Dadansoddi ar gyfer Polisi
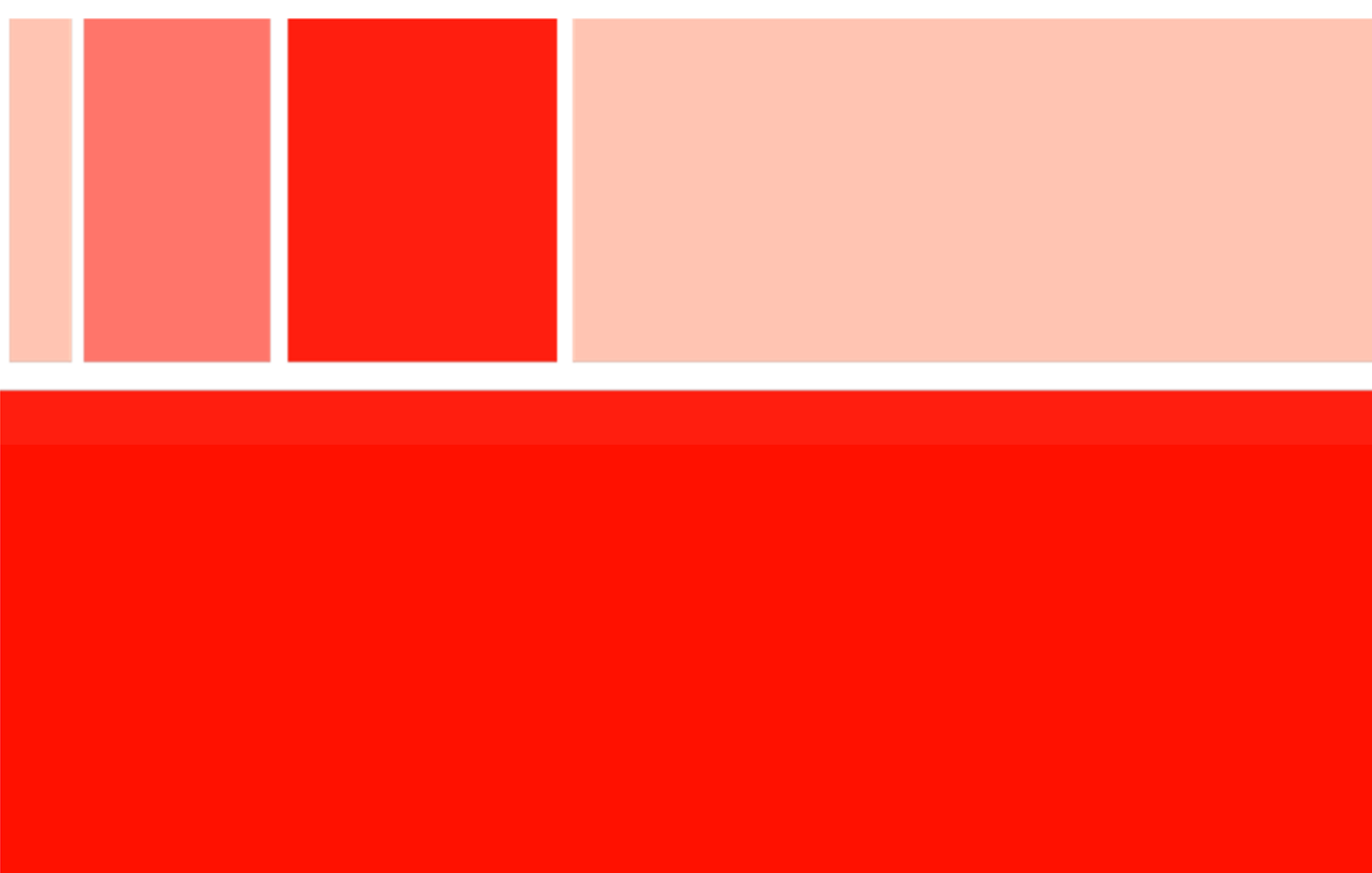
YMCHWIL GYMDEITHASOL
Y LLYWODRAETH

GSR

GOVERNMENT SOCIAL RESEARCH

Analysis for Policy

Llywodraeth Cymru
Welsh Government

National Survey for Wales, 2019-20 Data user guide

National Survey for Wales, 2019-20 Data user guide

For further information please contact:

Siobhan Evans

National Survey team

Knowledge and Analytical Services

Welsh Government

Cathays Park, Cardiff

CF10 3NQ

Tel: 03000 256685

Email: siobhan.evans@gov.wales

Contents

1. Introduction

The National Survey for Wales is a major study of adults across Wales. The survey covers a range of topics, and asks people about their experience of and satisfaction with public services, their local area and other aspects of their lives. The survey provides results at both national and local authority level. The information collected is used to inform the development of policy and the delivery of public services.

From the outset, it was intended that the data should be made publicly available for secondary analysis. The aim of this document is to provide potential users and other interested parties with guidance on how to undertake basic analysis.

This document provides some background to the survey and the survey methodology, describes the datasets available, and provides some example SPSS and Stata syntax to help guide analysis.

A more detailed account of how the survey was carried out can be found in the Technical Report. Further background information about the survey, including information provided to respondents and full questionnaires, is available on the National Survey web pages.

2. Survey design and methodology

2.1 Background to the survey
The National Survey for Wales began in 2012 as a random sample, 25 minute face to face interview. In 2015 the survey was reviewed and it was decided to amalgamate 5 large scale social surveys that were carried out in Wales into one.

