

Teaching resource: Analysing ethnic differences in health using data from Understanding Society

CITATION FOR THE TEACHING DATASET

University of Essex. Institute for Social and Economic Research.
(2020). Understanding Society: Ethnicity and Health Teaching Dataset
Wave 1, 2009-2010. [data collection]. UK Data Service. SN: 8465,
<http://doi.org/10.5255/UKDA-SN-8465-2>

CITATION FOR THE TEACHING RESOURCE:

Nandi, Alita and Wiltshire, Deborah. (2020). "Teaching Resource:
Analysing ethnic differences in health using data from Understanding
Society"

AUTHORS

Alita Nandi
University of Essex
Deborah Wiltshire
University of Essex



INTRODUCTION

In this worksheet we will help you to find out how to use data from a large scale survey, *Understanding Society*, to analyse the relationship between ethnicity and health. We will do this in stages. In **Part One** we will:

- Introduce the survey *Understanding Society*
- Explain the structure of this survey data
- Show you how to find out what information is collected in the survey

In **Part Two** we will introduce a data exercise in which we will show you how to do some simple analysis using data from this survey and the statistical software package, Stata.

There is a large body of evidence on the association between health and ethnicity (for a recent review see Toleikyte and Salway 2018). Some of the key reasons identified for ethnic differences in physical and mental health, including general wellbeing, are socio-economic status, health related behaviour and, experiences of harassment and discrimination (Nazroo 2003). A substantial proportion of ethnic minorities are also migrants, that is, those who were born in another country and who came to the UK at some point. Studies have also found that even though migrants tend to be younger and healthier than the UK born population (as these are the kinds of people who are more likely to move), after arriving to the UK their health often deteriorates, perhaps due to the anxiety and stress of living in a new place that often brings problems of unfamiliarity with the systems and procedures prevalent in the new country, lack of friendship and social support networks, and in some cases language difficulties.

In this exercise you will use data from Understanding Society to measure differences in health across ethnic groups and differences in health across sub-groups within each ethnic group (where sub-groups are based on sex, age, region of residence, country of birth).

PART ONE: WHAT IS UNDERSTANDING SOCIETY?

The survey

Understanding Society: the UK Household Longitudinal Study (UKHLS) is a survey of individuals living in private households in the UK. This study started in 2009 with individuals living in around 30,000 households across the UK. Individuals living in these sampled households are eligible for interviews every year. But remember not everyone who is eligible participates in a survey. Those who do participate and answer the questions asked by the interviewer are referred to as respondents. The purpose of the study is to allow researchers to investigate and understand how the lives of the UK's population are changing over time.

Every year **respondents** are asked questions about various aspects of their lives – their education, employment or other labour-market activity, their partnership and marital status, how many children they have and of what ages, their income and wealth, their health and wellbeing, their attitudes and values.

The first time this sample was interviewed is referred to as the first *wave*, the second time they are interviewed is the second *wave* and so on. In this study we will focus only on the data collected in the first wave. This means the analysis is 'cross-sectional', or a snapshot of one point in time, not longitudinal (based on change over time). In the next exercise we will look at changes over time.

Respondents - All 16+ year olds are eligible for these detailed interviews. But 10-15 year olds in these households are asked to fill a short questionnaire about issues that are relevant to them such as experiences of bullying, their health and wellbeing, their computer usage, their school experience, their educational and future aspirations.

Next you will look for information on health, ethnicity, age and sex on the survey's website. But before you do that you will need to know a few things about the structure of the data.

The structure: Data files, and variable and naming conventions

The information collected is made available in **data files**.

Items of information about respondents (e.g. gender or age) are called variables. The dataset consists of these variables

Data files are available in different formats: Stata, SPSS, TAB delimited.

for all respondents where each variable has a value. For example, if a respondent is 29 years old, there will be a variable called age which will have a value of 29 for this respondent. Some information like sex is not numeric but the information is represented as numbers. For example, in this dataset sex=1 means male and sex=2 means female. So, if this respondent is female then the value of the variable sex for this person will be 2.

