in which the survey is administered is well-documented as having the potential to alter how respondents answer questions. Some of these issues are related to sampling frames available for different modes and the response rates achieved (and therefore relate to the discussions in Section 4.2 above). Face-to-face surveys, like the National Survey usually obtain the highest response rates, with postal and web self-completion usually achieving the lowest. Other ‘mode effects’ are caused by the way in which questions are asked and presented in different modes. For instance, telephone interviews cannot use visual aids such as showcards (Dillman, 2000) and verbalised scale lengths need to be sufficiently short to be remembered. As a result, telephone surveys are more likely to use four rather than five-point scales, or to employ branching (where the scale is split into two

²⁰ One person or all person sample designs have different statistical properties, resulting in different standard errors. However, if appropriately handled, the differences in design should not introduce bias and therefore should not affect comparability.

²¹ Comparisons with birth cohort studies like the Millennium Cohort Study involve restricting the National Survey sample to families with dependent children of the same age (which will vary depending on the cohort fieldwork wave.)

questions²²) (Krosnick and Berent, 1993). But, again, this is an issue we cover below within Section 4.4.3.

However, beyond issues around sampling and survey administration, mode effects are also thought to be caused by the relationship that the respondent has with the interviewer in an interviewer-administered survey versus completing self-completion questionnaires. Giving responses to another person can lead to social desirability or acquiescence bias – that is, for people to respond (either consciously or unconsciously) in a way that is seen as being socially acceptable (i.e. less likely to admit to prejudice, bad habits, etc). It is generally thought that people can feel less comfortable answering questions which focus on personal or sensitive issues, and are therefore more likely to provide socially desirable responses in an interviewer-administered study (Dillman, 2000). Acquiescence bias is thought to be exacerbated when the interview is face-to-face compared to the telephone. We might even expect differences within self-completion surveys – postal, completed in CASI during a longer interview, or online, depending both on people’s feelings of privacy and confidentiality, and how they react to the use and layout of different media. Similarly face-to-face interviewing is viewed as better than other modes (particularly self-completion) at minimising ‘satisficing’ – that is, ensuring respondents make efforts to understand the question and consider their responses (Jackle et al 2008).

We do not attempt to cover mode effects in detail in this report. This is an issue written about in detail in a linked review for the Welsh Government on mixed modes (Nicolaas, 2012). Moreover, almost all of the comparator surveys we identified for the National Survey are fielded face-to-face. The exceptions to this are the Annual Population Survey (which is a mix of face-to-face and telephone), Census (self-completion, either paper or online) and the GP Patients Survey (postal self-completion). A number of studies (e.g. British Social Attitudes survey) include a self-completion element, completed either in the presence of the interviewer or after they have left. It would appear therefore that mode effects will not be one of the major issues in terms of question comparability with the National Survey. Where we do occasionally compare a National Survey question with a self-completion question, we discuss this in Chapter 5. Although there is a strong body of evidence which points to the fact that there are mode effects (and why these happen), assessing the potential impact of mode on the survey findings is a matter of judgement, especially when comparing across two surveys which are different in other respects²³.

4.3.2 Focus of the interview

In a general survey like the National Survey, respondents are asked about a wide range of issues about which they are unlikely to be ‘primed’ in advance of the interview. This is not the case of other potential comparator surveys, which have a clear focus on a particular topic. For instance, we might expect some variation in the profile of people who take part in the Health Survey for England versus those who take part in a general purpose survey such as the National Survey. This first issue can be corrected, to a certain extent, by weighting (although some differences will be less observable such as concern about health

²² Eg Are you satisfied or dissatisfied with.... Is that very satisfied or fairly satisfied?

²³ Controlled experiments have measured the extent to which different modes result in different response distributions (e.g. Jackle et al 2008), but it is not possible to replicate this outside of an experimental setting.

issues). And, likewise, people's responses to a question about their overall health might be different when they are primed to talk about their health (as in the Health Survey for England). This latter issue is less easily measurable, and needs to be included as a caveat when making comparisons between these different types of surveys. The comparator surveys we identified for the 20 National Survey questions on which we focus include a range of general-purpose and topic specific surveys.

4.3.3 Timing of the survey

Any assessment of the comparability of surveys needs to take account of the timing of the data collection (unless the interest is in change over time). If surveys have different fieldwork periods, differences in responses, particularly to attitudinal questions, may be a reflection of respondents' awareness of pertinent events happening locally or nationally, the position in the economic cycle or seasonal effects. So, if there is a choice to be made between surveys with identical questions, the survey with the closest fieldwork period to the National Survey should be taken. (However, in reality, this decision would need to be taken in the context of other comparability issues, as discussed in Section 4.5.)

The extent to which this is an issue varies depending on the focus of the question: it is dependent on how quickly we might expect a particular measure to change over time, and the extent to which it might be affected by external factors, such as major events, changes in government administrations, and so on. The more affected a question is likely to be, the more important it is to ensure that comparisons are made between data collected at similar points in time. When there has been a one-off event which may change people's attitudes to a particular question overnight (e.g. as the MPs expenses scandal did to levels of trust in politicians), it is probably appropriate to restrict analysis to National Survey data collected over an exactly the same time period to the comparator surveys (e.g. British Social Attitudes survey is conducted between June and September, so could be compared to National Survey data collected over the same four-month period).

There is a wealth of evidence from the British Social Attitudes survey and other surveys about fluctuating trends in attitudes depending on these issues. Indeed, for the European Social Survey, the research organisation in each country records 'major events' which are used to understand fluctuations over time.

4.3.4 Concluding comments

Our review of other similar surveys suggests that the Welsh Government will usually be comparing the National Survey with other face-to-face computer-assisted surveys, minimising the need to take account of mode effects between it and its comparator surveys. However, there will be issues to take into account in terms of data collection periods and survey foci.

4.4 Question design

4.4.1 Question order, and position in the interview

The way in which people respond to certain questions can be affected by previous questions they have been asked in the survey (e.g. Schuman and Presser, 1981). In our review of 20

questions in the National Survey, we found several examples where the same, or a very similar question, is fielded in another survey, but its position in relation to other pertinent questions makes comparisons between the two surveys less than perfect. For instance, the National Survey and the European Social Survey include an identical question about public satisfaction with the Government. However, the European Social Survey asks this question just after a series of questions on political affiliation and, importantly, views on the economy. We may well expect respondents, having just focused on their views on the previous questions, to answer about their views of the Government in a somewhat different light to in the National Survey (where the question is immediately preceded by questions on local authority provision). Similarly, while the National Survey fields a question asking respondents to rate their own health at the start of a series of questions about health care, the British Social Attitudes survey series asks this question just after asking about people's experience of the NHS.

Unfortunately, these kinds of order effects cannot be formally measured, or controlled for in analysis. Rather, they are issues to be aware of when making comparisons between surveys (and considered during survey development where there are definite plans to compare with another survey).

Another issue to watch in the comparison of questions is whether the questions are asked in a standard or randomised order. Some surveys will randomise a list (or 'battery') of questions using the same response scale (e.g. a satisfaction or agree to disagree scale) in order to avoid any order effects within the survey. We found one example among the 20 survey questions where, although the questions (on local authority provision) were identical to those in the Scottish Household Survey, the Scottish survey had randomised the order in which they were asked, but the National Survey had not (see Section 5.4.3).

4.4.2 Question phrasing

The issues in Sections 4.4.2 and 4.4.3 apply only when comparing similar, but not identical, questions and response categories.

Sometimes two similarly, but not identically phrased, questions can elicit different responses. The impact of minor differences in wording is necessarily assessed on a case by case basis (several examples are discussed in Chapter 5). Here we pick up on two particular issues which are common reasons why seemingly very similar questions are in fact different, and likely to elicit different responses -

- The way in which attitudinal, Likert-style scales are introduced in the question text can alter the distribution of responses. For instance, even when the same response scale is used, some questions introduce the start of the scale (e.g. 'how satisfied are you') while others introduce both ends (e.g. 'how satisfied or dissatisfied are you'). The latter arguably implies that both views are equally valid, and thereby reduces acquiescence bias.
- Similarly, it is important to check the context in which questions are asked, including the use of introductory text. The National Survey includes two or three questions (e.g. views on health care and on the education system) which start with an introduction on what comes within the definition of health care or education. These introductions were included after cognitive testing found that respondents did not necessarily take account

of all aspects of the services in question, more often thinking mainly or solely about their own experiences. For example for education, they thought about primary and secondary schools and not nurseries or universities. The additional text should increase the validity of the questions. However, it reduces the direct comparability with other surveys (in this case the European Social Survey) fielding identical questions but without the introductory text (see Sections 5.5.2 and 5.5.2).

Again, these kinds of order effects cannot be formally measured, or controlled for in analysis. Rather, they are issues to be aware of when making comparisons between surveys.

4.4.3 Response scales

Very much linked to the question wording, the precise wording and format of the response scales can affect how people respond to particular questions, and thus the comparability of questions using different response scales. There are several examples of this in Chapter 5. Here we do not focus on the appropriateness of different types of scales (on which there is a wide literature) but rather on the potential effect of different scale lengths and designs, and what, if anything can be done to increase the comparability between different scale lengths.