From 2016-17 onwards the survey was expanded to include topics previously covered by the Welsh Health Survey, Active Adults Survey, Arts in Wales Survey, and Welsh Outdoor Recreation Survey

The National Survey for Wales in its new form is a random sample, 45 minute face-to-face survey of around 12,000 adults each year (aged 16+) living in private households across Wales. The aim of the survey is to provide representative, reliable and up-to-date information about the people of Wales down to local authority level.

The survey covers a wide range of topics to allow links to be made between different policy areas. The 2019-20 topics were:

- Welsh language – use in local services and with local authorites
- Well-being
- Internet use and skills
- Online public sector services
- Local authority services
- Primary and secondary education
- Childcare
- General health and illnesses
- Health services – GP, Hospitals + non-GP services
- Health information – advice, support and online
- Medicine and side effects
- Arthritis and bone conditions
- Dental health
- Lifestyle behaviours (fruit and vegetables, alcohol smoking, exercise, BMI)
- Tax Devolution
- Material deprivation (household, pensioner and child)
- Universal Credit
- Volunteering and caring
- Employment, Fair work, ageism and bullying
- Planning for the future
- Employability, support and skills development
- Transport
- Arts attendance and participation
- Heritage, museums and libraries
- Sport participation and latent demand
- Loneliness
- Physical punishment of children

The [full questionnaire](#) is available. The topics included in the questionnaire change from year to year to keep up with the changing needs of survey users.

Results of the survey are reported in a First Release (published July 14th 2020) and series of statistical reports, which are published on the National Survey releases. Tables and charts of the data can also be found on our results viewer.

2.2 Sampling
The survey is designed to provide a nationally-representative sample of the adult population of Wales living in private households. Results can be seen at different geographical levels such as local authority and local health board level (although some topics are only asked of a subsample of respondents, where results are only required at regional or national level).

Each year, over 24,000 addresses are selected at random from the Royal Mail's Small User Residential Postcode Address File. Interviewers visit each address, collect basic household composition information and then randomly select one adult (aged 16+) to interview. Over 12,000 interviews were achieved in 2019-20. The sample size for each local authority was proportional to the population of that local authority; however a minimum effective sample size was set in each local authority of 250 interviews. The response rate in 2019-20 was 59%.

2.3 Fieldwork periods
The survey runs continuously with questionnaire content changing at the start of each financial year. The 2019-20 topics were included from the beginning of April 2019 until the mid-March 2020 (fieldwork finished two weeks earlier than expected due to the coronavirus situation).

2.4 Technical Report
Full details of how the survey was carried out can be found in the Technical Report. The Technical Report provides information on the survey design, how the questionnaire was developed, fieldwork and data processing procedures, as well as information on weighting and survey response rates.

2.5 Quality Report
A Quality Report is also available, containing more detailed information on the quality of the survey as well as a summary of the methods used to compile results.

3. Overview of the datasets

3.1 Datasets
There are four basic units of analysis for the National Survey. Information is collected about the respondent, the household, the other members of the household, and the address. These are kept in 3 separate datasets as described below.

Respondent file
This is the main dataset to use. It contains information collected from the respondent about the respondent and the household.

This dataset contains all respondents' answers to the survey including their demographic information such as their age, sex, economic status, general health, as well as their views about public services and their local area. This dataset also contains information about the household itself such as area details, tenure, household composition, and whether the household has internet access or broadband.

The dataset also contains weights which can be applied to the data so that the information represents either all adults or all households in Wales, as appropriate for each topic. More details on weights and how to use them is provided in section 4.1.

This file is available is available to download at the UK [Data Service](#). More detailed variables, such as Lower Super Output Area (LSOA) for each case, is available on request: please contact the National Survey team at [surveys@gov.wales](#).

All people file
This dataset contains basic information about all members of the household who live at the selected addresses. For example, we ask for the age and gender of all household members as well as their relationship to the selected respondent. This dataset may be made available upon request for research purposes only: please contact the National Survey team at [surveys@gov.wales](#).

Sample file
This dataset contains anonymised information about all addresses that were selected for interview. For example whether they took part in the survey, refused or whether the property was derelict etc. It also contains some information coded by the interviewer about their impressions of the property. This dataset may be made available upon request for research purposes only: please contact the National Survey team at [surveys@gov.wales.](#)

