Welsh Health Survey

2009

WHS simplified introductory dataset

User guide

Content of this guide

This guide describes a simplified introductory dataset for the Welsh Health Survey (WHS). The dataset contains a selection of key variables for adults – with fewer variables than the main WHS dataset, it is easier to use and may be particularly helpful for new users.

Background to the Welsh Health Survey

WHS covers a range of health related issues for people living in Wales, including health status, lifestyle and health behaviours, and use of health services. WHS in its current form began in 2003. The survey is funded by the Welsh Assembly Government (WAG) and is carried out by the National Centre for Social Research (NatCen).

Each year, data is collected from a representative sample of around 15,000 adults and 3,000 children in Wales, using a short household interview and individual self-completion questionnaires completed by all adults and up to two children in each household.

The content of the adult questionnaire has remained fairly stable since 2003, covering:
- health status, including general health and wellbeing, illness, eyesight and hearing
- health-related behaviours such as smoking, drinking, fruit and vegetable consumption, Body Mass Index and physical activity
- health services including GP, hospital visits and use of other services like dentists, opticians and NHS Direct.

Variables on the dataset

The simplified introductory dataset covers adults only. The main WHS adult dataset deposited with the UK Data Archive contains over 400 variables. This dataset is a slimmed down version containing around 60 key variables – this makes it easier to use, particularly for new or occasional users. All adult respondents are included in the dataset, and so any results produced from it should match those derived from the more detailed dataset and those shown in the WHS annual report.

A list and description of the variables is given in appendix 1.

Weighting the dataset

Respondents to the survey did not have an equal chance of selection (for instance, the probability of selecting an address varied by local authority). There were also differences in response rates between groups. Weights were calculated for WHS to correct for this. The final weight for adults is the variable **wt_adult** - this should be used when analysing the data to avoid bias in the estimates produced.

Missing values

Missing values occur when a respondent does not answer a particular question. This may happen for a number of reasons, including refusal or inability to answer a question, or where it does not apply to them. The following conventions are used in the dataset:

-1   not applicable: signifies that a particular variable did not apply to the respondent (eg for men, where the question was only asked for women)

-8 or -9   don't know, can't say, no answer, refused

It is common practice to exclude missing values from analyses, although they may occasionally be of interest.

Frequencies

As a first step, it is often useful to look at the 'frequencies' of the variables which are of interest in an analysis. Frequency tables for a selection of variables are at appendix 2, and may be a useful check for the new user.

Other useful sources of information

The ESDS webpages for WHS include some additional information, including an overview of the main datasets and documentation.
http://www.esds.ac.uk/government/whs/

There is also a series of ESDS guides, including introductory guides to using the health surveys and to the analysis packages SPSS and Stata.
http://www.esds.ac.uk/government/resources/

Additional information about WHS, including background, content, results and technical report, is available on the Welsh Assembly Government webpages.
http://new.wales.gov.uk/topics/statistics/theme/health/health-survey/?lang=en

Appendix 1 – list of variables

Variable	Label *(& any additional details)*
Serial numbers	
archpsn	Scrambled Person Serial
archhsn	Scrambled Household Serial
Household classification	
tenure3	(D) 3 class Housing Tenure
nssec3	(D) NS-SEC 3 classes
whsyear	whs year
Individual classification	
sex	(D) Sex
age5yrm	(D) 5 Year Age Bands with 75+ merged
carer	(D) Whether a carer
Employment and education	
ecstat3	(D) Economic status - 3 classification
qualhi	(D) Level of Highest qualification *(3 categories)*
Anthropometric (body) measurements	
htcm	(D) Height : in cm computed from feet/inches if necessary
wtkg	(D) Weight : in kg - computed from stones/pounds if necessary
bmilev2	(D) Body Mass Index classification (excl pregnant women) *(4 categories)*
bmiowob2	(D) BMI Overweight or obese (excl pregnant women)
bmiobes2	(D) BMI Obese (excl pregnant women)
Health service use	
gpbi	(D) Talked to GP in last 2 weeks – binary
cas12mbi	(D) Attended casualty in last 12 months – binary
out12mbi	(D) Outpatient in last 12 months – binary
inpatbi	(D) Inpatient in last 12 months – binary
denbi	(D) Visited a dentist in the last 12 months – binary
optibi	(D) Visited an optician - binary
pharbi	(D) Used a pharmacist - binary
Medicines	
prescmbi	(D) Regular prescribed medication - binary
Illnesses	
strokbi	(D) Ever had stroke - binary
hbpbi	(D) Currently treated for high blood pressure - binary
heart	(D) Any heart condition
asthmabi	(D) Currently treated for asthma - binary
resp	(D) Any respiratory condition currently treated
mental	(D) Any mental condition currently treated
arthbi	(D) Currently treated for arthritis - binary
diabbi	(D) Currently treated for diabetes - binary
illoth	Other chronic illness - currently treated
chronic	(D) Any chronic illness currently treated
Accidents	
accbi	(D) Hospital for accident, injury or poisoning in last 3 months - binary
Other conditions	
eyesight	(D) Eyesight difficulty - binary
hearbi	(D) Difficulty with hearing - binary
teethbi	(D) How many natural teeth do you have - binary
Health and well-being	
fphth	(D) Fair or poor health - binary
sf36pcs	(D) SF36 Physical Component Score *(Note: for this variable, look at mean score)*
sf36mcs	(D) SF36 Mental Component Score *(Note: for this variable, look at mean score)*
lltibi	(D) Limiting long-term illness - binary