Think of the data file as an excel spreadsheet where the rows consist of items of information about a *person* and the columns make up the list of *variables* (the items of information). Here is an example, of a data file we will use in this exercise, **A_INDRESP**

PIDP	A_HIDP	A_SEX_DV	A_AGE_DV
1	1001	2	29
2	1001	1	26
3	1001	2	3

Here each row refers to one person who is identified by the variable **PIDP**, which is a number that is unique for each person. The variable **A_HIDP** is a unique number assigned to each household. So, we can say that all three individuals in this table belong to the same household (as they all have the same value for **A_HIDP**). The variable **A_SEX_DV** captures the reported sex of the individual (where 1 means male and 2 means female) and **A_AGE_DV** captures the reported age in years of the individual at the time of the interview. So, from Table 1 we can say that there are 3 individuals living in this household, one of whom is a 26 year old man while the other two are women aged 29 and 3 years.

When people are interviewed again the next year the information collected are stored in another file, called **B_INDRESP**. Here is an example of that file,

PIDP	B_HIDP	B_SEX_DV	B_AGE_DV
1	2345	2	30
2	2345	1	27
3	2345	2	4

As you can see the root name of the file (INDRESP) has not changed but the letter prefix has changed from A to B. The variable names also follow the same naming convention: same root name but a different letter prefix for each wave.

*The variable **PIDP** does not change over time.* It is the unique number assigned to each person and is fixed over time. You will always be able to identify the person using this number. That is why this variable has *no wave prefix*.

As the second wave interview is a year later the age has increased by one year for each person.

So, if these people are interviewed again the following year their information will be stored in a data file **C_INDRESP** and will look like this.

PIDP	C_HIDP	C_SEX_DV	C_AGE_DV
1	9611	2	31
2	9611	1	28
3	9611	2	5

Looking for variables

A number of different questions are asked about respondents' physical and mental health and their subjective wellbeing. This information is then recorded in a number of variables in data files. In this section we will help you answer the following questions:

- What are the variable names?
- How will you look for them?
- Where will you look for them?

First go to the study website and click on the DATASET DOCUMENTATION tab at the top. It will take you to the [dataset documentation page](#) where you will see different data boxes.

The screenshot shows the 'DATA & DOCUMENTATION' section of the Understanding Society website. At the top, there's a navigation bar with links for TOPICS, DATA & DOCUMENTATION, RESEARCH, PARTICIPANTS, ABOUT US, and HELP & SUPPORT. Below this is a header image with the text 'DATA & DOCUMENTATION' and a sub-header 'User guides, fieldwork documents, questionnaires, technical reports.' A paragraph states: 'Data collected from the survey's thousands of participants is securely stored by the UK Data Service, from where researchers can access it online. All of the associated documentation is available here.' Below this are six boxes representing different data sources: Main survey, Health and biomarkers, Innovation panel, Ethnicity and immigration, Linked Data, and COVID-19. Each box contains a brief description of the data.

Which one should you choose? Click on “Main survey”. The other tabs refer to other surveys or datasets that are related to the main survey. Ignore those for now. You can see an overview of the study [here](#). For a more in-depth understanding of the study read the [user guides](#). For now just focus on finding health variables in the survey. There are different ways to do this search.

Method 1: Go to the [main survey page](#) and click on **Variable Search** on the right panel.

The screenshot shows the 'MAIN SURVEY' page of the Understanding Society website. The page has a blue header with the Understanding Society logo and navigation links. Below the header, there's a breadcrumb trail: 'Home > Data and documentation'. The main content area is titled 'Main survey' and contains a paragraph describing the study. To the right of the main content is a sidebar with a list of links: OVERVIEW, CITATION, SURVEY TIMELINE, USER GUIDES, VARIABLE SEARCH, SYNTAX, QUESTIONNAIRES, TECHNICAL REPORTS, FIELDWORK DOCUMENTS, LONG TERM CONTENT PLAN, and QUALITY PROFILE. The 'VARIABLE SEARCH' link is circled in red.

This will take you to the [variable search page](#) with a search box.



[TOPICS](#)
[DATA & DOCUMENTATION](#)
[RESEARCH](#)
[PARTICIPANTS](#)
[ABOUT US](#)
[HELP & SUPPORT](#)

MAIN SURVEY

Home > Data and documentation > Main survey > Variable search

[Variable search](#)
[Datafiles](#)
[Index terms](#)
[Questionnaire modules](#)

Variable search

Find the variables you need for your research by searching by variable name, by data file or by index term. If you're new to using Understanding Society our [Getting Started Guide](#) will help you start working with the dataset. [Our list of key variables for the analysis of individual response data](#) may also help you.