There are a number of factors to consider –

- *The length of the scale:* The National Survey includes a mix of shorter four or five-point scales and longer 11-point scales (chosen to ensure comparability with particular questions asked in other surveys). There are several issues which mean that it is problematic to compare these with similar questions on other surveys which have different scale-lengths. On occasions, it may be possible to group longer scales into smaller categories to make them comparable with shorter scales, but this depends on a number of other factors raised in the bullet points below.
- *The use of a mid-point:* Because there are debates around whether it is best to provide respondents with the opportunity to give a ‘neutral’ mid-point reply (e.g. neither agree nor disagree) or whether it is best to avoid satisficing by ‘forcing’ respondents into giving a reply, different survey researchers make different decisions during survey design. Overall, researchers tend to favour mid-point scales in face-to-face surveys, while questions designed for use over the phone are more likely to choose four-point scales (on the basis that five or more points are more difficult to remember, and three points is rather short). The use or otherwise of a mid-point necessarily alters the distribution of responses and makes any comparisons between odd and even length scales problematic: it is not possible to predict how people would have responded given the presence/absence of mid-point. Only one of the 20 National Survey questions we reviewed used a four-point scale (on GP appointments). Most of the comparator surveys also used scales with mid-points, although there are a small number of examples which we discuss in Chapter 5.
- *Verbal or numeric scales:* The choice of whether to use a verbal scale (i.e. all the points have a description) or a numeric scale with descriptions only for the end points tends to be correlated with the length of the scale, with scales longer than five using the latter approach. Moreover, cross-national surveys, like the European

Social Survey, favour numeric scales because of the difficulty in accurately translating the nuanced differences between verbalised scale ratings. The comparison of such scales is problematic. We cannot say with any assurance where people who would say they were 'fairly dissatisfied' (i.e. fourth point on a five-point scale) would place themselves on a scale of 0 to 10 where 0 is 'very dissatisfied' and 10 is 'very satisfied'. 'Fairly dissatisfied' could potentially equate to 1, 2, 3 or 4 on the 0 to 10 scale. However, this may be an issue for further analysis once data from the National Survey are available. Note that most verbalised scales run from positive to negative (e.g. very satisfied to very dissatisfied), on the basis that this is cognitively easier for respondents. However, questions with numeric scales tend to run from negative to positive, in order to more naturally match the numeric score (i.e. it makes intuitive sense for 'extremely bad' to score 0 and 'extremely good' to score 10). The direction of the scale may have some effect on the distribution of responses, although this is not easily measured given the associated differences in the scale properties.

- *Extremity of scale ends:* The end points of longer numeric scales are often fixed as 'extremely x' rather than 'very x' (the usual approach for verbalised scales). Again, this changes the distribution of responses (with respondents less likely to choose either end point).
- *Scale presentation:* There is a body of literature about the presentation of scales, presenting the case for horizontal and/or vertical designs. There is also variation in preferences for the use of numbers on showcards (e.g. 1. Strongly agree 2. Tend to agree, etc versus Strongly agree Tend to agree, etc), with some arguing that the use of numbers makes respondents more passive and less likely to concentrate on the full scale. While minor differences in question wording on the face of it, these should be taken into account in any question comparison using showcards or administered as self-completion. Similarly, if the comparison is with a telephone survey, we must take account of the fact that telephone respondents cannot visualise the full scale in the same way as other respondents, and that, to accommodate the telephone approach, some surveys adopt 'branching' where scales are split into two or more questions (a recent experiment of the effect of this is discussed in Nicolaas et al, 2011).
- *Offering 'don't know' as an option:* Surveys vary in whether they offer 'don't know' as a valid response option (with self-completion surveys being more likely to do so). While we know that there are a greater number of 'don't knows' when the option is explicitly offered, it is hard to predict how respondents would have answered in the absence of the 'don't know' option. On some occasions, 'don't know' responses are combined with the middle option (e.g. neither agree nor disagree). Again, this is an issue for consideration in any question wording comparison (although most of the comparator questions we found for the 20 questions used the same approach as the National Survey).

4.4.4 Concluding comments

This section arguably highlights the most difficult challenges for the Welsh Government when assessing which questions provide suitable comparisons with the National Survey. Our review of 20 questions identified several ways in which the position or wording of the survey questions compromised comparability, in different degrees. This is an issue we

return to in Chapter 6. These issues are also useful when considering questionnaire development for future waves of the National Survey, particularly where there is a need to be able to make robust comparisons with other surveys.

4.5 Summary of the key considerations for the National Survey

In the sections above, we have highlighted a wide range of factors which can affect the comparability of survey questions. In this final section of Chapter 4, we provide some advice on the relative weight that should be given to each of these factors when deciding if and when a question provides an appropriate comparison for the National Survey. Because the issues will vary depending on the nature of the question (e.g. attitude questions will be more affected by their position in the interview than a factual measure), we cannot provide a definitive 'decision-tree' for deciding on the most appropriate comparisons. Rather, we briefly raise a few key issues which, in combination with the review of the example questions in Chapter 5, should be helpful when making decisions on other questions within the National Survey or other surveys.

Firstly, although our review only looked at surveys fielding questions comparable to 20 of the National Survey questions, it appears that most comparable surveys are face-to-face interviewer administered surveys like the National Survey and that, therefore, **in most cases mode effects will not be an issue**. Likewise, **the majority of surveys use comparable sampling methods**. There are some **differences in response rates**, but these **can be taken into account in analysis, at least to some extent**.

So, the primary issues to take into account in the assessment of the comparability of the questions are –

- Survey focus
- Fieldwork period
- Position of the question in the interview
- Question phrasing
- Response scale

While it is impossible to measure the effect of each of these (particularly as we are often looking at a combination of more than one comparability issue), our view is that, as a general rule for all questions, **most attention should be paid to ensuring close comparability with the question wording and response scale**. **The focus of the survey and position of the question in the interview are important for questions which might arguably be affected by these issues** – so subjective questions more than objective measures of behaviour, experience or socio-demographics. And the **fieldwork period is especially important for variables for which we may expect fairly rapid changes** over time.

5 A comparison of the 20 National Survey questions

5.1 Introduction

In this chapter, we focus on the 20 questions from the current National Survey interview introduced in Section 3.2. We provide details of comparable questions in other surveys (full details of which are in Appendix 2). We advise on how and when it is appropriate to compare findings from the National Survey with these comparator surveys. We provide the necessary caveats when making these comparisons, as well as explanations for why we may expect differences in the distribution of responses between the National Survey questions and their close comparators in other surveys. We organise the 20 questions into the groups outlined in Section 3.2, and divide our discussions into (a) making cross-country comparisons and (b) making comparisons ‘within-Wales’ either over time or in the same timeframe.

It is important to note that we cannot always be sure that the comparator questions will continue to be fielded (in exactly the same format) in the comparator surveys, or of the frequency of fielding. This is especially the case for surveys which vary their content more regularly (e.g. British Social Attitudes Survey). We have written this report on the assumption that questions, and their frequency of fielding, remain similar to now. However, this will need to be kept under review. Also, as raised in Section 3.2, our discussion around the potential questions to which the National Survey questions will get compared assumes no aggregation of time series data (e.g. to increase the Welsh sample size in GB or UK based studies). Thus, when looking at the extent to which the National Surveys can be compared to other Welsh data from the same or similar questions, we focus on studies with a Welsh sample size which is sufficiently large to be used without the aggregation of waves (see Section 3.2 and Appendix 1). Lastly, this chapter is based on an assessment of the questions (and the originating survey design) in the absence of any National Survey data with which to test out various assumptions. This is something to which we return in Chapter 6.

5.2 Health and well-being

5.2.1 Overview

The National Survey includes two questions – one on self-perceived health and one on life satisfaction - which are frequently used, both in general-purpose population surveys and in surveys and evaluations which focus on specific population groups and particular policy areas. The National Survey data on these two measures can be used as a baseline measure for the Welsh population, and a time series tracking perceptions of health over time. Equally, the data will be valuable in positioning Wales among other GB, UK or European countries. A review of other general population surveys found that there are a number of very similarly structured questions on self-perceived health and life satisfaction. We provide recommendations as to which surveys would provide the best comparisons – within

Wales and cross-nationally. We also highlight issues which make the data from some surveys less comparable than others.

5.2.2 Self-perceived health

Q45		How is your health in general; is it...	
d	READ OUT. SINGLE CODE ONLY		
A	Very good	1	GO TO Q46
B	Good	2	GO TO Q46
C	Fair	3	GO TO Q46
D	Bad	4	GO TO Q46
E	Very bad	5	GO TO Q46
	Don't know (SPONTANEOUS ONLY)	-9	GO TO Q46
	Refused	-99	GO TO Q46

The following surveys provide directly comparable data with the National Survey -

- ❖ The Integrated Household Survey
- ❖ The European Social Survey
- ❖ ONS Opinions Survey
- ❖ Continuous Household Survey
- ❖ Citizenship Survey

Comparisons with other GB / UK countries

The question on the National Survey is directly comparable to a number of other surveys run in GB or UK countries. The following surveys all use comparable sampling methods and field the question on self-perceived health as a face-to-face question -

- *UK:* Integrated Household Survey
- *UK:* European Social Survey (see below)
- *Britain:* ONS Opinions Survey
- *England and Wales:* Citizenship Survey²⁴
- *England only:* Health Survey for England²⁵
- *Scotland only:* Scottish Health Survey
- *Northern Ireland only:* Continuous Household Survey

Given the sample size of the Integrated Household Survey (around 45,000 individuals in the UK), we recommend its use as the main source of comparison for other UK countries. However, all of the surveys in the above list provide comparability. An assessment of the position of the question in the surveys' interviews does not suggest that responses will be particularly affected by preceding questions. However, as the focus of the Health Surveys *may* influence how people respond to this question, the other, more general-purpose, surveys may provide the best comparisons²⁶.

²⁴ For time series and, potentially, ongoing if the replacement survey is comparable.

²⁵ When it was running, the Citizenship Survey fielded an identical question in England and Wales. The new survey may therefore contain the same question – in which case providing comparable data which matches the National Survey fieldwork.

²⁶ Analysis of the IHS needs to restrict the population to age 16+ to match the National Survey.

Note that the British Social Attitudes Survey series, Understanding Society and the 2011 Census all field *similar* questions. However, they do not provide as good a GB/UK comparison as the surveys above.

British Social Attitudes includes a question which –

- asks respondents to consider their health in the context of ‘someone of their age’, which is a substantively different question to the more standard question;
- immediately follows a series of questions about attitudes towards NHS health care, which may, arguably, alter their perceptions of their own health. (In the National Survey, the question precedes questions on attitudes to the health services.)

Understanding Society uses a response scale which is different to the National Survey (running from ‘excellent’ to ‘poor’ rather than ‘very good’ to ‘very bad’) (see Section 4.4.3 for a discussion of the potential effect of different scales).

