Understanding Society: *Waves 2 and 3 Nurse Health Assessment, 2010-2012*

USER GUIDE

Version 4, *June 2025*

Institute for Social and Economic Research
University of Essex
Colchester
Essex

Understanding Society

Economic and Social Research Council

Contents

1. Overview

Understanding Society -- the UK Household Longitudinal Study (UKHLS) -- is a large longitudinal survey of approximately 40,000 households in the United Kingdom (England, Scotland, Wales and Northern Ireland). All members of responding households recruited at the first round of data collection and their descendants form the core sample and are visited at one year intervals to collect information on changes to their household and individual circumstances. The study in its current form started in 2009. A similar household survey, the British Household Panel Survey (BHPS), interviewed UK residents from 1991 to 2009, and surviving members of this panel joined Understanding Society in 2010. The overall purpose of the study is to provide high quality longitudinal data about 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 well-being of the UK population.

The study is primarily funded by the Economic and Social Research Council with the scientific leadership by the Institute for Social and Economic Research, University of Essex. To learn more about the study take a look at the User Guide (which provides the basic information about the study such as study overview, data collection, data structure, weights etc), questionnaires and the interactive variable search.

In 2010-2012 adult participants who had given consent received a health assessment visit from a registered nurse around 5 months after their Wave 2 interview (and in the case of BHPS sample members this was after their Wave 3 interview). The purpose of this health assessment was to provide objective measures of health risks and factors which may be clinical precursors to major chronic health conditions. A range of biomedical measures were collected from around 20,000 adults, which included blood pressure, weight, height, waist measurement, body fat, grip strength and lung function.

This document describes the biomedical measures collected and is accompanied by four other user guides which describe the biological datasets derived from the blood sample collected at the nurse visit (from 13,130 adults). The Biomarkers, Proteomics, Polygenic scores and Epigenetic ageing algorithms (clocks) datasets are outlined in the blood sample section below. In addition to this data, the main survey also collects a lot of information about health and wellbeing such as diagnosed health conditions, multiple self-reported physical and mental health and wellbeing measures.

2. Data collection

2.1 Overview of data collection

The fieldwork organisation was the National Centre for Social Research (NatCen). The nurse health assessments were conducted approximately five months following completion of the main survey interviews.

The measures included:

- Anthropometrics – height, weight, percent body fat, and waist circumference (See further details in the Anthropometrics section)

- Blood pressure (See further details in the blood pressure section)
- Grip strength (See further details in the grip strength section)
- Lung function (See further details in the lung function section)
- Blood samples which were analysed to produce different sets of measures. See the Blood samples Section for further details of these measures and links to the specialist user guides for these measures.

At these health assessments, the nurses also collected information about prescription medications that the respondent was taking (see further details in the prescription medication Section).

Most of these measures are available from the UK Data Service in the dataset "Understanding Society: Waves 2-3 Nurse Health Assessment, 2010-2012" (**SN 7251**) under the End User License (EUL) agreement. Some prescription medication, usage information and the polygenic scores are only available in the Special License (SL) version of this data, "Understanding Society: Waves 2-3 Nurse Health Assessment, 2010-2012: Special Licence Access" (**SN 7587**). This document shows the variables unique to the SL version.

Please visit the Data Citation section where we have provided information on how to cite the data correctly (which will vary based on which sets of variables you include in your analysis).

Information about design decisions and implementation of the nurse health assessments can be found in an *Understanding Society* working paper.[1]

2.2 Eligibility rules and fieldwork period

The Understanding Society in its current form started in 2009 with two samples – the General Population Sample (GPS), a nationally representation sample of approximately 26,000 households from across the UK and an Ethnic Minority Boost sample of approximately 4,000 households. Due to the large sample size the data collection was spread across a 24-month period. But respondents were still interviewed at approximately one year intervals. This was achieved by an overlapping fieldwork period. There is an *Understanding Society* working paper on the sample design of *Understanding Society*.[2] The surviving sample of the BHPS joined the survey in 2010 as part of the 2nd wave.