Variable	Label	Datafile	Waves
a2	arf qletter.2: how many floors are there at the address?	hhsamp	4, 5
a31	locked common entrance	hhsamp	4, 5
a32	locked gates	hhsamp	4, 5
a33	security staff or gatekeeper	hhsamp	4, 5
a34	entry phone access	hhsamp	4, 5
a35	none of these	hhsamp	4, 5
a36	unable to obtain information	hhsamp	4, 5
a4	arf qletter.4: on what floor of the building is the address's main entrance?	hhsamp	4, 5

Datafile

indall	402
child	473
egoalt	27

Show more

Index term

Caring	166
Child Care	369
Children	1093

Show more

Wave

Wave 9 (I, 2017-18)	4171
Wave 8 (H, 2016-17)	3329
Wave 7 (G, 2015-16)	3895

Show more

If you type **health** in the **Search** box and click enter, you will see the search results.

[Variable search](#)
[Datafiles](#)
[Index terms](#)
[Questionnaire modules](#)

Variable search

Find the variables you need for your research by searching by variable name, by data file or by index term. If you're new to using Understanding Society our [Getting Started Guide](#) will help you start working with the dataset. [Our list of key variables for the analysis of individual response data](#) may also help you.

Variable	Label	Datafile	Waves
health	long-standing illness or impairment	indresp	1, 2, 3, 4, 5, 6, 7, 8, 9
health_tm	start time for health module (seconds)	indresp	1
sf1	general health	indresp	1, 2, 3, 4, 5, 6, 7, 8, 9, BH09, BH14
scsf1	general health	indresp	2, 3, 4, 5, 6, 7, 8, 9
jbendreas7	Health reasons	indresp	7, 8, 9
nxtendreas7	Health reasons	indresp	7, 8, 9
reasend7	Health reasons	indresp	2, 3, 6
stendreas7	Health reasons	indresp	7, 8, 9

Search found: **1039 items**

health

[Clear search](#)

Datafile

indall	2
child	15
hresp	15

Show more

Index term

Caring	23
Child Care	2
Children	160

Show more

Wave

Wave 9 (I, 2017-18)	436
Wave 8 (H, 2016-17)	318

Analysing ethnic differences in health using data from Understanding Society

The search result will show over 1039 items! You can filter the search by **Datafile**, **Index Term** and **Wave** on the right side panel. Next to each variable you will see its **Label** which is a short description of the variable, the **Datafile** in which it appears and the **Wave** in which it was asked.

The variable **health** which is available in **indresp** looks like a possible variable for measuring general health. But its **Label** reads “long-standing illness or disability”. So, this is not a measure of general health. If you go down the list you will see a variable called **SF1** with the **Label**, “general health”.

Method 2

You can go back to the [variable search page](#) and now click on **Index Terms** tab at the top. This will take you to a page with different index terms or keywords including a few with **Health** in their title.



Index terms

Our Index Terms cover all the thematic areas in the Study. Use the Index to identify the variables most relevant to your research interests and to find other variables with related data throughout the dataset.