The 2011 Census uses a question which is identical to the National Survey. However, it is self completion rather than administered by interviewers (see Section 4.3.1 for a discussion of mode effects). There is evidence from the Health Survey for England (Tipping et al, 2008) that respondents answering this question are more likely to rate their health as good if asked the question face-to-face rather than as a self-completion question. The authors suggest that the differences may be accounted for either by social desirability bias (where respondents are more likely to give a favourable answer to the interviewer) or by different interpretations of the scale depending on whether it is heard or read. This previous study would suggest that any comparisons between the National Survey and the Census of this question should be made with caution.

Comparisons with other European countries

The best pan-European comparator survey on self-perceived health is the European Social Survey²⁷. As with the UK surveys above, it fields precisely the same question as the National Survey²⁸, with the same scale, and employs the same mode and comparable sampling strategies. In both surveys, the questions are fielded part way through the survey, with none of the immediately preceding questions focusing on issues which we might expect to produce order effects.

Comparisons within Wales

The Integrated Household Survey provides the best data for a short-term time series for Wales (since 2009). The Citizenship Survey could also be used. Neither Living in Wales nor the Welsh Health Survey provide data that can be used in conjunction with the National Survey to measure changes in self-perceived health over time. The Living in Wales survey did not include a question on respondents’ perceptions of their own health. The Welsh Health Survey uses a question which is worded a little different and – most importantly – uses a different scale (running, like in Understanding Society, from ‘excellent’ to ‘poor’

²⁷ Analysis of the European Social Survey needs to restrict the population to age 16+ to match the National Survey.

²⁸ Although the European Social Survey uses the phrase ‘Would you say it is...’ in comparison with the National Survey’s ‘Is it...’.

rather than ‘very good’ to ‘very bad’)²⁹. In addition, the question on the Welsh Health Survey is preceded by a set of questions about recent ailments: we might confidently surmise that this may affect how people respond to a question about how they rate their overall health.

5.2.3 Overall life satisfaction

Q13 Overall, how satisfied are you with your life nowadays?

d Interviewer instruction: READ OUT IF NECESSARY ‘where nought is ‘not at all satisfied’ and 10 is ‘completely satisfied’

0-10 scale		GO TO Q14
Don't know (SPONTANEOUS ONLY)	-9	GO TO Q14
Refused	-99	GO TO Q14

- ❖ The Integrated Household Survey provides directly comparable data with the National Survey.
- ❖ The European Social Survey provides a *reasonably* close pan-European comparison.

Comparisons with other GB or UK countries

Again the Integrated Household Survey³⁰ provides the best comparison with this National Survey question. It is fielded identically, and the sample design and mode are comparable³¹.

We also identified a range of other general population surveys that include this question, or a very similar one. However, none of these other surveys are *precisely* comparable with the National Survey – either the question is slightly different, or the scale length or composition varies, or the survey mode. On balance, we view the ONS Opinions Survey, the European Social Survey and the Scottish Health Survey as providing cross-national comparisons about which we can have a *fair* degree of, but by no means absolute, confidence.

- **The ONS Opinions Survey** includes an identical question to the National Survey – including both the question wording and the scale composition. Its sample size means that it can be used alongside the National Survey to make comparisons with England, or with GB as a whole. However, the life satisfaction on the Opinions Survey is asked as a self-completion question (on the computer). While there is no particular reason to think that respondents will answer this question very differently in different modes, there is nonetheless a risk that they will do so. Moreover, there is evidence that people may use response scales differently if they are seen visually (in the case of the Opinions Survey) in comparison to a ‘read out’ question.
- As a comparator to the National Survey question on life satisfaction, **the European Social Survey’s**³² question has the advantage of being fielded as a face-to-face survey,

²⁹ The Welsh Health Survey may provide the most robust time series on self-perceived health.

³⁰ We include the Annual Population Survey within this.

³¹ As mentioned before, analysis of the IHS data needs to be restricted to the 16+ population.

³² Analysis of the European Social Survey needs to be restricted to the 16+ population.

using comparable sampling methods. It is certainly the best comparator question for European countries outside of the UK. However, any comparisons need to be made with the following cautions –

- The European Social Survey provides respondents with a showcard – so the scale is provided visually as well as orally. As mentioned above, visual aids can help respondents and, therefore, questions using showcards cannot be directly compared to those without them;
 - The scale ends in the European Social Survey are more ‘extreme’ than those in the National Survey. Respondents are asked to view a score of 0 as being ‘extremely dissatisfied’ while 10 is ‘extremely satisfied’. In comparison, the scale in the National Survey question runs from ‘not at all satisfied’ to ‘completely satisfied’. Research has highlighted the fact that fewer people will choose the end points under the European Social Survey model, so direct comparability between the two surveys will be difficult. If one were to group the scale into three or four bands this might make comparison more accurate.
 - In the National Survey, the question on life satisfaction is placed early in the survey interview, with little chance that people’s responses will be affected by preceding questions. In comparison, the life satisfaction on the European Social Survey comes after a set of questions on political trust.
- **The Scottish Health Survey and Scottish Social Attitudes Survey** mirror the European Social Survey in terms of the response scale and the use of showcards, so the same caveats apply. Moreover, in the Scottish Health Survey, the question is preceded by questions on the respondents’ general health which we may expect to have an effect on people’s responses on their overall life satisfaction. Likewise in the Scottish Social Attitudes Survey it follows other questions on satisfaction with standards of living.

A number of other surveys, such as British Social Attitudes and the European Quality of Life Survey, include very similar questions, but use a scale from 1 to 10 (instead of 0 to 10). The use of a scale with an even number of response categories (i.e. without a mid-point) makes direct comparison difficult (see Section 4.4.3).

Comparisons within Wales

Note that this question has not been fielded in a previous Welsh national study. The National Survey 2009-10 Pilot included a life satisfaction question using a five-point scale. The Integrated Household Survey provides a short time series from 2009.

As this question is likely to be used in an increasing number of surveys over time, it is important to check for these issues (particularly scale length and end points and position in the interview) when making comparison.

5.3 Living in the community

5.3.1 Overview

We reviewed two National Survey questions on people’s experiences of living in their local or wider community: the extent to which people feel they belong to their local community

(defined as within 15 to 20 minutes walking distance from home) and experience of harassment, discrimination or abuse. We found very few questions which could be used to make even the broadest comparisons with other UK or European countries, or over time within Wales.

5.3.2 Feelings of belonging

CORE SHOWCARD B To what extent would you agree or disagree that you belong to your local area?		
d	SINGLE CODE ONLY	
A	Strongly agree	1
B	Tend to agree	2
C	Neither agree nor disagree	3
D	Tend to disagree	4
E	Strongly disagree	5
	Don't know/No opinion (SPONTANEOUS ONLY)	-9
	Refused	-99
D		

❖ We found no surveys which provide sufficiently comparable data with the National Survey for robust analyses.

Comparisons with other GB or UK countries

We found two surveys which have the potential to provide similar, but by no means identical, data for England and (in one survey) Scotland. However, even with up to date data, they could be used to make only the very broadest comparisons with the National Survey –

- **The Citizenship Survey** included questions on ‘how strongly you belong’ to ‘your immediate neighbourhood’ and ‘your local area’, using a scale running from ‘very strongly’ to ‘not at all strongly’. If the new survey to be commissioned later this year includes these questions they provide a broad comparison with the National Survey. If the new survey has a sufficiently large sample size in Wales, it would be worth conducting some analysis of the ‘local area’ question to test the extent to which these two questions produce a similar ‘headline’ story on community cohesion.
- **The British Social Attitudes survey series** has previously run a question asking about ‘a sense of attachment and belonging...to your local area’, using a scale running from ‘very closely’ to ‘not at all closely’. It provides very loose comparability. Moreover, this question has not been fielded since 2003, so would only be useful if there were plans to field again.

Comparisons with other European countries

Again, there are data only to make the very broadest comparisons between Wales (using the National Survey question) and other European countries. Eurobarometer has asked how attached people feel to their ‘town or village’, using a scale from ‘very attached’ to ‘not at all attached’. This question was last fielded in 2007, and we do not know about future

plans for fielding it. The random route sample design for Eurobarometer also reduces its comparability with the National Survey.

Comparisons within Wales

The current question was adapted from a question in the Living in Wales survey, which asked about ‘your neighbourhood’ rather than ‘your local area’. As such, the Living in Wales survey provides the best data for comparisons over time³³. However, we know from other developmental work that different respondents can make quite different judgements as to what counts as their local area or their neighbourhood. Indeed, the Citizenship Survey had two questions which distinguished between ‘your immediate neighbourhood’ and ‘your local area’ with the latter defined as within a 15-20 minute walk from home. The National Survey attempts to minimise between-respondent variation by defining the local area as within 15 to 20 minutes’ walk from home. Conversely, the Living in Wales survey prefaces the questions on neighbourhood by talking about ‘the area where you live’. So, we may expect some changes in the distribution of responses due to this wording change between the two surveys: something that would usefully be assessed once sufficient data from the National Survey are available. That said, within Living in Wales the question follows a series of questions about feelings of safety in their neighbourhood (as opposed to questions about access to local amenities in the National Survey). We might expect these preceding questions to have some effect on how respondents answer about belonging to their neighbourhood.

5.3.3 Experience of harassment and discrimination

Q24		SHOWCARD 5 In the last 12 months have you personally experienced discrimination, harassment or abuse for any of these reasons? Please just tell me the letter next to the reasons that apply.	
d	MULTICODE. CAPI TO ONLY HAVE LETTERS FOR PRIVACY		
A	Your accent	1	
B	Your ethnicity	2	
C	Your age	3	
D	Your language	4	
E	Your colour	5	
F	Your nationality	6	
G	Your mental ill health	7	
H	Any other health problem or disability	8	
I	Your sex	9	
J	Your religious belief or faith	10	
K	Your sexual orientation	11	
L	Where you live	12	
M	Other (please specify only if you wish to)	13	
N	I have not experienced this	14	GO TO Q26
	Don't know (SPONTANEOUS ONLY)	-9	GO TO Q26
	Refused	-99	GO TO Q26

❖ We found no surveys which provide sufficiently comparable data with the National Survey for robust analyses.