3.2 Structure of datasets
The following section provides some background information on how the datasets are organised.
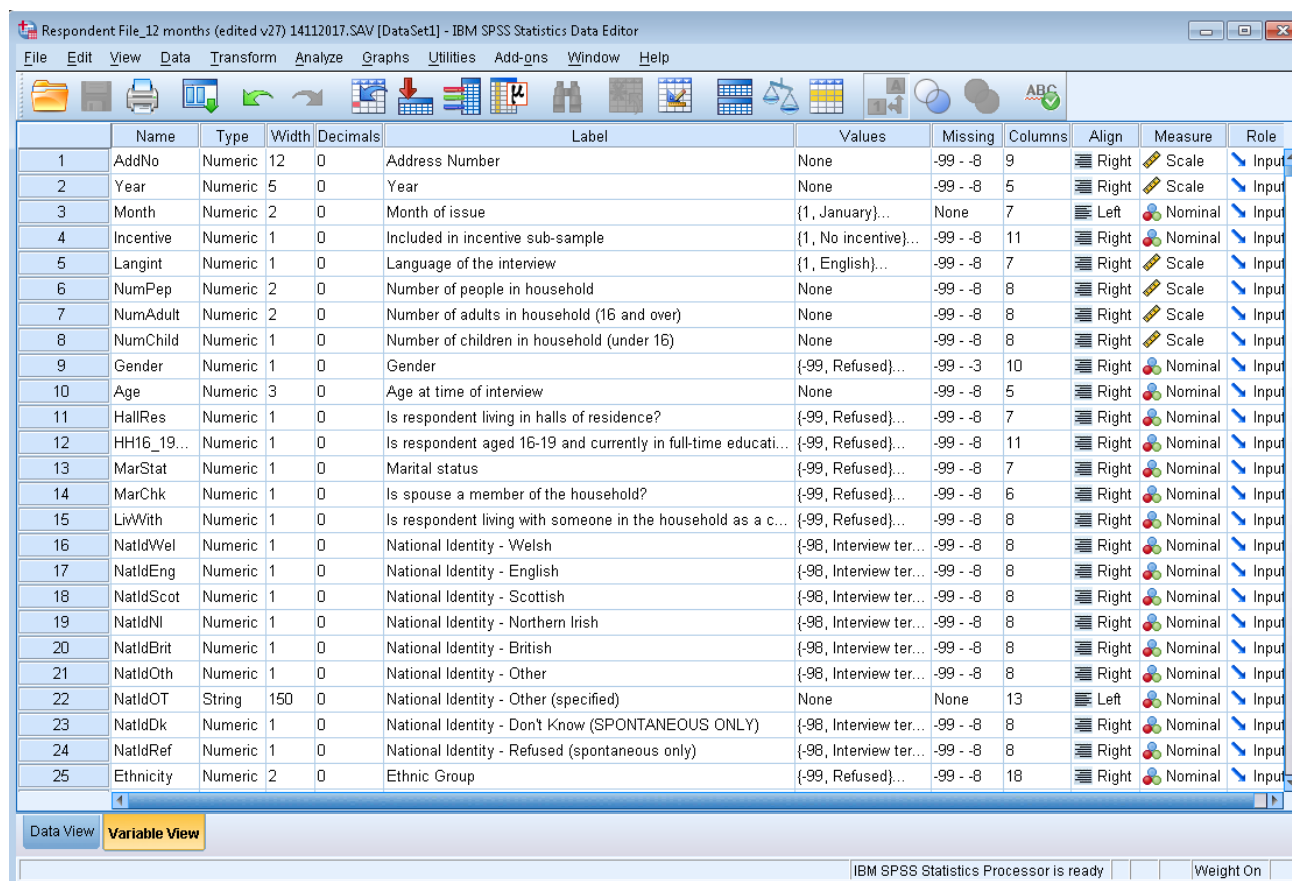
Case identification
Each address has been assigned its own unique identification number, stored as 'AddNo'. In the All people file each person in a household (address) is given their own identifier, 'Person'. In the Sample file multiple Households at an address point are given a unique identifier, 'HhNo'. To ensure anonymity the unique identifier (AddNo) in the respondent file, deposited with UKDS, has been recoded and stored as 'CaseNo'.

Variable names
The variables have been named to reflect the questions asked and the group of questions they belong to. A full list of variables can be viewed in the accompanying documentation "variable catalogue". This is a good starting point when deciding which survey results to analyse.

A screenshot of the respondent dataset below shows the type of variables included in the National Survey, and how they are presented in SPSS.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	AddNo	Numeric	12	0	Address Number	None	-99 - -8	9	Right	Scale	Input
2	Year	Numeric	5	0	Year	None	-99 - -8	5	Right	Scale	Input
3	Month	Numeric	2	0	Month of issue	{1, January}...	None	7	Left	Nominal	Input
4	Incentive	Numeric	1	0	Included in incentive sub-sample	{1, No incentive}...	-99 - -8	11	Right	Nominal	Input
5	Langint	Numeric	1	0	Language of the interview	{1, English}...	-99 - -8	7	Right	Scale	Input
6	NumPep	Numeric	2	0	Number of people in household	None	-99 - -8	8	Right	Scale	Input
7	NumAdult	Numeric	2	0	Number of adults in household (16 and over)	None	-99 - -8	8	Right	Scale	Input
8	NumChild	Numeric	1	0	Number of children in household (under 16)	None	-99 - -8	8	Right	Scale	Input
9	Gender	Numeric	1	0	Gender	{-99, Refused}...	-99 - -3	10	Right	Nominal	Input
10	Age	Numeric	3	0	Age at time of interview	None	-99 - -8	5	Right	Nominal	Input
11	HallRes	Numeric	1	0	Is respondent living in halls of residence?	{-99, Refused}...	-99 - -8	7	Right	Nominal	Input
12	HH16_19...	Numeric	1	0	Is respondent aged 16-19 and currently in full-time educati...	{-99, Refused}...	-99 - -8	11	Right	Nominal	Input
13	MarStat	Numeric	1	0	Marital status	{-99, Refused}...	-99 - -8	7	Right	Nominal	Input
14	MarChk	Numeric	1	0	Is spouse a member of the household?	{-99, Refused}...	-99 - -8	6	Right	Nominal	Input
15	LivWith	Numeric	1	0	Is respondent living with someone in the household as a c...	{-99, Refused}...	-99 - -8	8	Right	Nominal	Input
16	NatIdWel	Numeric	1	0	National Identity - Welsh	{-98, Interview ter...	-99 - -8	8	Right	Nominal	Input
17	NatIdEng	Numeric	1	0	National Identity - English	{-98, Interview ter...	-99 - -8	8	Right	Nominal	Input
18	NatIdScot	Numeric	1	0	National Identity - Scottish	{-98, Interview ter...	-99 - -8	8	Right	Nominal	Input
19	NatIdNI	Numeric	1	0	National Identity - Northern Irish	{-98, Interview ter...	-99 - -8	8	Right	Nominal	Input
20	NatIdBrit	Numeric	1	0	National Identity - British	{-98, Interview ter...	-99 - -8	8	Right	Nominal	Input
21	NatIdOth	Numeric	1	0	National Identity - Other	{-98, Interview ter...	-99 - -8	8	Right	Nominal	Input
22	NatIdOT	String	150	0	National Identity - Other (specified)	None	None	13	Left	Nominal	Input
23	NatIdDk	Numeric	1	0	National Identity - Don't Know (SPONTANEOUS ONLY)	{-98, Interview ter...	-99 - -8	8	Right	Nominal	Input
24	NatIdRef	Numeric	1	0	National Identity - Refused (spontaneous only)	{-98, Interview ter...	-99 - -8	8	Right	Nominal	Input
25	Ethnicity	Numeric	2	0	Ethnic Group	{-99, Refused}...	-99 - -8	18	Right	Nominal	Input