Caring	Financial Management: Loan Repayments	Incomes: Benefits and Allowances and Pensions
Child Care	Financial Management: Material Wellbeing	Incomes: Grants for Education
Children	Financial Management: Pensions	Incomes: Household Income
Coded variables	Financial Management: Personal Spending	Incomes: Rents, Savings, Investments
Cognitive ability	Financial Management: Problems	Incomes: Windfalls
Computers and Computing	Financial Management: Savings and Bank Accounts	Interview Characteristics and Conditions
Consents	Friendship	Key Linking Variable
Crime	Gender Roles	Leisure Activity
Derived Variables	Geographic Location	Life Events
Education: Background and attainments	Geographic Mobility	Marital and Cohabitation History
Education: Expectations	Harassment	Marital Status
Education: Recent Education and Training	Health behaviour	Neighbourhood and Residence
Employment History: Labour Force Status Spells	Health: Accidents, Illness	Newspaper Readership
Employment History: Reasons for Leaving and Taking Jobs	Health: Childrens Health	Parents and Children: Parenting Styles
Employment History: Size, Sector and Duties	Health: Drinking	Person Identifiers
Employment: Attitude to Work and Incentives	Health: Effect on Daily Life, Employment	Personality traits
Employment: Benefits Receipt	Health: Height and Weight	Physical Characteristics
Employment: Expectations	Health: Hospital and Clinic Use	Political Support and Behaviour
Employment: Hours Worked and Overtime	Health: Medical Consultations	Population weights: Cross-sectional analysis
Employment: Industrial and Occupational Classification	Health: NHS vs Private	Population weights: Longitudinal analysis
Employment: Labour Force Status	Health: Personal Health Condition	Psychological Characteristics
Employment: Length of Job Tenure	Health: Smoking	Relationship between Household Members
Employment: Not Working/Seeking Work	Health: Subjective Wellbeing	Religion
Employment: Prospects and Training	Health: Use of Health and Welfare services	Retirement
Employment: Second Job	Household Changes	Roles of Partners/Spouses: Domestic Work
	Household Composition	Sampling Factors

If you click on **Health: Personal Health Condition** you will be taken to a page with a large number of health related variables. If you look down the list you will find the variable, **SF1** again.

sf1	general health	indresp	1, 2, 3, 4, 5, 6, 7, 8, 9, BH09, BH14
ypffd	how often eat fast food, takeaways	youth	BH14, BH15, BH16, BH17, BH18
ypfrut	how often eat fresh fruit or veg	youth	BH14, BH15, BH16, BH17, BH18
ypfrutppd	portions of fresh fruit and vegetables per day	youth	1, 2, 4, 6, 8, 9

Method 3

A third option is to first find the variable name in the questionnaires and then to search for this variable via variable search box (as shown in Option 1). Questionnaires are the list of questions along response options, filtering or routing rules (i.e., who gets asked each question) that respondents were asked in the exact order in which they were asked. These are provided as pdf files. Go back to the [main survey page](#) and click on **Questionnaires** on the right panel.

Home > Data and documentation

Main survey

Understanding Society, the UK Household Longitudinal Study, is a longitudinal survey of the members of approximately 40,000 households (at Wave 1) in the United Kingdom. Households recruited at the first round of data collection are visited each year to collect information on changes to their household and individual circumstances. Interviews are carried out face-to-face in respondents' homes by trained interviewers or through a self-completion online survey. Young people aged 10-15 complete a youth questionnaire, whilst respondents aged 16 and over complete the adult survey.

The overall purpose of Understanding Society is to provide high-quality longitudinal data on subjects such as health, work, education, income, family, and social life to help understand the long-term effects of social and economic change, as well as policy interventions designed to impact upon the general wellbeing of the UK population. To this end, the Study collects both objective and subjective indicators and offers opportunities for research within and across multiple disciplines including sociology and economics, geography, psychology and health sciences.

The Understanding Society main survey sample consists of a large General Population Sample plus three other components: the Ethnic Minority Boost Sample, the former British Household Panel Survey sample and the Immigrant and Ethnic Minority Boost Sample.

The data from Understanding Society are available via the UK Data Service.

- OVERVIEW
- CITATION
- SURVEY TIMELINE
- USER GUIDES
- VARIABLE SEARCH
- SYNTAX
- QUESTIONNAIRES**
- TECHNICAL REPORTS
- FIELDWORK DOCUMENTS
- LONG TERM CONTENT PLAN
- QUALITY PROFILE

This will take you to a page with different tabs on the left-hand side, each tab referring to a different wave.

Questionnaires

Youth and Adult

We use different types of questionnaire depending on who is answering the question. Every section of the questionnaire, and each question, is answered voluntarily.

- Each adult participant, aged 16 and above, answers a questionnaire asking them about different aspects of their life.
- For households where there are children under the age of 10, parents or carers answer a set of questions about the children in their care.
- There is also a household information questionnaire which one adult in the house completes on behalf of the whole household.
- Young people aged 10-15 complete a short questionnaire looking at different parts of their life, after their parent or carer has given permission for them to complete it.