Smoking	
smok	Smoking status
smokec	(D) Currently smoke either daily or occasionally
smokstat	(D) Smoking status *(3 categories)*
exinnsm	(D) Exposed to smoke indoors - non-smokers only
alcohol	
freqalc	How often had an alcoholic drink in last 12 months
alc5	(D) Maximum drank last week (incl never drinks) *(5 categories)*
alcagbi	(D) Maximum daily alcohol consumption: above guidelines - binary
alcbibi	(D) Maximum daily alcohol consumption: binge - binary
Fruit & vegetables	
fv5aday2	(D) No of portions of fruit and vegetables eaten yesterday *(grouped)*
fv52	(D) Eaten 5+ fruit or veg the previous day – binary
Physical activity	
exercise	(D) At least 30 mins mod/vigorous exercise on 5+ days
exergrp	(D) No of days at least 30 mins mod/vigorous exercise *(grouped)*
Weighting	
wt_adult	wt_adult *(the variable used to weight the dataset for analyses)*

A copy of the questionnaire showing exact question wording and the full range of questions is available in the documentation for the main dataset or on the Welsh Assembly Government WHS webpages.

A note on derived variables

A number of the variables in the dataset have labels with the prefix "(D)". This indicates that they are derived variables, and are based on other variables. Some of the common types of derived variables are as follows:

- recoding a variable with answer options yes / no to a binary 1/0 scheme to make subsequent analysis easier (eg the variable *hbpbi* shows whether a respondent reports being treated for high blood pressure with code 1 signifying 'yes' and code 0 signifying 'no');
- grouping answer options to produce less detailed summary variables (eg the variables *smokec* and *smokstat* show less detailed information about smoking status than the original variable *smok*);
- combining answers from more than one variable (eg the variable *exercise* shows whether a respondent meets the weekly guidelines for physical activity, and is based on answers to a series of different questions about activity undertaken during the past week).

A full specification for all derived WHS variables is provided in the documentation for the main dataset.

Example syntax

The examples below show the syntax used to generate 10 year age bands (AGE10YRM) and three broader age bands (AGE3).

AGE10YRM: (D) 10 year age bands with 75+ merged

1	16-24
2	25-34
3	35-44
4	45-54
5	55-64
6	65-74
7	75+

AGE3: (D) Age in 3 bands with 65+ merged

1   16-44
2   45-64
3   65+

Appendix 2 – sample frequency tables

These frequency tables show information for a selection of variables. They are based on weighted data. They can be produced in SPSS using the commands:

> *weight by wt_adult.*
> *freq varname.*

(where *varname* is the variable of interest)

hbpbi (D) Currently treated for high blood pressure - binary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	12397	77.4	80.1	80.1
	1 Yes	3078	19.2	19.9	100.0
	Total	15475	96.6	100.0	
Missing	-9 No answer/refused	543	3.4		
Total		16018	100.0		

heart (D) Any heart condition

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	13599	84.9	91.2	91.2
	1 Yes	1317	8.2	8.8	100.0
	Total	14917	93.1	100.0	
Missing	-9 No answer/refused	1101	6.9		
Total		16018	100.0		