Single-code response categories
When a question has been asked in the survey, where only one answer is appropriate, (e.g. How long ago did you give up smoking – 'Less than a month ago', '1 month to 1 year ago', 'More than a year ago'), results are stored as a single-code response category in one variable.

Multi-code response categories
Some questions enable the respondent to provide more than one answer and the interviewer is instructed to 'code all that apply' (e.g. last month, which of these websites have you visited? - 'Facebook', 'Flickr', 'YouTube' etc).

In such cases there will be a variable for each answer category. For example, 'Facebook', 'Flickr' and 'YouTube' will all have a separate variable and be coded as binary variables (i.e. 1 if the respondent has selected the answer, and 0 if the respondent has not selected the answer). The variable names for multi-code response categories consist of the same name plus a number at the end to represent the answer – i.e. 1 for the first possible answer, 2 for the second possible answer etc.

In the above example the following system is applied:
IntSocTyp1 = Facebook
IntSocTyp2 = Twitter
IntSocTyp3 = Flickr
IntSocTyp4 = YouTube
IntSocTyp5 = None of these
IntSocTypDk = Don't know
IntSocTypRef = Refused

Other multi-code variables follow a similar format.

3.3 Derived variables
In addition to the questions directly asked of respondents, the data file also includes a series of derived variables. Derived variables are those which are calculated from questionnaire variables (raw data) and/or data from non-survey sources in order to enable easier and more detailed analysis of the datasets. Some relate to the household, some to the area, and others to individuals.

For example, the 'local authority' variable is derived from the address information and the 'Tenure' variable is derived from the respondent's answer to questions about their home.
All derived variables have the letters 'Dv' at the start of their variable name.

A full list of derived variables can be viewed in the variable catalogue – in many cases these are the best variables to use for cross-analysis purposes. The details of the derivation for individual variables can be obtained on request by emailing the National Survey team at surveys@gov.wales.

3.4 Missing values
All respondents have a value recorded against them for each variable. However in some cases the value is referred to as a "missing value" code. When carrying out analyses, missing values are typically excluded. More information on the types of missing value codes in the dataset is given below.

Not selected in subsample
For some topics within the survey, the results were only required at a national or regional level. Therefore in order to use the survey time most effectively, some question sets have been asked of a random subsample of respondents.

When looking at the questionnaire, the subsampling is shown in the questionnaire routing at the start of each subsampled topic as shown below:

IntSocTyp

NEW **In the last month, which, if any, of the following websites have you used or visited?**
SHOWCARD 18

MULTICODE

1	Twitter 1	
2	Facebook 2	
3	Flickr 3	[Module ends]
4	YouTube 4	
5	None of these 5	
	Don't know (SPONTANEOUS ONLY) -9	

Where this routing appears, it applies to the rest of the questions in that module unless otherwise indicated. More details on which topics are subsampled are given in Annex A of the questionnaire. A different set of weights will be required to apply to the questions that have been subsampled. The appropriate weight to use is added to the label of the subsampled variables in the dataset. More information on these can be found in the section on weights.

Questions not asked due to routing
Not all questions are applicable to all respondents. It is possible that a whole section of the questionnaire is not suitable for some respondents (e.g. questions for parents with children at school). In other cases it may be an individual question that is not appropriate, (e.g. a respondent has stated that they don't use the internet, so they are therefore not asked about the websites they visited). In either of these cases their 'answers' are given the value label -8, 'Question not asked due to routing'.

Interview terminated early
A respondent may have decided to end the interview before the last question is reached. In such a circumstance, their answers up to that point are included in the data set and from that point onwards all of their 'answers' are given the value label -98, 'Interview terminated early'.

Don't know and refusal codes
For the majority of questions respondents are not explicitly given the options 'don't know' or 'refuse'. Respondents may however spontaneously respond "don't know" or refuse to answer a question. These are recorded by the interviewer and are classified as follows:
    Refused -99
    Don't know -9.

All of the missing values identified above are coded as 'missing values' in SPSS.