³³ This question was also fielded in the National Survey 2009-10 Pilot.

Comparisons with other countries

There are no questions which provide a direct comparison with the National Survey question above, and the only reasonably close comparator question is fielded only in Scotland, in the Scottish Household Survey.

The Scottish Household Survey provides respondents with a showcard showing exactly the same list of issues as the National Survey question (above). However, it asks two separate questions: the first about 'being unfairly treated or discriminated against' and the second about 'experiencing harassment or abuse'. Although any comparison between Scotland and Wales could combine the responses to the two Scottish questions to make it fairly comparable to the Welsh question, there are nonetheless wording differences which may alter people's propensity to report an issue. In particular, the term 'unfairly treated' is a milder term than the other three: its absence from the Welsh question may lead to the under-reporting of incidences in comparison with the Scottish question.

In terms of comparisons with England, **the Citizenship Survey** included a question on harassment due to skin colour, ethnic origin or religion. While it may be possible to derive the proportion of people in England versus Wales experiencing discrimination due to these three issues, there are other inconsistencies in the question which mean that the data are insufficiently comparable to draw any firm conclusions: the Citizenship Survey asks about the past two years (rather than 12 months) and restricts incidents to the local area.

Eurobarometer includes a question which measures the prevalence of discrimination or harassment on the grounds of seven of the issues included in the National Survey question. Very broad comparisons (given the difference in question wording, and variation in sampling methods) could be made with Wales and other European countries.

Comparisons within Wales

The National Survey will form the start of a new time series on discrimination in Wales.

The **Living in Wales Survey** included a question which asked about the incidence of discrimination, harassment or victimisation. However, it is incomparable with the National Survey on a number of dimensions –

- It uses the term victimisation rather than abuse, which may be interpreted differently;
- It asks about the past five years;
- It asks a dichotomous question, and then follows up with an unprompted question about the reason for the discrimination/harassment/victimisation. This unprompted approach is likely to under-report incidents in comparison with the National Survey showcard approach.

The **National Survey 2009-10 Pilot** included a question which asked about discrimination and harassment over the past year about a sub-set of the similar (but not identical) dimensions included in the National Survey question. However, again, certain factors make it incomparable with National Survey –

- It was included a self-completion question. Given the sensitivity of the issue, this may well alter the prevalence of people reporting incidences;
- The question included 'violence' rather than just harassment and discrimination.

5.4 Local authority services

5.4.1 Overview

The National Survey includes a set of questions on people's attitudes towards the services provided by their local authority. They are drawn from a set of questions fielded in the Scottish Household Survey, enabling good comparisons with Scotland on these issues. Questions fielded in England and previously within Wales over time are less directly comparable.

5.4.2 Provision of high quality services

I'm now going to ask you a few questions about the services provided by [name of local authority]. [name of local authority] runs services including street lighting, road maintenance, parks and leisure facilities, housing, refuse collection, and recycling.

Q83 CORE SHOWCARD B To what extent do you agree or disagree with the following statement: my council provides high quality services.

d			
A	Strongly agree	1	GO TO Q84
B	Tend to agree	2	GO TO Q84
C	Neither agree nor disagree	3	GO TO Q84
D	Tend to disagree	4	GO TO Q84
E	Strongly disagree	5	GO TO Q84
	Don't know/No opinion (SPONTANEOUS ONLY)	-9	GO TO Q84
	Refused	-99	GO TO Q84

❖ The Scottish Household Survey provides very closely comparable data with the National Survey.

Comparisons with other GB or UK countries

Having identical questions in the National Survey and the Scottish Household Survey means that comparisons can be made with confidence between the views of people in Scotland and Wales on local authority provision. However, two caveats should be highlighted, both of which may have a minor (but we would not expect major) effect on the distribution of responses -

- The National Survey precedes its questions on local authority services with a clarification of what services people should take into account when answering the question. This was added after cognitive testing found that people did not necessarily take account of the full set of relevant services. The National Survey question is strengthened as a result, despite slightly reducing the comparability with Scotland;
- The Scottish Household Survey includes this question among a battery of questions (which include three of the four questions on information and involvement in the following section), the order of which varies randomly between respondents. The randomisation will have been introduced in Scotland in order to mitigate against any order effects among the questions – something which may therefore be an issue in the National Survey. If this was thought to be an issue in Wales, and the Welsh Government wanted to improve the exact comparability between the data from the two countries,

this question could be placed alongside the four questions below and asked of respondents in a random order.

From our review, it appears that comparisons with England are problematic. Both the Place Survey and the Citizenship Survey included questions measuring people’s satisfaction with how the local authority ‘runs things’. However, neither survey is now running. The new survey to replace the Citizenship Survey may include this question – although it would only provide a comparison with the National Survey question in the broadest sense. Previous waves of the Citizenship Survey provide the only Welsh data we found on people’s views of their local authority provision – although the differences in the question and the small sample size make this of limited value. We found no pan-European surveys with relevant questions.

5.4.3 Information and involvement

Q86 CORE SHOWCARD B I am going to read out four statements which might be used to describe things [name of local authority] does. For each of these please tell me to what extent you agree or disagree that it applies to [name of local authority].

My council is good at letting local people know how well it is performing.

I would like more information on how my council is performing.

I can influence decisions affecting my local area.

I would like to be more involved in the decisions my council makes that affect my local area.

D			
A	Strongly agree	1	GO TO Q87
B	Tend to agree	2	GO TO Q87
C	Neither agree nor disagree	3	GO TO Q87
D	Tend to disagree	4	GO TO Q87
E	Strongly disagree	5	GO TO Q87
	Don't know/No opinion (SPONTANEOUS ONLY)	-9	GO TO Q87
	Refused	-99	GO TO Q87

❖ The Scottish Household Survey and Continuous Household Survey provide very closely comparable data with the National Survey.

Comparisons with other GB or UK countries

Three of these four questions (‘I would like more information on how my council is performing’ is not included) form part of a longer battery on the Scottish Household Survey on public attitudes towards being involved and informed about local services (together with the service quality question in Section 5.4.2 above). As such there is directly comparable data with Scotland, apart from the caveats above³⁴.

³⁴ The Scottish Health Survey also includes a single comparable question on the ability to influence decisions affecting the local area.

The **Continuous Household Survey** has previously (though not in the latest round) fielded almost the same question on people's ability to influence decisions that affect their area, using the same response scale, providing a pretty close comparison in Northern Ireland on that one question³⁵.

Unfortunately, again, there are few data for England, despite the National Survey question on people's ability to influence decisions affecting the local area being an ONS harmonised question³⁶:

- The Place Survey and Citizenship Survey did field this same question about people's ability to influence decisions affecting their local area. However, neither used the same response scale, but rather used a four-point scale from 'definitely agree' to 'definitely disagree'. So again, the new survey replacing the Citizenship Survey may provide this partially comparable data for England.
- Similarly the Citizenship Survey (and therefore potentially the replacement survey) and the Place Survey asked a trichotomous "yes/no/depends on the issue" question about whether they would like to be more involved in decisions that affect the local area.

Comparisons within Wales

The Living in Wales Survey included a question on whether people felt that they were kept informed of the performance of the local authority's services. Although this used a dichotomous yes/no response, it would be appropriate to compare its results *in broad terms* to the National Survey question on how well the council is at keeping people informed of its performance.

5.5 Education and children's development

5.5.1 Overview

We have reviewed two questions which measure people's satisfaction with education in Wales, as well as a question asking parents about their aspirations for their children post-compulsory education and a series of questions which collect data on what parents do with younger children at home to help with their development. The opportunities for comparative work are greatest with the first question on overall satisfaction.

³⁵ There may be some effect of being immediately preceded by questions about membership of organisations and feelings of local safety.

³⁶ ONS leads a cross-governmental programme of work to standardise questions for use in National Statistics. The aim is to make it easier for users to draw clearer and more robust comparisons between data sources: <http://www.ons.gov.uk/ons/guide-method/harmonisation/harmonisation-index-page/index.html>

5.5.2 Satisfaction with the education system

The Welsh Government is interested in your views about education in Wales. When answering these questions please consider all aspects of the education system, from nursery schools through to universities as well as adult education courses. For this question I'd like you to give an answer on a scale of nought to ten, where nought is 'extremely bad' and ten is 'extremely good'.

Q89 SHOWCARD 14 Please say what you think overall about the state of education in Wales nowadays?

d SINGLE CODE ONLY. REMINDER: Answer is to be given on a scale of nought to ten, where nought is 'extremely bad' and ten is 'extremely good'.

0-10 scale		GO TO Q90
Don't know (SPONTANEOUS ONLY)	-9	GO TO Q90
Refused	-99	GO TO Q90
Section not applicable	-88	GO TO Q90

❖ The European Social Survey provides very closely comparable data with the National Survey.

Comparisons with other GB or UK countries and with wider Europe

This question is taken directly from the European Social Survey, which is therefore the best source of comparison data from England³⁷ and other European countries. There are two caveats that should be given where comparing this question in the National Survey and the European Social Survey, the first of which is the more important -

- The National Survey question is prefaced with a clarification about what counts as 'education' for the purposes of the question. It prompts people to think about the full range of education provision, including nursery and adult learning. This introduction was added in the light of cognitive testing which showed that some respondents did not take account of the full range of educational services when answering the question. We might reasonably surmise that, in the absence of the introduction, the European Social Survey respondents think about a more narrow range of provision (probably that closest to their own experience). It is hard to predict how the inclusion of the introduction will affect the distribution of responses, but this could be explored further once data from the National Survey are available.
- In the National Survey, the question comes relatively late in the interview – after a series of questions asking parents about their views of the education of their children. It is possible that having been recently asked to focus their thoughts on these experiences, their response to the overall satisfaction question might be affected. Again, the size and type of effect is not measurable.