WAVE 11 (2019-2021)

WAVE 10 (2018-2020)

WAVE 9 (2017-2019)

WAVE 8 (2016-2018)

WAVE 7 (2015-2017)

WAVE 6 (2014-2016)

WAVE 5 (2013-2015)

WAVE 4 (2012-2014)

Wave 11

- [Mainstage - questionnaire](#)
- [Mainstage - youth self-completion questionnaire](#)

- OVERVIEW
- CITATION
- SURVEY TIMELINE
- USER GUIDES
- VARIABLE SEARCH
- SYNTAX
- QUESTIONNAIRES**
- TECHNICAL REPORTS
- FIELDWORK DOCUMENTS
- LONG TERM CONTENT PLAN
- QUALITY PROFILE

Click on Wave 1 and you will see three documents. You can look at all of them. But for this exercise, we suggest you click on **Mainstage – questionnaire (CAPI)** (CAPI is where an interview is undertaken using a laptop rather than paper questionnaire).

Once you open this file, you will see that the first page is the content page which list all the question **Modules**. A *module* is a group of questions on a particular topic.

UK Household Longitudinal Study Mainstage Questionnaire Wave 11, v01

Household Grid	1
Household Questionnaire	15
Individual Questionnaire	41
Demographics	43
Initial Conditions	48
Family Background	59
Ethnicity and National Identity	69
Language	74
Religion	77
Migration History	81
Partnership History	89
Fertility History	102
Health and Disability	115
Caring	135
Employment Status History	139
Current Employment	143
Employees	148
Self-Employment	160
Job Satisfaction	170

Click on the module that reads **Health & Disability** and it will take you to that section of the questionnaire. You will see all the questions included in this Module including **SF1**.

Finding information about variables without opening a datafile

You can find out some information about a variable from the questionnaire and more information from its online variable description page.

In the questionnaire, just below the variable name you will see the exact text of the question that was asked (marked **Text**) and below that you will find Interviewer Instructions if any. Just below that are the response options, and a field marked **Universe** which shows the filtering or routing rule, that is, who gets asked that question.

If you click on **SF1** on the [variable search page](#) (using Method 1 or 2) you will arrive at the [variable page for SF1](#) where you will find a lot of information about the variable.

The screenshot shows the 'Variable search' tab with the following content:

- Variable label or a short description:** sf1, general health
- The data file in which you will find this variable:** Datafile, indresp
- Index Terms associated with the variable:** Index term, Health: Personal Health Condition
- The text of the question:** proxy_w9.sf1, General health, Text: In general, would you say NAME's health is...
- Who was asked this question?:** Universe: if GRIDVARIABLES.modetype = 1 (Mode is face-to-face) and if IProxy = 1 (Able to do a proxy interview for respondent)

Annotations (red arrows) point from the text on the right to the corresponding elements in the interface:

- Variable label or a short description → sf1
- The data file in which you will find this variable → indresp
- Index Terms associated with the variable → Health: Personal Health Condition
- The text of the question → Text: In general, would you say NAME's health is...
- Who was asked this question? → Universe: if GRIDVARIABLES.modetype = 1 (Mode is face-to-face) and if IProxy = 1 (Able to do a proxy interview for respondent)

Important information about a variable is provided in a **Variable Note**. As you can see here, the **Universe** shows that the question was asked of proxy respondents only. But you can see from the **Variable Note** that in the first wave this question was asked of everyone. From onwards the second wave it was decided to ask this question within the self-completion questionnaire and that variable was named **SCSF1**. This question in its current form continued to be asked of the proxy respondents. So, the **Universe** here indicates that this was being asked of proxy respondents. If you check the Wave 1 questionnaire, you will see that the **Universe** indicates it was asked of everyone.

Proxy respondent and proxy questionnaire: When a person cannot give an interview, their spouse or adult children are asked if they would complete a short questionnaire, focussing on factual questions only, on their behalf. This is the proxy questionnaire and the respondent on whose behalf the interview is done is the proxy respondent.

Below this you will see wave the cascading wave blocks with the title **Wave specific data**. If you click on a specific wave you will see the frequency of the variable for that wave. By default the most recent wave in which the variable was asked will appear (and only the waves in which the variable was asked will appear). In this case it is Wave 9.