The nurse health assessments were not conducted in Northern Ireland or with members of the Ethnic Minority Boost sample. Only the GPS Year 1 sample and a sub-sample of the GPS Year 2 sample (specifically, 0.81 of the primary sampling units (PSU's) in England) and the BHPS samples were eligible. Further, eligibility criteria were completion of a full face-to-face interview in the corresponding wave in English, being adult (aged 16 or older) and living in England, Scotland or

[1] McFall SL, Booker CL, Burton J, Conolly A. Implementing the biosocial component of Understanding Society – nurse collection of biomeasures. Understanding Society working paper series 2012: Available from: https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2012-04.

[2] Lynn P. Sample design for Understanding Society. Understanding Society Working Paper Series 2009: Available from: https://research.understandingsociety.ac.uk/publications/working-paper/2009-01.

Wales. Respondents could request survey materials and a nurse visit in Welsh, as required by the Welsh Language Act. However, such requests were not made. Women who were pregnant at the time of the nurse assessment were classified as not eligible. Theses eligibility rules are also provided in a diagrammatic format in Figures 2 & 3 in the Response Outcomes section.

2.3 Data collection procedures, protocols and ethical approval

2.3.1 Nurse Training for data collection procedures and protocols

The nurse interviewers were registered on the Nursing and Midwifery Council (NMC), members of the Royal College of Nurses (RCN) or Royal College of Midwives (RCM), and competent in venepuncture. Nurse interviewers were required to have satisfactory immunity to hepatitis B, tuberculosis and rubella and to complete a satisfactory Enhanced Criminal Records Bureau check.

Nurses had a four-day generic training on the NatCen working principles, laptop use and CAPI training and protocols for commonly collected bio-measures and a one-day study-specific briefing. Survey nurses were periodically observed by supervisors and received feedback about their adherence to the study protocol.

The nurse protocols for specific measures and specific project instructions are available online. The document describes, in full, the protocols mentioned in the sections below for: anthropometric measurements, grip strength, lung function, blood pressure and blood sampling.

2.3.2 Ethical approval

Approval from the National Research Ethics Service was obtained for the collection of biosocial data by trained nurses in Waves 2 and 3 of the main survey. (*Understanding Society* - UK Household Longitudinal Study: A Biosocial Component, Oxfordshire A REC, Reference: 10/H0604/2).

2.3.3 Consent

Approximately five months after the interview, participants received an advance letter and small leaflet describing the nurse visit by post. These advance materials can be found online in the fieldwork documents.

The initial contact was often made by telephone unit staff who introduced this study component, answered non-clinical questions about the study, and explained that the nurse will be calling to make an appointment.

At the visit, participants received additional leaflets about the specific measures and the collection of blood samples. The survey nurses were briefed to give the participants time to read the leaflet and consent form and to consider their decision before signing the consent form or orally agreeing to a procedure. Participants were free to consent or not to any procedure. Direct measurements like grip strength were made following oral consent while written consent was obtained for invasive procedures (blood samples). The participants received a copy of the signed consent form which contained information about how to withdraw their consent to blood storage, if they decided to do so after the nurse had left the household. Respondents aged 16 and 17 years old were asked to consent to their own participation. However, nurses were advised to check with parents (when present), as a matter of courtesy, before taking a blood sample from this age group.

2.3.4 Interview and data collection process

Data collection began in May 2010 for eligible participants interviewed in January 2010. Blood samples were not collected until the second month because the storage facility was not initially ready to receive the specimens.

The interview was conducted with Computer Assisted Personal Interviewing (CAPI). The average nurse visit required approximately 60 minutes if all measurement and samples were taken.

Participants received a Measurement Record Card with their blood pressure, height, weight, waist circumference, percent body fat, and grip strength. There was no additional feedback of other measurements, for example, the results of blood tests. Participants received a £10 voucher after the nurse visit.

2.4 Questionnaires

The nurses followed the instructions in the measurement protocols and recorded responses in the CAPI program. Minor changes, which have been annotated in the questionnaire, were made in the CAPI instrument for the second year of Wave 2 and Wave 3. The CATI questionnaire was used by the telephone unit in explaining the study.