Two other studies include similar, but not identical questions on public attitudes towards the education system –

³⁷ The sample sizes for Scotland and Northern Ireland are small. However, comparisons between Wales and the rest of the UK would be possible. Analysis of the European Social Survey would need to be restricted to respondents aged 16+ to be comparable with the National Survey.

- **The British Social Attitudes survey** series has previously fielded³⁸ a question on satisfaction with the ‘quality of education in Britain’s state schools,’ and a separate question asking how good or bad state education is. Neither question provides direct comparability. The first question (a) focuses on ‘quality’ and, importantly, on schools, rather than on education more widely and (b) uses a five-point verbalised satisfaction scale, rather than the 11-point ‘good to bad’ scale used in the National Survey. The second question (a) again focuses only on schools rather than the education system more widely and (b) measures ‘Very good to very bad’ using a five-point verbalised scale with less extreme end points than the National Survey.
- **The European Quality of Life Survey** included a question on the quality of the education system – again a different measure to the ‘good to bad’ measure on the National Survey.

Comparisons within Wales

There are no prior surveys with sufficiently large Welsh samples with which to provide a retrospective time series for the National Survey question on satisfaction with education. In 2007 the Welsh Life and Times Study ran the British Social Attitudes questions above.

5.5.3 Satisfaction with secondary schools

Q36	CORE SHOWCARD D Thinking about the last 12 months, how satisfied or dissatisfied are you with [name]’s secondary school?		
	INTERVIEWER INSTRUCTION: If the child has moved school in the last 12 months, ask about the most recent school.		
D	SINGLE CODE ONLY		
A	Very satisfied	1	GO TO Q37
B	Fairly satisfied	2	GO TO Q37
C	Neither satisfied nor dissatisfied	3	GO TO Q37
D	Fairly dissatisfied	4	GO TO Q37
E	Very dissatisfied	5	GO TO Q37
	Don't know/No opinion (SPONTANEOUS ONLY)	-9	GO TO Q37
	Refused	-99	GO TO Q37
	Question not applicable	-8	GO TO Q37
	Section not applicable	-88	GO TO Q37

❖ We found no surveys which provide sufficiently comparable data with the National Survey for robust analyses.

This question is asked for parents with an 11 to 19 year old child. We found no questions comparable to this question – either within Wales or for comparison with other countries.

³⁸ 2007 most recently.

5.5.4 Post compulsory school aspirations

Q42			
SHOWCARD 8 When [name] reaches 16 and can leave school, what would you like [him/her] to do?			
d	INTERVIEWER INSTRUCTION: If respondent chooses more than one option ask them to pick their main or preferred option. SINGLE CODE ONLY		
A	Continue in full time education	1	GO TO Q43
B	Go on a training course or start an apprenticeship	2	GO TO Q43
C	Get a full-time paid job (either as an employee or self-employed)	3	GO TO Q43
D	Work, unpaid, in the family business full-time	4	GO TO Q43
E	Work, unpaid, as a volunteer full-time	5	GO TO Q43
F	Start a family	6	GO TO Q43
G	Join the armed forces	7	GO TO Q43
H	Other (please specify)	8	GO TO Q43
	Don't know (SPONTANEOUS ONLY)	-9	GO TO Q43
	None of these things (SPONTANEOUS ONLY)	10	GO TO Q43
	It's my child's decision/it's up to my child (SPONTANEOUS ONLY)	11	GO TO Q43
	Refused	-99	GO TO Q43
	Section not applicable	-88	GO TO Q43

❖ We found no current surveys which provide sufficiently comparable data with the National Survey for robust analyses.

This question is identical to one fielded in the Families and Children Study (FACS) in 2008. FACS is no longer running however, meaning that the only opportunity for comparison with FACS would be over time. However, the Welsh sample size in FACS was less than 500 and would be considerably smaller for this question given the focus on families with 14 and 15 year olds³⁹, making the ability to make comparisons infeasible (without the aggregation of data over waves).

³⁹ If any comparisons were made, users should note that, in the National Survey, the question is asked of the eldest child in the 11 to 15 year olds age bracket.

5.5.5 Home learning activities

Q43 SHOWCARD 9 [Name] has been chosen at random. Thinking about [name], how often do you...**IF ONLY ONE CHILD THEN: [Name] has been chosen at random. Thinking about [name], how often do you...**

d SINGLE CODE FOR EACH

Look at books or read stories to [him/her].

Help [him/her] to recognise or learn about letters, numbers or shapes.

Help [him/her] to read or write.

Help [him/her] with school work.

d

A	Every day	1	GO TO Q44
B	Several times a week	2	GO TO Q44
C	Once or twice a week	3	GO TO Q44
D	Once or twice a month	4	GO TO Q44
E	Less often	5	GO TO Q44
F	Not at all	6	GO TO Q44
	Don't know (SPONTANEOUS ONLY)	-9	GO TO Q44
	Refused	-99	GO TO Q44
	Not applicable	-8	GO TO Q44
	Section not applicable	-88	GO TO Q44

❖ The Childcare and Out of School Activities Survey provides *reasonably* comparable data for two of the home learning questions.

Comparisons with other GB or UK countries

These questions are similar to those fielded in the Millennium Cohort Study (MCS), the DfE Childcare and Out of School Activities Survey and (previously) FACS. Arguably, the DfE survey should provide the best up to date 'snapshot' picture for comparison with the home learning activities of parents in England and Wales. The children in the MCS are now too old for these questions to be fielded, and FACS is no longer running.

The DfE survey includes two comparable questions: one on looking at books and reading stories and one on recognising letters, words numbers or shapes. However, slight differences need to be recognised –

- In the National Survey, these questions are asked about three to seven year olds. In the DfE study, they are asked about two to five year olds. Any comparative analysis should compare three to five year olds in each country;
- The National Survey questions ask 'how often' the parents do various activities with their child; the DfE study asks a dichotomous 'yes/no' question then follows up with a frequency scale for those who say yes. There is potential for acquiescence bias in the first question on the DfE study, and the frequency scale could appear to assume some level of frequency, thus depressing the chances of parents reporting that they do not do the activity at all.

- The response scales are different, with the DfE survey providing a more detailed breakdown. However, it is possible to combine groups to do comparable analysis.

Comparisons within Wales

To look at trends in parental involvement with their children’s learning within Wales, there are limited data. As far as we are aware, the Welsh equivalent of the Childcare and Out of School Activities Survey has not included these home learning questions. The Millennium Cohort Study sample size would make its use possible, if the analysis of the National Survey was restricted to the age group of the Cohort children at each wave. However, the only comparable question in one on reading to their children (for which the equivalent National Survey question asks how often people ‘look at books or read stories’ to their child. We are not sure how useful such analysis would be.

5.6 Health care

5.6.1 Overview

We reviewed four questions which measure public views of health services in Wales. The first is an overall satisfaction measure, asked of all respondents. The others are asked only of people with direct experience of the respective services in the past year: one question about hospital care, and two questions about their GPs. Again, the best comparable data we found is for the overall opinion measure, although we found some reasonable comparisons for the GP and hospital care questions.

5.6.2 Satisfaction with health services

The Welsh Government is interested in your views about health services in Wales. When answering these questions please think about all aspects of NHS services, from GPs to pharmacies, NHS dentists and opticians, community health services, and hospitals. For this question I’d like you to give an answer on a scale of nought to ten, where nought is ‘extremely bad’ and ten is ‘extremely good’.

Q91	SHOWCARD 14 Please say what you think overall about the state of health services in Wales nowadays?		
d	SINGLE CODE ONLY. REMINDER: Answer is to be given on a scale of nought to ten, where nought is 'extremely bad' and ten is 'extremely good'.		
	0-10 scale		GO TO Q92
	Don't know (SPONTANEOUS ONLY)	-9	GO TO Q92
	Refused	-99	GO TO Q92
	Section not applicable	-88	GO TO Q92

❖ The European Social Survey provides very closely comparable data with the National Survey.

Comparisons with other GB or UK countries and with wider Europe

Like the question on attitudes towards the education system, this question is taken directly from the European Social Survey, which therefore provides the best source of comparison

data from England⁴⁰ and other European countries. There are two caveats that should be given where comparing this question in the National Survey and the European Social Survey, the first of which is the more important-

- The National Survey question is prefaced with a clarification about what counts as 'health services' for the purposes of the question. It prompts people to think about the full range of healthcare provision, including pharmacies and dentists. This introduction was added in the light of cognitive testing which showed that some respondents did not take account of the full range when answering the question. We might reasonably surmise that, in the absence of the introduction, the European Social Survey respondents think about a more narrow range of provision (probably that closest to their own experience). It is hard to predict how the inclusion of the introduction will affect the distribution of responses, but this could be explored further once data from the National Survey are available.
- In the National Survey, the question comes relatively late in the interview – after a series of questions asking people about their experience of health services during the past year. It is possible that having been recently asked to focus their thoughts on these experiences might affect how people respond to the overall satisfaction question. Again, the size and type of effect is not measurable.

Four other studies include similar, but not identical questions on public attitudes towards the health system which can provide only the broadest of comparisons –

- The **British Social Attitudes survey series** has previously fielded⁴¹ a question on satisfaction with how the NHS runs using a five-point verbalised scale from 'very satisfied' to 'very dissatisfied', rather than the 11-point 'extremely bad' to 'extremely good' scale used in the National Survey.
- The **Public Perceptions of the NHS Survey** fields an almost identical question to the British Social Attitudes Survey.
- The **European Quality of Life Survey** included a question on the 'quality of health services' – using a 0-10 scale, again a different measure to the 'extremely bad' to 'extremely good' scale used in the National Survey.
- **Eurobarometer** uses a four-point verbalised scale from 'very good' to 'very bad' to rate health care provision.

Comparisons within Wales

Time series analysis is limited to only the broadest of comparisons. In 2007 the Welsh Life and Times Study ran the British Social Attitudes questions above.