Wave specific data

Wave 9				<i>l_sf1, l_indresp (2017-2018)</i>
Frequencies				
Value label	Value	Absolute frequency	Relative frequency	
missing	-9	1	0.0%	
inapplicable	-8	34959	96.96%	
refusal	-2	1	0.0%	
don't know	-1	1	0.0%	
Excellent	1	235	0.65%	
Very good	2	324	0.9%	
Good	3	281	0.78%	
Fair	4	150	0.42%	
or Poor?	5	103	0.29%	
	Total	36055	100.0%	

Wave 8	<i>h_sf1, h_indresp (2016-2017)</i>
Wave 7	<i>g_sf1, g_indresp (2015-2016)</i>
Wave 6	<i>f_sf1, f_indresp (2014-2015)</i>
Wave 5	<i>e_sf1, e_indresp (2013-2014)</i>
Wave 4	<i>d_sf1, d_indresp (2012-2013)</i>
Wave 3	<i>c_sf1, c_indresp (2011-2012)</i>
Wave 2	<i>b_sf1, b_indresp (2010-2011)</i>
Wave 1	<i>a_sf1, a_indresp (2009-2010)</i>
Wave B14	<i>bn_sf1, bn_indresp (2004)</i>
Wave B9	<i>bl_sf1, bl_indresp (1999)</i>

Click on the tab for Wave 1 to look at the frequency distribution of this variable for Wave 1

Then the Wave 1 tab will open.

Wave 2

b_sf1, b_indresp (2010-2011)

Wave 1

a_sf1, a_indresp (2009-2010)

Frequencies

Value label	Value	Absolute frequency	Relative frequency	
missing	-9	63	0.12%	
refusal	-2	18	0.04%	
don't know	-1	13	0.03%	
excellent	1	9583	18.79%	
very good	2	16285	31.94%	
good	3	14035	27.52%	
fair	4	7309	14.33%	
or Poor?	5	3688	7.23%	
	Total	50994	100.0%	

Wave B14

bn_sf1, bn_indresp (2004)

Wave B9

bl_sf1, bl_indresp (1999)

- This is a table of frequency of responses. In a data file, the information collected is recorded as numerical values (In this case, 1 2 3 4 5 -1 -2 -9) and then the data providers attach value labels so you can see what these values mean.
- What this table shows is that in Wave 1, 50994 individuals were interviewed of which 9583 individuals when asked to report on their general health said it was “Excellent”, 3688 said it was “Poor” and so on... This table also shows that 13 people said they didn’t know and 18 refused to answer. But there are 63 cases where the value of the variable is “missing”. What does that mean? It means that due to some coding or scripting error this information is not available for these 63 people.
- Different surveys have different ways of presenting the missing values. This study presents them as negative values. For all variables the same convention is used:
 - 1 (Don’t Know) means the person did not give a valid answer because they didn’t know
 - 2 (Refusal) means they refused to answer
 - 9 (Missing) means something went wrong in recording or coding the information

For some of the other variables, you will also find two other negative values -7 and -8. These don’t show up here as there are no such cases:

- 8 (Inapplicable): This value is assigned to a question when that question was not asked of a respondent because they did not qualify for it. For example, a question about the usual monthly pay, **W_PAYGU**, is only asked of those who say they are in paid employment. All those who said they are not in paid employment get a value of -8 for **W_PAYGU**.
- 7 (Proxy): This value is assigned to a question when that question was not asked of the respondent because they were being interviewed by proxy and this question was not included in the proxy questionnaire.

The variable **A_SF1** we discussed, was a simple question where individuals had to choose one answer. Their answers were then recorded as 1, 2, ...5. But what if they could choose more than one answer?

Let’s look for another question “**DISDIF**” – you can find it using any of the methods above. If you find it in the questionnaire you will see an item below the question text reads “CODE ALL THAT APPLY” Here is an extract from the questionnaire:

Disdif. *Type of impairment or disability*

Source
FRS (adapted)

Scripting Notes
Code 96 is exclusive

Text
Does this/Do these health problem(s) or disability(ies) mean that you have substantial difficulties with any of these areas of your life? Please read out the numbers from the card next to the ones which apply to you.