Figure 1 is a markup of a page in the CAPI questionnaire showing variable name, value labels, the universe for questions, question text, and instructions to the nurse. The “universe” shows who is eligible for the question and so provides important information for identifying subgroups who were asked the question or had the measurement procedure. The codes for different types of missing values are not shown in the questionnaire. Codes are:

-9	Missing by error.
-8	Not applicable to the person or because of routing
-2	Refused
-1	Don’t know.

These values have not been specified as missing in Stata or SPSS. However, these statistical packages have commands to assign values to missing for many variables simultaneously.

Figure 1. Markup of CAPI questionnaire

RespHts = **Variable name – Must add a wave prefix to the variable name**

NURSE: Measure standing height and code below. Include 'disguised' refusals such as 'It will take too long', 'I have to go out' etc. at code 2: Height refused**.**

= Instruction to the nurse

1. Height Measured

2. Height refused

3. Height attempted, not obtained= **Value Labels**

4. Height not attempted

[NODK, NORF]

IF RespHts = 1 THEN **= Universe for next items**

Height = **Variable name – Note no wave prefix**

NURSE: Record standing height (in centimeters).

Please record height with one decimal digit, using the full stop as decimal point.

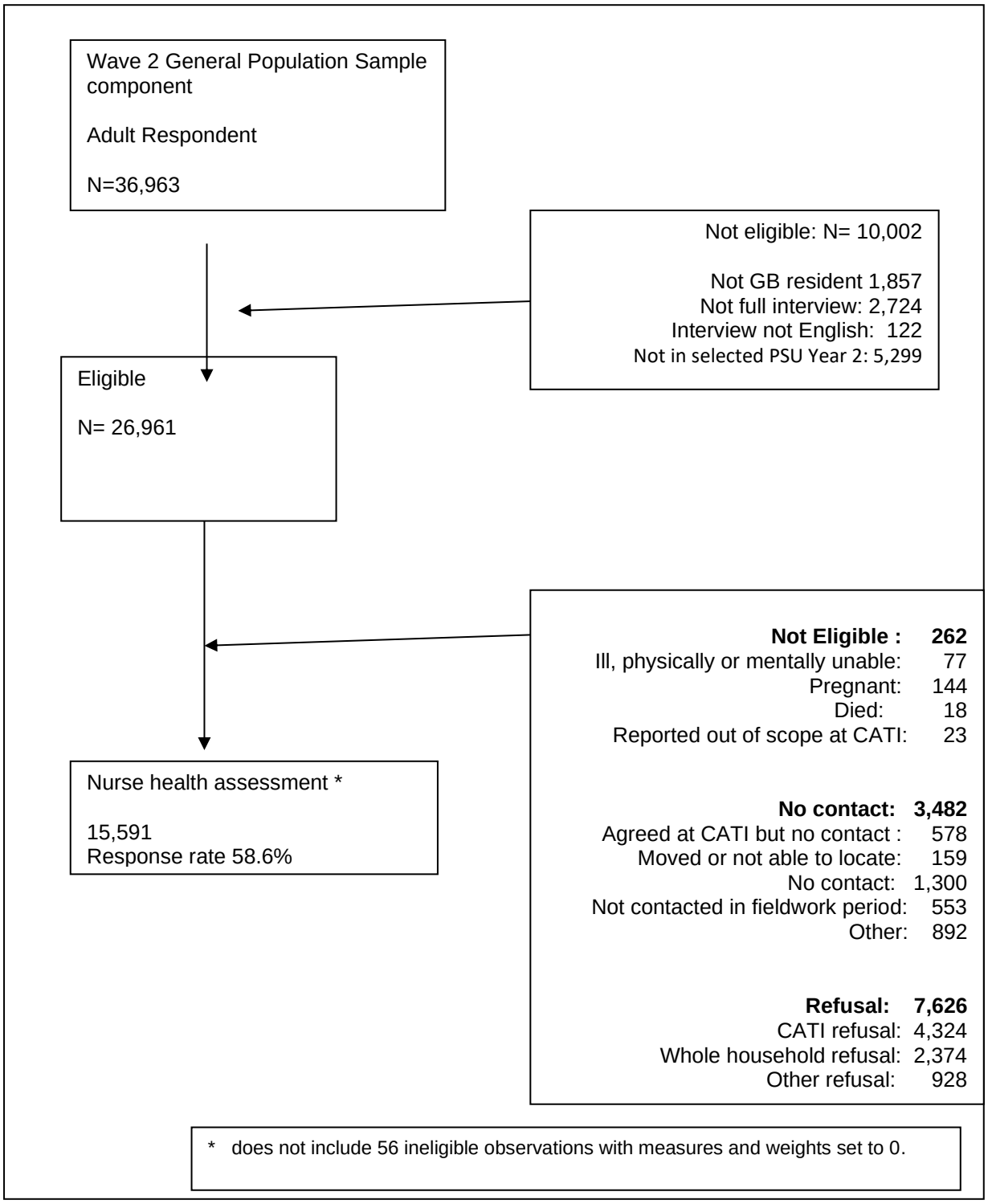
RANGE: 60.244 [NODK, NORF] = **Nurse records numeric**