The screenshot below shows how three response options for a variable have been coded as missing values and how they are displayed when running a frequency in SPSS:

Public service communications - Visited any of these websites for personal purposes in the last 12 months [SampleIntWeight]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	1425	13.6	64.3	64.3
	No	791	7.5	35.7	100.0
	Total	2216	21.1	100.0	
Missing	Interview terminated early	17	.2		
	Not selected in sub-sample	8246	78.6		
	Don't know (SPONTANEOUS ONLY)	14	.1		
	Total	8277	78.9		
Total		10493	100.0		

4. Analysing the data

4.1 Understanding the weights

When analysing the data, the intention is often to make statements about the characteristics, attitudes or experiences of all adults in Wales – not just the people in the sample. The survey weights can help you to do this, by ensuring that the sample mirrors the age and sex profile of the general population of people and therefore that the results are representative of the general population. However, before undertaking any analysis, it is important to understand which weight you need to use for which analysis.

For some topics within the National Survey, the results were only required at a national or regional level. Those topics have therefore been asked only of a random subsample of respondents. As each of these sets of sub-sampled questions were asked of a different sub-sample, each sub-sample requires a different weight in order to make sure that it reflects the age and sex profile of the general population.

Below is a summary of the weights that are found in the respondent file, what they do and when they should be applied. For more information on how the weights were constructed please refer to the relevant Technical report.

Variable name	What the weight does	When to use this weight
SampleAdultWeight	Adjusts the sample to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size.	**This is the weight to use for most analyses about people (rather than households)**. It should be used when the questions have been asked of everyone and not a subsample.
SampleHhWeight	Adjusts the sample to have the same tenure/local authority distribution as households in Wales; provides a weighted sample size equal to the unweighted sample size.	**This is the weight to use for results about households** (only used occasionally where question refers to the household e.g. does household have internet access)
SampleDentalWeight	Adjusts the random sub-sample of approx. 8,000 asked the Dental question set to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size.	This is the weight to use when analysing the questions in the Dental health subsample

Variable name	What the weight does	When to use this weight
SampleHospitalWeight	Adjusts the random sub-sample of approx. 10,000 asked the Hospital question set to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size.	This is the weight to use when analysing the questions in the Hospital subsample
SampleGPWeight	Adjusts the random sub-sample of approx. 6,600 asked the GP services question set to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size.	This is the weight to use when analysing the questions in the GP services subsample
SamplePopHlthWeight	Adjusts the random sub-sample of approx. 6,000 asked the population health question set to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size.	This is the weight to use when analysing the questions in the Diet, Alcohol, Smoking, & Physical activity subsample
SampleOnlineHlthWeight	Adjusts the random sub-sample of approx. 6,000 asked the online health questions to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size.	This is the weight to use when analysing the questions in the Online information on health and care subsample
SampleSatWGWeight	Adjusts the random sub-sample of approx. 4,000 asked the Welsh Government satisfaction question questions to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size.	This is the weight to use when analysing the questions in the Welsh government satisfaction subsample
SampleTravelWeight	Adjusts the random sub-sample of approx. 2,000 asked the Active travel questions to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size	This is the weight to use when analysing the questions in the satisfaction with Active travel subsample

Variable name	What the weight does	When to use this weight
SampleTaxWeight	Adjusts the random sub-sample of approx. 2,000 asked the Tax devolution question set to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size.	This is the weight to use when analysing the questions in the Tax devolution subsample
SampleHeritageWeight	Adjusts the random sub-sample of approx. 6,000 asked the Heritage question set to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size.	This is the weight to use when analysing the questions in the Heritage subsample
SampleArchivesWeight	Adjusts the random sub-sample of approx. 2,000 asked the Archives question set to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size.	This is the weight to use when analysing the questions in Archives subsample
SamplePunishWeight	Adjusts the random sub-sample of approx. 3,000 non-parents and all parents asked the Physical punishment of children question set to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size.	This is the weight to use when analysing the questions in Physical punishment of children subsample
SampleMatDepWeight	Adjusts the random sub-sample asked the borderline material deprivation question set to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size..	This is the weight to use when analysing the questions in the borderline material deprivation subsample

Variable name	What the weight does	When to use this weight
SamplePenDepWeight	Adjusts the random sub-sample asked the borderline pensioner deprivation question set to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size.	This is the weight to use when analysing the questions in the borderline pensioner deprivation subsample
SampleChildDepWeight	Adjusts the random sub-sample asked the borderline child deprivation question set to have the same age/sex distribution as the general population; provides a weighted sample size equal to the unweighted sample size.	This is the weight to use when analysing the questions in the borderline child deprivation subsample
popwtesp	Adjusts the sample to have the same age/sex distribution as the European Standard Population	When wishing to age-standardise your results – e.g. for population health results

4.2 Analysis of respondents

Many questions in the survey have not been asked of a sub-sample. When analysing a question that has not been subsampled, if the topic relates to people (rather than households) then you should apply the weight **SampleAdultWeight** to produce proportions or means.

For those questions that have been asked of a subsample, you should apply the weight that applies to that particular subsample. In order to assist data users to use the correct weight, the variable names of the subsampled variables contain the appropriate weight in square brackets.