resp (D) Any respiratory condition currently treated

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	13155	82.1	86.5	86.5
	1 Yes	2050	12.8	13.5	100.0
	Total	15205	94.9	100.0	
Missing	-9 No answer/refused	813	5.1		
Total		16018	100.0		

mental (D) Any mental condition currently treated

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	13781	86.0	90.0	90.0
	1 Yes	1531	9.6	10.0	100.0
	Total	15312	95.6	100.0	
Missing	-9 No answer/refused	706	4.4		
Total		16018	100.0		

arthbi (D) Currently treated for arthritis - binary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	13431	83.8	87.0	87.0
	1 Yes	2013	12.6	13.0	100.0
	Total	15444	96.4	100.0	
Missing	-9 No answer/refused	574	3.6		
Total		16018	100.0		

diabbi (D) Currently treated for diabetes - binary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	14841	92.7	93.7	93.7
	1 Yes	990	6.2	6.3	100.0
	Total	15831	98.8	100.0	
Missing	-9 No answer/refused	187	1.2		
Total		16018	100.0		

lltibi (D) Limiting long-term illness - binary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	11325	70.7	72.6	72.6
	1 Yes	4282	26.7	27.4	100.0
	Total	15607	97.4	100.0	
Missing	-9 No answer/refused	411	2.6		
Total		16018	100.0		

smok Smoking status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 I smoke daily	3126	19.5	19.8	19.8
	2 I smoke occasionally but not every day	647	4.0	4.1	23.9
	3 I used to smoke daily but do not smoke at all now	2425	15.1	15.3	39.2
	4 I used to smoke occasionally but do not smoke at all now	1877	11.7	11.9	51.1
	5 I have never smoked	7730	48.3	48.9	100.0
	Total	15806	98.7	100.0	
Missing	-9 No answer/refused	200	1.3		
	-8 Don't know	12	.1		
	Total	212	1.3		
Total		16018	100.0		

smokec (D) Currently smoke either daily or occasionally

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	12032	75.1	76.1	76.1
	1 Yes	3773	23.6	23.9	100.0
	Total	15806	98.7	100.0	
Missing	-9 No answer/refused	200	1.3		
	-8 Don't know	12	.1		
	Total	212	1.3		
Total		16018	100.0		

alcagbi (D) Maximum daily alcohol consumption: above guidelines - binary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	8459	52.8	55.4	55.4
	1 Yes	6818	42.6	44.6	100.0
	Total	15277	95.4	100.0	
Missing	-9 No answer/refused	741	4.6		
Total		16018	100.0		

alcbibi (D) Maximum daily alcohol consumption: binge - binary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	11077	69.2	72.5	72.5
	1 Yes	4201	26.2	27.5	100.0
	Total	15277	95.4	100.0	
Missing	-9 No answer/refused	741	4.6		
Total		16018	100.0		

fv52 (D) Eaten 5+ fruit or veg the previous day - binary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	10034	62.6	64.7	64.7
	1 Yes	5480	34.2	35.3	100.0
	Total	15514	96.9	100.0	
Missing	-9 No answer/refused	504	3.1		
Total		16018	100.0		

exercise (D) At least 30 mins mod/vigorous exercise on 5+ days

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	11026	68.8	70.8	70.8
	1 Yes	4540	28.3	29.2	100.0
	Total	15566	97.2	100.0	
Missing	-9 No answer/refused	452	2.8		
Total		16018	100.0		

bmiowob2 (D) BMI Overweight or obese (excl pregnant women)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	6267	39.1	42.8	42.8
	1 Yes	8367	52.2	57.2	100.0
	Total	14634	91.4	100.0	
Missing	-9 No answer/refused	1381	8.6		
	-8 Don't know	3	.0		
	Total	1384	8.6		
Total		16018	100.0		

bmiobes2 (D) BMI Obese (excl pregnant women)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	11523	71.9	78.7	78.7
	1 Yes	3111	19.4	21.3	100.0
	Total	14634	91.4	100.0	
Missing	-9 No answer/refused	1381	8.6		
	-8 Don't know	3	.0		
	Total	1384	8.6		
Total		16018	100.0		

gpbi (D) Talked to GP in last 2 weeks - binary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No	12997	81.1	82.0	82.0
	1 Yes	2852	17.8	18.0	100.0
	Total	15849	98.9	100.0	
Missing	-9 No answer/refused	169	1.1		
Total		16018	100.0		