Figure 2, a flow diagram of participation, begins with individuals potentially eligible for the nurse health assessment via their participation in the general population sample component of the Wave 2 adult interview. The top set of non-eligible individuals were not issued to the nurse assessment component of the study, either for resource reasons or because they did not complete a full interview. The box on the lower right shows the small number of additional persons identified and classified as not eligible during the invitation to participate. The rate of refusal among eligible participants was 28.6%. The overall response rate to the nurse health assessment, defined as providing one or more measurement, was 58.6%.

Figure 3 summarises participation of adults interviewed in Wave 3 of *Understanding Society*. The top right box has ineligible individuals (persons not in England, Wales or Scotland, those who completed a proxy or telephone interview and those interviewed in a language other than English). Additional persons were identified as not eligible closer to the time of the nurse health assessment. The rate of refusal among eligible participants was 19.4%. The proportion participating, that is completing one or more health measurements was 57%. For the 20,700 adults who participated in the nurse visit 19,121 were eligible. From the 14,433 who consented, 13,130 had blood taken at the nurse visit, a response rate of 91%. This equates to a response rate of 63.4% for those who participated in the nurse visit (20,700) or 68.7% for those who were eligible to participate (19,121). Full details on

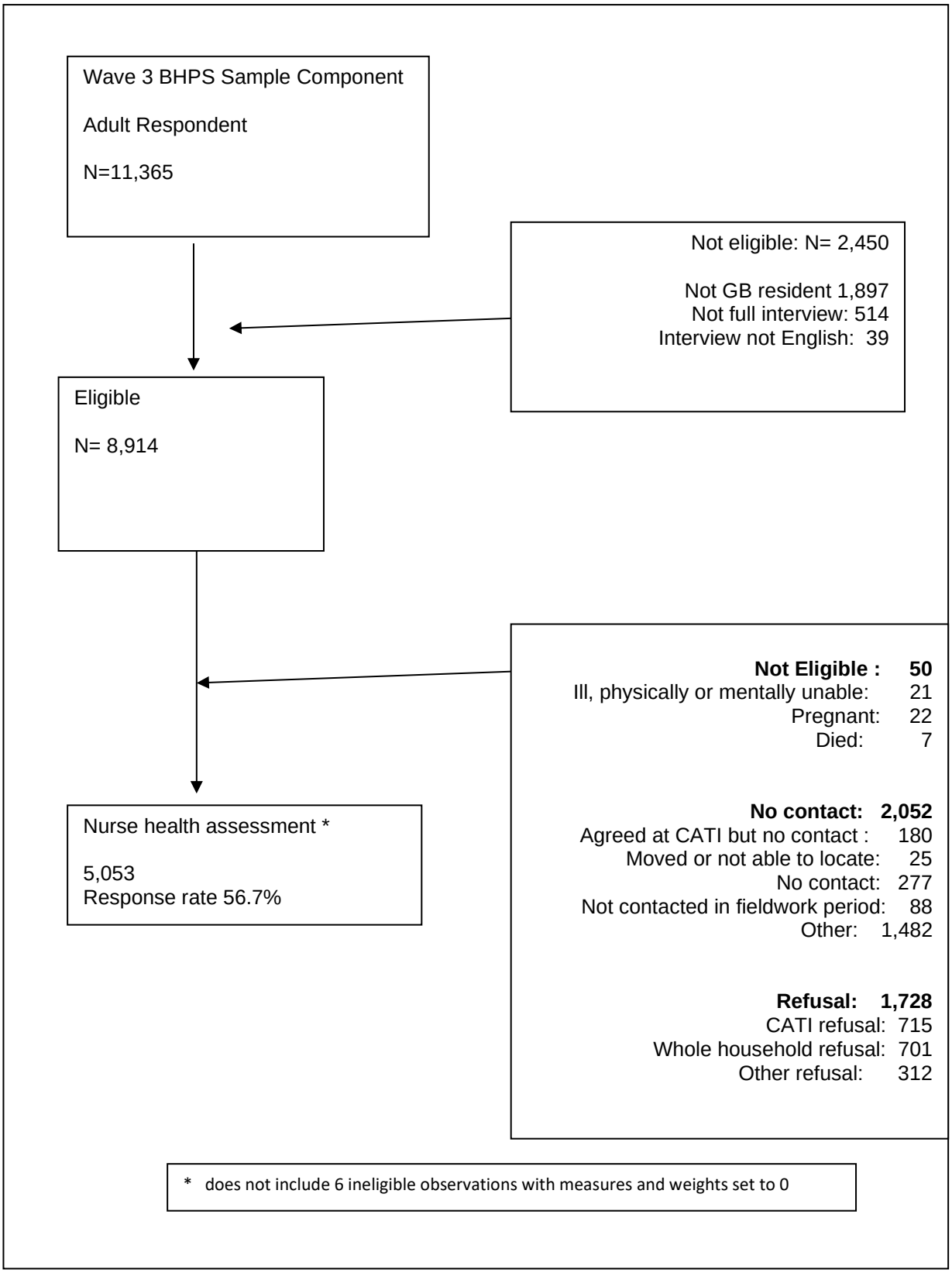
eligibility, missing cases and participation rates for the blood sample collection are available in Table 1 in the Biomarker User Guide and Glossary[3].