Use SampleGPWeight when analysing this variable

Use SampleHospWeight when analysing this variable

GpDigResp	GP - Treated with dignity and respect [SampleGPWeight]
GpOverSat	GP - Overall satisfaction with care received [SampleGPWeight]
HspHadApp	Hospital - Had an appointment at an NHS hospital in last 12 months [SampleHospWeight]
HspPatType	Hospital - Type of patient [SampleHospWeight]
HspKnewInfo	Hospital - At the start of the appointment, the health professional knew all the relevant information about
HspDigResp	Hospital - Treated with dignity and respect [SampleHospWeight]
HspOverSat	Hospital - Overall satisfaction with care received [SampleHospWeight]
AmbContact	Ambulance service - Last 12 months, personally contacted the emergency ambulance service in Wales

No square brackets, therefore it's not-subsampled - use SampleAdultWeight

All of the weights included in this dataset weight the results up to the size of the sample that were asked the question (i.e. 12,393 for non-sub-sampled questions in 2019-20). If instead you would like to produce figures on the estimated numbers of people or households in Wales with a particular characteristic, please contact: surveys@gov.wales

4.3 Analysis of households
When producing results for households, you should apply the weight **SampleHhWeight** to produce the result for the number of households in the sample and return the proportions and means for any given variable.

Household analysis should only be carried out when the intention is to make statements about the characteristics or experiences of households in the sample. For example, for questions on household internet access or household tenure (renting, owner-occupier, etc.).

4.4 Analysis of all people in households
If you would like to produce results about people living in households with certain household characteristics, we advise that you contact the National Survey team at surveys@gov.wales to obtain the 'All people file' for this analysis.

To illustrate, if you wanted to know the percentage of people aged 60+ who live in a household that has internet access it would be possible to take the figure from the 'respondent file'. However, for these kinds of questions, the answers will apply to everyone in a household, not just the respondent. Therefore, given that other household members are excluded from the respondent file, it's more appropriate to take the figure from the 'All people' file.

4.5 Analysis of multiple response variables
If you are using SPSS to analyse the data, the questions that have more than one possible response category (multi-code), have been defined as multiple response sets. Therefore when you use the 'Custom tables' command a selection of additional variables appear at the bottom of the variable list, which are 'multiple response sets'. The screenshot below is an example of how they look in SPSS Custom Tables:
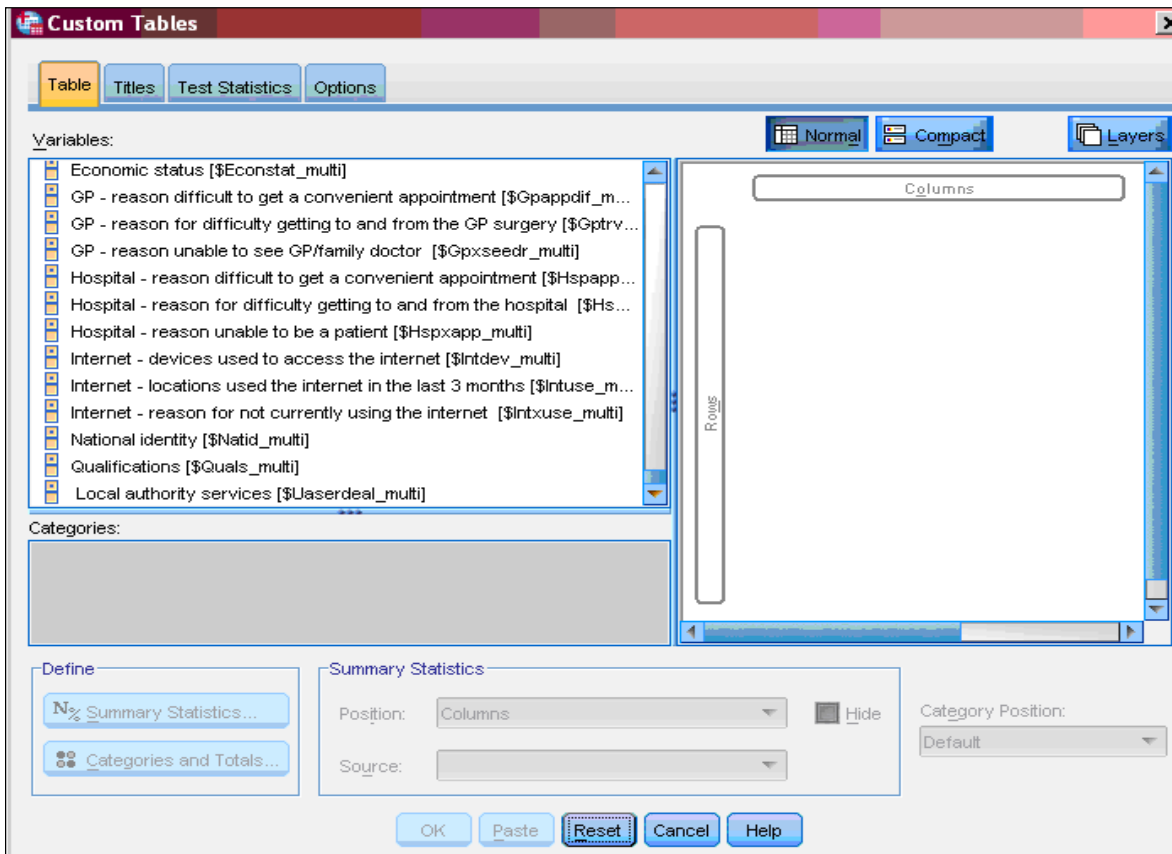
It is advisable to use the multiple response sets for cross-tabulating multi-code response categories. This is because answer categories such as 'don't know' will be removed from the base, to provide you with percentages of those who gave a genuine answer.

5. Methodological limitations

It should be noted that the figures derived from the National Survey are estimates. As with any sample survey, the survey estimates are subject to sampling and non-sampling error and a range of other methodological limitations due to the design of the survey.