Interviewer Instruction
PROBE: WHICH OTHERS?
CODE ALL THAT APPLY

Showcard
TBC

Options

1	Mobility (moving around at home and walking)
2	Lifting, carrying or moving objects
3	Manual dexterity (using your hands to carry out everyday tasks)
4	Continence (bladder and bowel control)
5	Hearing (apart from using a standard hearing aid)
6	Sight (apart from wearing standard glasses)
7	Communication or speech problems
8	Memory or ability to concentrate, learn or understand
9	Recognising when you are in physical danger
10	Your physical co-ordination (e.g. balance)
11	Difficulties with own personal care (e.g. getting dressed, taking a bath or shower)
12	Other health problem or disability
96	None of these

“CODE ALL THAT APPLY” means that a person can choose more than one response options. Or they can choose 96 which means none of the 12 options were relevant for the person. So, how can this information be recorded in the data file? The response to this question is provided as a series of 13 variables each of which has a value 1 if it was mentioned and 0 if it was not. These are named **A_DISDIF1... A_DISDIF12 A_DISDIF96**.

You will also notice that there is an item called **Universe** just below the question:

Universe
If (Health = 1) // Has a long-standing illness or disability

What does that mean? It means that this question was NOT asked of everyone, but only those who said they had a long-standing illness or disability, that is, chose the response option 1 to the question **HEALTH**. If you go back a bit further up in the questionnaire you can find the question **HEALTH**.

Now look for one of these variables, say, **A_DISDIF1** using the variable search box. After you find it click on it, and then click on the Wave 1 tab, you will see,

Wave 1				a_disdif1, a_indresp (2009-2010)
Frequencies				
Value label	Value	Absolute frequency	Relative frequency	
inapplicable	-8	30850	60.5%	
proxy	-7	3262	6.4%	
refusal	-2	4	0.01%	
don't know	-1	5	0.01%	
not mentioned	0	10686	20.96%	
Mentioned	1	6187	12.13%	
Total		50994	100.0%	

The “Frequencies” table shows that 50994 individuals were interviewed but 30850 people were never asked this question (“inapplicable”) and for 3262 people we don’t have this information as this question was not included in the proxy questionnaire (“proxy”). It also shows that 6187 people did choose this option “mobility (moving around at home and walking)” while 10686 people did not choose this option.

So, basically this question was NOT ASKED of $30850+3262 = 34112$ people and WAS ASKED of $4+5+10686+6187 = 16882$ people.

Exercise 1: Does this mean that 16882 people answered Yes to the question **A_HEALTH**? Check the Frequency table for **A_HEALTH**! Actually 17745 respondents including proxy respondents answered Yes to this question. What happened to $17745 - 16882 = 863$ people? Why were they not asked this **DISDIF** question? These 863 people are proxy respondents: remember although **A_HEALTH** was asked of proxy respondents the follow-up **DISDIF** was not. You can check this in **Exercise 4**.

Exercise 2: Find variables to identify ethnic group, age, sex, region of residence, whether born in the UK. Write down the names of these variables.

When searching for variables you may sometimes come across different variables measuring the same thing. Then which one do you choose? This will mostly happen if you have a “derived variable” version of the same variable. For example, **A_DVAGE** and **A_AGE_DV** both measure age but **A_AGE_DV** is the one derived by the data team.

HELP AND SUPPORT

Understanding Society has a wealth of information on our [website](#):

It is a highly comprehensive online source of information regarding its variables, methodology, survey design and implementation. It is also an up to date source of training courses, webinars, videos, data releases and other relevant news regarding longitudinal research.

Email [User Support](#) if you have any comments or suggestions about this teaching resource.

ACKNOWLEDGEMENTS

This is a collaborative project by Understanding Society and the UK Data Service.

Acknowledgements go to Malcom Brynin (University of Essex) and Annette Pasotti (University of Essex).

Understanding Society

Institute for Social and Economic Research (ISER)
University of Essex
Wivenhoe Park
Colchester
CO4 3SQ
Tel: +44 (0) 1206 872957

www.understandingsociety.ac.uk

 @usociety

 Understanding Society –
UK Household Longitudinal Study



University of Essex

Understanding Society has been commissioned by the Economic and Social Research Council (ESRC). The Scientific Research Team is led by the Institute for Social and Economic Research (ISER) at the University of Essex.