5.6.3 Satisfaction with hospital care

The National Survey includes a measure of satisfaction with hospital care among those who had an appointment within the last 12 months, asking people to report on their most recent hospital appointment -

⁴⁰ The sample sizes for Scotland and Northern Ireland are small. However, comparisons between Wales and the rest of the UK would be possible. Analysis of the European Social Survey would need to be restricted to respondents aged 16+ to be comparable with the National Survey.

⁴¹ 2007 most recently.

Q81 CORE SHOWCARD D Overall, how satisfied or dissatisfied were you with the care you received?

d SINGLE CODE ONLY

A	Very satisfied	1	GO TO Q83
B	Fairly satisfied	2	GO TO Q83
C	Neither satisfied nor dissatisfied	3	GO TO Q83
D	Fairly dissatisfied	4	GO TO Q83
E	Very dissatisfied	5	GO TO Q83
	Don't know/Can't remember (SPONTANEOUS ONLY)	-9	GO TO Q83
	Refused	-99	GO TO Q83
	Question not applicable	-8	GO TO Q83

❖ The Public Perceptions of the NHS and the Living in Wales Surveys provide *reasonably* comparable data with the National Survey

Comparisons with other GB or UK countries

There are no surveys which provide a direct comparison with this question. However, the Public Perceptions of the NHS Survey does provide a fairly good comparison with England. It fields a very similar question in a face-to-face survey using the same response scale, asking respondents to focus on their last visit. Although it is asked of all respondents, the survey also collects information about whether people have attended hospital in the past year – so comparative analysis could be restricted to those who have. However, there are two, related, issues which may slightly reduce the ability to compare these two survey questions: the Public Perceptions Survey uses quota sampling. Given its primary focus is on attitudes towards the NHS it is feasible that people will be more likely to take part if they have particular experiences they would like to discuss (either positive or negative). Self-selection into surveys is much more of an issue for quota surveys (see Section 4.2.1).

The other survey which we found to be fielding a similar question was the British Social Attitudes survey series. However, we would recommend its use over the Public Perceptions Survey, as it does not ask whether people have had recent experience of hospitals. It is not possible to restrict the sample to make it comparable with the National Survey.

Comparisons within Wales

The Living in Wales Survey included a question which is closely comparable to the National Survey question (with slight differences in wording). We can be pretty confident that the two surveys could be used in combination to produce a time trend of satisfaction with hospital care in Wales.

5.6.4 Satisfaction with GP care

The National Survey includes a satisfaction measure among respondents who have seen a GP in the previous 12 months to report on their last visit –

Q57 CORE SHOWCARD D Overall, how satisfied or dissatisfied were you with the care you received?

d SINGLE CODE ONLY			
A	Very satisfied	1	GO TO Q59
B	Fairly satisfied	2	GO TO Q59
C	Neither satisfied nor dissatisfied	3	GO TO Q59
D	Fairly dissatisfied	4	GO TO Q59
E	Very dissatisfied	5	GO TO Q59
	Don't know/Can't remember (SPONTANEOUS ONLY)	-9	GO TO Q59
	Refused	-99	GO TO Q59
	Question not applicable	-8	GO TO Q59

❖ The Public Perceptions of the NHS Survey, the Living in Wales Survey and, possibly, the Scottish Household Survey provide *reasonably* comparable data with the National Survey

Comparisons with other GB or UK countries

Again, there are few questions with which to make cross-country comparisons within GB (and none with Northern Ireland or wider Europe) –

- **The Public Perceptions of the NHS Survey** can be used to make reasonably confident comparisons with England (with the caveats about quota survey design raised above). Both surveys field questions asking people to rate their last GP visit using the same five-point verbalised satisfaction scale. There are slight differences in question wording, with the National Survey asking about satisfaction with ‘the care you received’ and the Public Perceptions of the NHS survey asking about satisfaction with ‘this last visit as a patient’.
- It may be possible to make comparisons with Scotland using the **Scottish Household Survey**. However, the relevant question has not been fielded since 2004, so further investigation would be required to see if there are plans to field this question in the future. The two questions are very similar though not identical: the National Survey asks respondents about their satisfaction with the ‘care’ they received on their last visit, whereas the Scottish Household Survey asks about their satisfaction with the ‘service’ they received on their last visit. It is possible that a respondent might feel they received excellent care, but that there were aspects of the overall service they were dissatisfied with, leading to differences in the distribution of responses between these questions.
- The **GP Patient Survey in England** fields a series of self-completion questions about ‘how good’ respondents’ last GP visit was in terms of, for example, being treated with care and concern, being given enough time, being listened to, and so on. These questions are sufficiently different to the satisfaction question in the National Survey to provide only the broadest comparisons with between Wales and England.

Comparisons within Wales

The Living in Wales Survey provides a fairly robust time series preceding the National Survey. It asks a similar question – although using the term ‘service’ rather than ‘care’. Given the Living in Wales Survey uses the same mode and sampling procedure as the National Survey, it provides a better time series comparison than the self-completion GP Access Survey, which asks a general satisfaction question about GP care.

5.6.5 Convenience of GP appointments

Respondents who have seen a GP in the previous 12 months are asked to report on the ease or difficulty of getting a convenient appointment time on their last visit –

Q52 CORE SHOWCARD A How easy or difficult was it to get an appointment at a time that was convenient to you?			
SINGLE CODE ONLY			
A	Very easy	1	GO TO Q54
B	Fairly easy	2	GO TO Q54
C	Fairly difficult	3	GO TO Q53
D	Very difficult	4	GO TO Q53
	Don't know/Can't remember (SPONTANEOUS ONLY)	-9	GO TO Q54
	Refused	-99	GO TO Q54
	Question not applicable	-8	GO TO Q54

❖ The Living in Wales study provides *reasonably* comparable data with the National Survey.

Comparisons across GB or the UK

We found no surveys with precisely comparable questions to the National Survey question on the ease of making conveniently-timed appointments with the GP, but found three with similar questions. Any comparisons – either over time or across countries – will necessarily be made in the knowledge of the differences in the question wordings and methodology. We make no particular recommendations on the best comparisons, but rather outline the advantages and disadvantages of each –

For GB comparisons:

- The **British Social Attitude survey series** includes two Likert-style questions asked of all respondents about the extent to which they agree or disagree that it is easy to arrange an appointment with the GP (a) during the working day and (b) at the evening or weekend⁴². The survey does not include a question about whether respondents have been to their GP with the past year, so it is not possible to compare the answers of respondents in the National Survey (where only respondents having seen a GP or family doctor in the past year are asked about appointments).
- The **GP Patient Survey in England** includes a question asked of respondents who attempted and managed to get an appointment with their GP how convenient the appointment was. It should be noted that this question is asked of respondents who have seen their GP within the last year, but also of respondents who last saw their GP 'more than 12 months ago'. For the purposes of comparison therefore, analyses should only be conducted on those respondents whose most recent appointment with their GP was within the last year. On face value this question may be more comparable with the National Survey than the BSA questions. However, one should be cautious about the

⁴² This question is not fielded every year, so if it were to be used as a GB comparison, it may be necessary to accept that slightly different fieldwork periods are being compared, or to only make comparisons in years where BSA fields the question.

fact that the survey is a self-completion postal survey with a response rate of less than 50 per cent – so both mode and representativeness are potential issues affecting comparability.

Comparisons within Wales

The Living in Wales Survey includes a very similar question to the National Survey – although the question asked more broadly about appointments with ‘health care professionals at your GP surgery’. Given it may be more or less easy to secure convenient appointments with staff such as nurses or health visitors, the question is not directly comparable, although, with caveats, could be used as a time trend measure. The GP Access Survey asks (among a battery of questions) how easy people find it to book appointments with their GP. Again, this question is only of limited use in making time series comparisons with the National Survey.

5.7 Views on other government policy issues

5.7.1 Overview

In addition to the measures above we reviewed two other questions – a measure of how well people think the Welsh Government is ‘doing its job’ and a measure of how good or bad the transport system is within Wales. While the first has a very good comparator which will allow the Welsh Government to review its performance alongside England and other European countries, we found no questions with which to make comparisons about the transport system.

5.7.2 Satisfaction with the Welsh Government

Q87 SHOWCARD 12 For this question I'd like you to give an answer on a scale of nought to ten, where nought is 'extremely dissatisfied' and 10 is 'extremely satisfied'. Now thinking about the Welsh Government, how satisfied are you with the way it is doing its job?

D SINGLE CODE ONLY

	0-10 scale	GO TO Q88
Don't know (SPONTANEOUS ONLY)	-9	GO TO Q88
Refused	-99	GO TO Q88

❖ The European Social Survey provides very closely comparable data with the National Survey.

Comparisons with other GB or UK countries and with wider Europe

This question is taken directly from the European Social Survey, which is therefore the best source of comparison data from England⁴³ and other European countries. However, there is one important caveat that should be given where comparing this question in the National Survey and the ESS. In the European Social Survey, the question is immediately preceded by a series of questions asking about political affiliation and, importantly, the economy. It is

⁴³ The sample sizes for Scotland and Northern Ireland are small. However, comparisons between Wales and the rest of the UK would be possible. Analysis of the ESS would need to be restricted to respondents aged 16+ to be comparable with the National Survey.

feasible that the focus of these preceding questions may affect how people then respond when asked to make a judgement on how well the government is doing its job. The position of the question is quite different in the National Survey interview, where it follows a series of questions about local authority service provision. The size and type of any resulting effect is not measurable, but the difference should be borne in mind.

We found no other questions sufficiently similar to enable comparisons with other countries. The British Social Attitudes Survey series asks a question about how good or bad a job the UK Government has done over the past four years, using a three-point scale. The differences in time period and response scale make any comparison difficult. Likewise, the Ipsos Political Monitor includes a dichotomous ‘satisfied vs dissatisfied’ question on how well the Government is running the country, but uses a quota sample design.

Comparisons within Wales

We found no comparable questions with a sufficiently large sample size in Wales for any within-Wales comparisons. The National Survey 2009-10 Pilot included a similar question which used a five-point satisfaction scale.