Sampling error arises because the estimates are based on a random sample of the population rather than the whole population. Non-sampling error covers all other differences between the survey estimates and true population (e.g. coverage error, non-response error, measurement error and processing error). More information about sampling and non-sampling errors can be found in the National Survey Quality report.

The National Survey sample is stratified at local authority level. Whilst the proportion sampled in each local authority is proportional to the amount of people living in the local authority, there is a minimum effective sample size of 250 addresses in each local authority regardless of its population size, plus the sample was boosted for authorities like Powys to ensure that results are robust at Local Health Board level. Therefore there was a different probability of being selected in different local authorities. One of the effects of using this complex design is that standard errors for the survey estimates are generally higher than the standard errors that would be derived from a simple random sample of the same size. When producing confidence intervals, which are based on standard errors, it is therefore advisable to take the survey design into account. If the survey design is not taken into account then the confidence

intervals will appear smaller than they should, suggesting that the results are more precise than they actually are. More information can be found in the [Technical report](#).

The sample was also clustered by 'sampling points' within each quarter, but unclustered over the whole year. Therefore, for datasets of the whole year there is no need to account for clustering. However, analysis covering less than a full survey year should take clustering into account. Please [contact us](#) to discuss if you wish to do this.

To produce confidence intervals it is necessary to use a statistical package that can take account of the complex design of the survey. Stata and SPSS (using the complex samples add on module) both take account of the survey design. Chapter 6 provides some Stata syntax to carry out analysis which accounts for the survey design.

6. Example syntax for analysis

In this section, we provide some example syntax for two common software packages, Stata and SPSS, to guide some of the more common types of analysis that users may wish to undertake.

The syntax for both packages will show how to weight the data, produce simple cross tabulations of the data and how to recode variables into new categories in both Stata and in SPSS.

We recommend any analysis conducted should take account of the complex survey design. We use Stata to do this and to calculate confidence intervals and coefficients of variance. Therefore the following section shows how to take the complex survey design into account before undertaking analysis, using Stata.

6.1 Stata
To take account of the survey design in Stata, it's important to use the 'Svy' command. Before producing any survey results, you must indicate, which weight you would like to use and how the sample is stratified and clustered.

This code below tells Stata the survey design - that you wish to weight the data by 'SampleAdultWeight' and that the sample is stratified by local authority:

```
svyset [weight= SampleAdultWeight], strata (DvLA)
```

Once this syntax has been run, all subsequent analysis will take the complex sample design into account in producing standard errors / confidence intervals.

We remove common 'missing values', such as 'Don't' know' and 'Refused' from our variables when running analysis. To replace all 'missing values' with alphabetical letters, you can use the following syntax:

```
foreach v1 of varlist _all {
capture confirm numeric variable `v1'
 if !_rc {
     di "`v1'"
           replace `v1' = .r if `v1'==-99
           replace `v1' = .s if `v1'==-88
           replace `v1'= .q if `v1'==-8

     }
     }
```

To provide a survey result as a 'mean' (for continuous variables) use the following syntax:

```
svy: mean WbSatLife
```

To provide a survey result as a 'proportion' (for categorical variables) use the following syntax:

```
svy: proportion UaQualServ
```

To produce a cross-tabulation of a 'mean' or a 'proportion' use the following syntax:

```
svy: proportion UaQualServ, over (Gender)
svy: mean WbSatLife, over (Gender)
```

To recode variables into new variables or to combine categories you can use (or adapt) the following syntax:

```
label define agreecon 1 "Agree" 2 "Neither_Agree_ nor_Disagree" 3 "Disagree

foreach v1 of varlist UaQualServ UaComPerf UaPerfInfo UaInfDecs UaMreInvol {

quietly generate `v1'_1 =.
quietly replace `v1'_1 = 1 if (`v1'==1) | (`v1'==2)
quietly replace `v1'_1  = 2 if (`v1'==3)
quietly replace `v1'_1  = 3 if (`v1'==4) | (`v1'==5)
label values `v1'_1 agreecon

}
```

The syntax above takes a variable with 5 answer categories "Strongly Agree", "Tend to Agree", "Neither Agree nor Disagree", "Tend to Disagree" and "Strongly Disagree" (e.g. UaQualServ) and collapses them into a new variable with 3 categories "Agree", "Neither Agree nor Disagree" and "Disagree", with "_1" at the end of each new variable name (e.g. UaQualServ_1).

6.2 SPSS

We use SPSS to carry out basic analysis such as frequencies and cross tabs. The means and proportions produced by SPSS will match the Stata outputs. It is only when producing confidence intervals and standard errors that you need to take the complex sample design into account.

Before producing any survey results, you must indicate, which weight you would like to use. The following syntax tells the package that you wish to weight the data by 'SampleAdultWeight':

```
WEIGHT BY SampleAdultWeight.
```

To recode variables into new variables or to combine categories you can use the following syntax:

```
RECODE UaQualServ (1=1) (2=1) (3=2) (4=3) (5=3) (ELSE=Copy) INTO UaQualServ_1.
EXECUTE.

VARIABLE LEVEL UaQualServ_1(NOMINAL).
VALUE LABELS UaQualServ_1
  1.00 'Agree'
  2.00 'Neither Agree nor Disagree'
```

```
   3.00 'Disagree'.
EXECUTE.
```

The syntax above takes the variable UaQualServ, with 5 answer categories "Strongly Agree", "Tend to Agree", "Neither Agree nor Disagree", "Tend to Disagree" and "Strongly Disagree" and collapses them into a variable UaQualServ_1, with 3 categories, and labels them as "Agree", "Neither Agree nor Disagree" and "Disagree".

An alternative way of creating the same proportions would be to use the Custom Tables option in SPSS to create subtotals, which combines answer options 1 and 2 together, and 3 and 4 together, as shown in the following syntax:

```
CTABLES
  /VLABELS VARIABLES=UaQualServ DISPLAY=LABEL
  /TABLE UaQualServ [COUNT COMMA40.0, COLPCT.COUNT 'Col %' F40.1]
 /CATEGORIES VARIABLES=UaQualServ [1, 2, HSUBTOTAL='Agree', 4, 5,
HSUBTOTAL='Disagree', 3] TOTAL=YES POSITION=AFTER.
```

Running a basic frequency will produce a proportion for a variable:

```
FREQUENCIES VARIABLES=UaQualServ
  /ORDER=ANALYSIS.
```

The Custom Tables option produces crosstabs, for example:

```
* Custom Tables.
CTABLES
  /VLABELS VARIABLES=UaQualServnew Gender DISPLAY=LABEL
  /TABLE UaQualServnew [COLPCT.COUNT PCT40.1] BY Gender
  /CATEGORIES VARIABLES=UaQualServnew ORDER=A KEY=VALUE EMPTY=INCLUDE
  /CATEGORIES VARIABLES=Gender [1, 2] EMPTY=INCLUDE.
```

The advantage of using custom tables is that it removes any variables coded as missing values from your output.

7. Published results

Throughout the year the National Survey team publish results from the survey in a number of forms.

A statistical first release is published in each year (for 2019-20 this was published on 14th July 2020). This presents summary findings of interviews carried out in the preceding financial year. We also publish more detailed topic specific bulletins thereafter. All releases can be found on the National Survey web pages.

We also publish an interactive results viewer, where users can select the topics and questions that they are interested in. Here's a screenshot of the results viewer's topic page:

Once you have selected the questions of interest, it is possible to view the results in chart and table format, by various breakdowns. Where the same question has been asked in previous years it is possible to view the results across years (from 2012-13 onwards).

Both charts and tables show the results published with 95% confidence intervals, which take into account the design of the survey. We have also colour coded the survey results in the tables to give an indication of how wide the confidence intervals are and therefore how reliable the data is.

Here's an example of a chart on the results viewer:

Life satisfaction

Choose breakdown: General health ▼

How to use
Browse other topics
See technical info

Show / hide error bars | See in questionnaire | Show as table

General health, by Life satisfaction

Bar chart values: 8.2, 7.9, 7.3, 6.2, 4.8

General health
- Very good
- Good
- Fair
- Bad
- Very bad

Life satisfaction

Here's an example of a table on the results viewer:

Have a sense of community (DV)

Choose breakdown: Age (5 year bands) (DV) ▼

How to use
Browse other topics
See technical info

See in questionnaire | Show as chart

Have a sense of community	Age (5 year bands)														
	16-24			25-44			45-64			65-74			75+		
	%	Lower CI	Upper CI	%	Lower CI	Upper CI	%	Lower CI	Upper CI	%	Lower CI	Upper CI	%	Lower CI	Upper CI
Yes	39	34	43	43	41	46	51	49	53	59	57	61	69	66	71
No	61	57	66	57	54	59	49	47	51	41	39	43	31	29	34

Sample size: 10,300

Margin of error:
- Precise (within +/- 2%)
- Fairly precise (+/- 2% to +/- 4%)
- Less precise (+/- 4% to +/- 6%)
- Least precise (more than +/- 6%)
- - Results based on < 30 people, not shown

8. Terms and definitions

Confidence intervals
A confidence interval gives a range within which the true value is likely to fall. In 95% of survey samples, the 95% confidence interval will contain the 'true' figure for the whole population (that is, the figure we would get if the survey covered the entire population). In general, the smaller the sample size the wider the confidence interval.

Demographic information
Information about people's characteristics (e.g. age, gender, ethnicity)

Complex sample
As the National Survey is stratified at local authority level, with different probabilities of selection for people living in different local authorities (i.e. it is a complex sample design), the standard errors for the survey estimates are generally larger than the standard errors that would be derived from a simple random sample of the same size.

Fieldwork
The collection of primary information; for the National Survey, this is the process of interviewing respondents.

Household
A household is defined as one person living alone or a group of people (not necessarily related) living at the same address with common housekeeping – that is, sharing a living room or sitting room or at least one meal a day.

ONS mid-year estimates
The annual population estimates produced by ONS.

Syntax
Computer language used to instruct a software package how to analyse report on the data.

Respondent
The adult aged 16 and over who is randomly selected to answer the survey questions.

Response rate
The percentage of people who actually took part in the survey, out of those who were selected to take part (the sample).[1]

Royal Mail's Small User Residential Postcode Address File
A dataset maintained by the Royal Mail which contains all known residential delivery points.

Sample
A subset of the population, selected to take part in a survey.

Sampling point
Sampling points are small areas (clusters), where a sample of addresses has been selected to take part in the survey.

Stratification
Dividing the population into subgroups for sampling.

---

[1] For more detail on how response rates are calculated, see the Technical report.

Sub-sample

A random subset of the 'sample', selected to take part in a survey.

Weighting

The survey results are weighted to reduce bias by compensating for any groups that are more likely to respond (e.g. older people) or less likely (e.g. younger people). Weighting ensures that the age and sex distribution of the responding sample matches that of the population of each local authority in Wales. In this way, the results are made more representative of the general population.