5.7.3 Satisfaction with transport provision

The Welsh Government is interested in your views about transport in Wales. When answering these questions please think about all aspects of transport – from roads, rail and bus services to cycling and walking. For this question I’d like you to give an answer on a scale of nought to ten, where nought is ‘extremely bad’ and ten is ‘extremely good’.

Q93	SHOWCARD 14 Please say what you think overall about the state of the transport system in Wales nowadays?		
d	REMINDER: Answer is to be given on a scale of nought to ten, where nought is 'extremely bad' and ten is 'extremely good'. SINGLE CODE ONLY.		
	0-10 scale		GO TO Q94
	Don't know (SPONTANEOUS ONLY)	-9	GO TO Q94
	Refused	-99	GO TO Q94
	Section not applicable	-88	GO TO Q94

❖ We found no current surveys which provide sufficiently comparable data with the National Survey for robust analyses.

While this question mirrors the structure of the ESS-based questions on the education system and health provision, we found no direct comparisons for this question in other surveys. Two questions – one on the European Quality of Life Survey and one on the Citizenship Survey – ask people to rate the quality of (in the case of the European Quality of Life Survey) or their satisfaction with (in the case of the Citizenship Survey) public transport. It should be noted that the Citizenship survey asks about ‘local public transport’. The narrower foci of these questions, and the difference in response scale, do not make for meaningful comparisons with the National Survey.

5.8 Demographics

5.8.1 Overview

We reviewed four demographic questions: the prevalence of Welsh language skills, education levels, financial stability and use of the internet.

5.8.2 Welsh language skills

Q10	Can you...
d	READ OUT EACH ONE
	Understand spoken Welsh
	Speak Welsh
	Read Welsh
d	Write Welsh
	Yes 1
	No 2
	Don't know (SPONTANEOUS ONLY) -9
	Refused -99

❖ The Annual Population Survey provides the best directly comparable data with the National Survey.

The National Survey measures the proportion of the Welsh population who understand Welsh, with four separate measures for understanding spoken Welsh, and for speaking, reading, and writing Welsh. We assume that Welsh Government's interest is in the comparability of the prevalence rates collected in the National Survey with other Welsh data.

There are similar or identical questions in –

- the Annual Population Survey
- the 2001 and 2011 Censuses
- Living in Wales, including the Welsh Language Usage Survey

If the Welsh Government wants to use the National Survey in combination with previous data to make comparisons over time, we recommend the use of previous waves of the Annual Population Survey for the following reasons.

Likewise, we would expect comparable results for data collected for the two surveys at similar points in time –

- The question wording and the mode in which the questions are asked are identical;

- The sample design is comparable⁴⁴;
- They are both face-to-face surveys;

The one caveat we place upon comparison between the two surveys is the position of the question in the two interviews. In the National Survey, this question is asked at the start of the interview, while in the Annual Population Survey it follows a set of questions on national identity. An ONS study by Haselden (2003) suggests that asking about national identity prior to asking about Welsh language skills potentially leads to respondents being more likely to say they are Welsh speakers.

This same study showed that Census data provide lower prevalence estimates than face-to-face surveys such as the Annual Population Survey, and we would therefore recommend caution if making comparisons between the National Survey and the Census on this issue. Although it is difficult to pinpoint the exact reasons for the differences in estimates between the Census and face-to-face survey data, Haselden suggests that the mode has an effect, probably partly due to satisficing and partly due to a tendency for respondents to read and complete self-completion forms less carefully (e.g. failing to tick all the codes that apply to them). The same caution would apply to comparison with the Welsh Language Survey which was asked as a self-completion supplement to the Living in Wales Survey.

The Living in Wales Survey (face-to-face element) and National Survey 2009-10 Pilot can be used alongside the National Survey questions to produce time trend analysis on the prevalence of Welsh speaking and people's fluency, though this would need to be limited to those aged 16 and over. The Living in Wales Survey asked about household members' ability to speak Welsh (but not understand, read or write). It also asked the same question as fielded in the National Survey about the respondents' fluency in speaking Welsh. However, any comparisons with Living in Wales should be made with one caveat: the household reference person reports whether each household member 'speaks Welsh' or not. Given the subjective nature of this question, it may be answered differently in proxy rather than in person.

In reality, when looking at trends in Welsh language ability over time, the Welsh Government may decide to use the Annual Population Survey alone (i.e. not in combination with the National Survey). In the Annual Population Survey, the question is asked of all adults *and* children, and it is therefore possible to produce estimates of Welsh language ability for the Welsh population aged three and over. In comparison, the National Survey provides estimates for adults aged 16 and over.

⁴⁴ Although the Annual Population Survey collects data about all household members (of all ages) and the National Survey collects data about one randomly selected adult (16+), provided that any comparisons are made (a) for adults 16+ and (b) applying the relevant survey weights, the two surveys will produce comparable estimates. See Section 4.2.3.

5.8.3 Educational qualifications

Q100 SHOWCARD 19 Please look at this card and tell me which of these qualifications you have. Please read out the letter next to the qualifications which apply.			
d	INTERVIEWER NOTE: IF UK QUALIFICATION IS NOT LISTED, SELECT ITS NEAREST EQUIVALENT. IF QUALIFICATIONS GAINED OUTSIDE THE UK, SELECT 'FOREIGN QUALIFICATIONS' AND THE NEAREST UK EQUIVALENTS (IF KNOWN). CODE ALL THAT APPLY		
A	1 - 4 O levels / CSEs / GCSEs (any grades), Entry Level, Foundation Diploma	1	GO TO Q102
B	NVQ Level 1, Foundation GNVQ, Basic Skills	2	GO TO Q102
C	5+ O levels (passes) / CSEs (grade 1) / GCSEs (grades A*- C), School Certificate, 1 A level / 2 - 3 AS levels / VCEs, Welsh Baccalaureate Intermediate Diploma	3	GO TO Q102
D	NVQ Level 2, Intermediate GNVQ, City and Guilds Craft, BTEC First / General Diploma, RSA Diploma	4	GO TO Q102
E	Apprenticeship	5	GO TO Q102
F	2+ A levels / VCEs, 4+ AS levels, Higher School Certificate, Progression / Welsh Baccalaureate Advanced Diploma	6	GO TO Q102
G	NVQ Level 3, Advanced GNVQ, City and Guilds Advanced Craft, ONC, OND, BTEC National, RSA Advanced Diploma	7	GO TO Q102
H	Degree (for example BA, BSc), Higher degree (for example MA, PhD, PGCE)	8	GO TO Q102
I	NVQ Level 4 - 5, HNC, HND, RSA Higher Diploma, BTEC Higher Level	9	GO TO Q102
J	Professional qualifications (for example teaching, nursing, accountancy)	10	GO TO Q102
K	Other vocational / work-related qualifications	11	GO TO Q102
L	Foreign qualifications	12	GO TO Q102
M	No qualifications	13	GO TO Q102
	Don't know (SPONTANEOUS ONLY)	-9	GO TO Q102
	Refused	-99	GO TO Q102

❖ The Census provides the best directly comparable data with the National Survey.

The National Survey uses the ONS harmonised question on education levels, which uses identical categories to the 2011 Census. There is a wide degree of variation in how other surveys collect education levels, and it is difficult to assess their direct comparability with the National Survey without undertaking some analysis of the ways in which the different surveys combine the categories and variations in how different category lists affect respondent comprehension or response. This may be an issue for further investigation once some National Survey data are available. In the meantime, we review the extent to which the National Survey should, in principle, mirror the findings from the Census in Wales.

While we might expect the National Survey findings to closely match those of the 2011 Census, the following mode effects may lead to differences -

- While respondents to both the National Survey (via a showcard) and the Census (via the self-completion form) have the full list of qualifications to select from, National Survey respondents will arguably provide a fuller (better) response than people give in the Census. In Section 4.3.1, we discussed the fact that self-completion modes are more

prone to respondents answering questions less thoroughly (satisficing), especially on more difficult questions like this education variable. In an interviewer-administered survey, interviewers are trained to prompt respondents to include all the relevant qualifications (which may include a number of ‘any others’ follow up prompts).

- On the other hand, there is a danger a list of this kind which includes a ‘no qualifications’ category right at the bottom of a long list may elicit some degree of acquiescence and/or social desirability bias among face-to-face survey respondents, over-reporting the incidence of having any qualifications.

5.8.4 Financial stability

Q117 SHOWCARD 22 Which one of the statements on this card best describes how well you are keeping up with your bills and credit commitments at the moment? Please just read out the letter that applies.

d CAPI TO ONLY HAVE LETTERS FOR PRIVACY. SINGLE CODE ONLY			
A	Keeping up with all bills and commitments without any difficulties	1	GO TO Q118
B	Keeping up with all bills and commitments but it is a struggle from time to time	2	GO TO Q118
C	Keeping up with all bills and commitments but it is a constant struggle	3	GO TO Q118
D	Falling behind with some bills or credit commitments	4	GO TO Q118
E	Having real financial problems and have fallen behind with many bills or credit commitments	5	GO TO Q118
F	Have no bills	6	GO TO Q118
	Don't know (SPONTANEOUS ONLY)	-9	GO TO Q118
	Refused	-99	GO TO Q118

❖ We found no current surveys which provide sufficiently comparable data with the National Survey for robust analyses.

We found no questions which are directly comparable with the National Survey question, with the exception of the National Survey Pilot in 2009-10. This provides a short time-series comparison within Wales.

This means that comparability with other countries is problematic. Some studies ask specifically about loan repayments (e.g. the Continuous Household Survey and the Life Opportunities Survey), which is a much narrower issue than ‘bills and credit commitments’. The European Union Statistics on Income and Living Conditions Survey asks about ‘total housing costs’, which excludes credit commitments unconnected to property. All of these surveys employ response scales that differ from that of the National Survey.

Arguably the best survey for contemporary cross-country comparisons is the European Social Survey which asks how respondents ‘feel about your household’s income nowadays’. Although the scale is very different to the National Survey question (a four-point verbalised scale from ‘living comfortably on present income’ to ‘finding it very difficult on present income’), it could be used to provide very broad comparisons.

5.8.5 Use of the internet

Q120	Do you personally use the internet at home, work or elsewhere?		
D	SINGLE CODE ONLY		
	Yes	1	GO TO Q123
	No	2	GO TO Q121
	Don't know (SPONTANEOUS ONLY)	-9	GO TO Q128
	Refused	-99	GO TO Q128

- ❖ The Oxford Internet Household Survey provides very closely comparable data with the National Survey.
- ❖ The European Social Survey and the Living in Wales Survey provide *reasonably* close comparisons.

Comparisons with other GB or UK countries

Although none of the surveys we reviewed included exactly the same question to the one in the National Survey above, the factual nature of this question means that we should be relatively confident in recommending the use of these questions alongside the National Survey to compare Wales with other UK or wider European countries. We would not expect order effects⁴⁵ for this kind of question (i.e. responses should be unaffected by the position of the question in the interview relative to other questions). Rather, in order to measure the prevalence of internet use in different countries accurately, we recommend comparisons to surveys with comparable sample designs and response rates.

Some questions ask slightly broader questions (e.g. capturing present and past internet use; asking about frequency of internet use). But, the questions are often worded such that analysis can be restricted to current use as in the National Survey question. However, the surveys vary in whether they capture personal or personal and work use.

One difficulty we found when assessing the comparability of questions is the slightly ambiguous nature of the National Survey question. It is not entirely clear whether the word 'personally' means 'you yourself' or 'for personal use'. However we note that this question is immediately preceded by a question which asks "[Do you/Does your household] have access to the internet at home?". Given that this preceding question concerns access to the internet, rather than type of use of the internet, we would expect most respondents to interpret 'personally' as meaning 'you yourself'.

The Oxford Internet Survey is arguably the best GB comparator question as it uses a similar question wording to the National Survey: 'Do you yourself personally use the Internet at home, work, school, college or elsewhere or have you used the Internet anywhere in the past?'⁴⁶. The Oxford Internet Survey question specifies 'you yourself,' removing any

⁴⁵ Or mode effects, although all of the surveys we identified were face-to-face.

⁴⁶ The OXIS question asks about present or past usage, so comparative analysis would be restricted to those who currently use it. Note that the eligible age group for the OXIS survey is those aged 14 and over. This means that any analysis of the two surveys should be restricted to 16 and over for comparability. It may also explain the direct reference to schools in the OXIS question.

potential ambiguity that might arise in the National Survey. However to the extent that this ambiguity is minimal in the National Survey comparisons should not be affected.

The fact that the Oxford Internet Survey prompts, in the same way as the National Survey, for use at home, work or elsewhere (plus school or college in the case of the Oxford survey) arguably makes it a better comparison than others which do not remind respondents to include usage anywhere. There are some differences –

- OXIS prompts about use at school or colleges, as well as home and work. Plus it has a follow up prompt which double checks on people who say they do not use the internet. These two factors might mean that the OXIS figures are slightly higher than the National Survey figures.
- The primary focus of OXIS is on internet use. Although it achieves a good response rate (72 per cent in the last survey), there is some risk that people are more likely to take part if they have an interest in the issues.

So, although the data from two surveys should be very closely comparable and allow for comparisons between England and Wales, these factors mean that OXIS may slightly overestimate the prevalence of internet usage compared to the National Survey.

The other UK survey which provides a relatively close comparison is the Annual Population Survey which asks when people last used the internet. Excluding those who say ‘never’ provides a pretty close comparison with the National Survey – with sample sizes large enough for each UK country. However, it does not prompt people to include their internet use at home, at work, or elsewhere. Moreover the Annual Population Survey question does not use the word ‘personally’, opting instead simply for ‘When did you last use the internet’. While we would not expect this to have a large impact on comparability, it is possible that there may be differences in interpretation between this question and the National Survey question^{47,48}.

The British Social Attitudes Survey and Scottish Household Survey both restrict their questions to use other than work. Given the potential ambiguity in the National Survey question we would not recommend these two surveys’ use for comparability. Likewise, the Continuous Household Survey in Northern Ireland asks about ‘having access’ to the internet, rather than use.

Comparisons with wider Europe

The best source of comparison we found across Europe more widely was the European Social Survey which asks about use of the internet at home or work for personal use. Again, it uses a frequency scale which can be used to derive an ‘ever use’ statistic which is comparable with the National Survey question. The extent to which it can be compared directly with the National Survey depends on how people interpret the term ‘personally’ and whether this equates to personal use as in the European Social Survey.

⁴⁷ In general, the APS tends to report higher rates of internet use than other surveys.

⁴⁸ We also reviewed the Ofcom Communications Market Report. However, differences in sample design (its data is based on a quota survey) and question wording (it asks about household use, then personal use) mean that we would not recommend its use in comparisons with the National Survey, given the better comparisons available.

Comparisons within Wales

The Living in Wales Survey provides a reasonably good comparison over time within Wales. It asks how often people access the internet, in the sense of 'use the internet,' as opposed to 'have access to the internet'. It can be used to derive an 'ever use' statistic. However, as before, as it does not use the term 'personally' the two questions may be interpreted slightly differently.

6 Concluding thoughts

6.1 Introduction

Chapters 4 and 5 have highlighted the wide range of issues which arise when attempting to make robust comparisons between questions asked in different surveys. It is almost impossible to have the ‘perfect’ comparison question, taking into account sample design, survey administration and question wordings. We have identified a number of questions among the 20 we were asked to review when comparisons can be made with a fair degree of confidence – either with other countries or within Wales over time. However, we have also shown how sometimes small differences (e.g. the position of the question, differences in wording) may have a large effect on how respondents react, and therefore on the ability to make robust comparisons. While we explain how some differences between surveys can be taken into account in data analysis (increasing comparability), many issues are not ‘solvable’ and must simply be recognised when assessing the findings across different surveys or over time.

Some surveys are designed and administered with comparability as the, or a, key priority. For instance, central to the European Social Survey is the need for direct comparability across the participating countries in relation to survey design and questions. Panel surveys, such as Understanding Society, or cross-sectional time series like British Social Attitudes try to ensure comparability across years. However, the National Survey is a multi-purpose survey with several aims in relation to informing government policy, only one of which is to provide comparability with other countries and over time. Often, the need for questions which work in the Welsh context at that point in time may be more important than comparability over time or, particularly, between countries. It is therefore important that the Welsh Government understands when and how it can make ‘less than perfect’ comparisons between the National Survey and other surveys, which was the one purpose of this report. The issues raised in the report may aid any future question development (especially when drawing on questions from other surveys), as well as in the analysis and interpretation of the National Survey data.

In the final sub-section below, we list a few thoughts about further work that could be done – either to further investigate the comparability of the National Survey questions or take into account in any future question development.

6.2 Thoughts on future work

In Chapter 5, there are a number of occasions where we *surmise* that different question wordings will produce somewhat different findings across the surveys. Once some National Survey data are available, it will be possible to test out some of these assumptions. Analysis of the National Survey data and data from comparator surveys would show whether, and how, the comparability issues we raise do seem to have an effect on the distribution of responses in different surveys.

The greatest opportunities for doing this use the small number of surveys like the Integrated Household Survey (see Section 3.3) which have sufficiently large Welsh sample sizes,

collected over a similar timeframe to the National Survey. However, most of the queries we raise about comparability concern surveys with sample sizes in Wales too small for this kind of analysis.

It would be possible to use other surveys, although in the main, this will necessarily be a relatively unscientific exercise. Although we can compare findings in Wales (using National Survey data) to findings in one or more other countries, unless we have an idea of the true level of difference we might expect between the countries, we will not know whether any differences in results are caused by incomparability or whether they are real between-country differences. Likewise, this applies to comparisons over time within Wales. However, if the questions are carefully selected on the basis that we would not expect large differences between countries and if the comparison is made with countries which we might expect to be similar (on the chosen variable) to Wales, we could be *reasonably* confident that any differences we find are probably due to issues of survey/question comparability.

This type of analysis is worthwhile in ‘flagging up’ issues, rather than formally testing the impact of differences in question wording, position and so forth. So, we also suggest that some thought is given to introducing some split-run experiments to the National Survey to test – with a random sub-set of respondents – whether some of the comparator questions identified on other surveys would in fact elicit different responses. Fielding both the National Survey question and the comparator question on the National Survey would allow for a formal test of the difference between them. This may be feasible given the large sample size of the National Survey, especially as the ‘test variant’ need only be tested on a sub-set (not necessarily half) of the sample.

References

Dillman, D (2000) *Mail and Internet Surveys: The Tailored Design Method. Introduction to tailored design*, Hoboken, NJ: Wiley

Haselden, L (2003) Differences in estimates of Welsh language skills, London: ONS

Krosnick, J, & Berent, M (1993) Comparisons of Party Identification and Policy Preferences: The Impact of Survey Question Format. *American Journal of Political Science*, 37(3), 941-964
JSTOR

Jackle, A, Roberts C & Lynn P (2008) Assessing the Effect of Data Collection Mode on Measurement, ISER working paper series 2008-08

Nicolaas, G (2012) Mixing Modes within a Social Survey: Opportunities and constraints for the National Survey for Wales

Nicolaas, G et al (2011) Is it a good idea to optimise question format for mode of data collection? Results from a mixed modes experiment, ISER Working Paper Series

Pudney, S (2010) An experimental analysis of the impact of survey design on measures and models of subjective well-being, *Understanding Society Working Paper Series No 2010-01*

Purdon, S (2012) Generating aggregate statistics from National Survey for Wales data

Schuman, H and Presser, S (1981) *Questions and answers in attitude surveys: experiments in question form, wording and context* New York: Academic Press

Tipping, S, Hope, S, Pickering K, Mindell, J and Erens, H (2008) An analysis of mode effects using data from the Health Survey for England 2006 and the Boost Survey for London, NHS Information Centre

Appendix 1: comparator survey spreadsheet

Appendix 2: comparator question spreadsheet