Figure 2. Flow diagram of participation in the nurse health assessment General Population Sample Component



[3] Institute for Social and Economic Research. (2024). Understanding Society: Biomarker User Guide and Glossary, Version 3, July, 2024, Colchester: University of Essex. https://www.understandingsociety.ac.uk/documentation/health-assessment/user-guide

Figure 3. Flow diagram of participation in the nurse health assessment BHPS Sample Component



The following sections, 3-6, describe the specific measures used during the nurse visit.

3. Anthropometric measures

A full description of these measures can be found in the Nurse protocols document.

3.1 Background: Anthropometric measures

Anthropometric measures assess body size and composition. Understanding Society has measures of height, weight, body fat, and waist circumference. In addition, body mass index (BMI) has been derived using height and weight. Anthropometric measures are relevant to obesity and overweight. However, they are also important measures of general health, are risk factors for multiple chronic conditions and are relevant to nutrition status. The anthropometric measures are genetically and environmentally determined.

Obesity or overweight are of interest because they are risk factors for disease and mortality. A number of studies have established that overweight or obesity is associated with cardiovascular risk, cardiovascular related mortality, cancer, disability during older age, and decreased life expectancy. Furthermore, obesity is associated with serious chronic conditions such as Type 2 diabetes, hypertension, and hyperlipidaemia (i.e., high levels of fat in the blood that can lead to narrowing and blockages of blood vessels), which are risk factors for cardiovascular disease.

Height has been negatively correlated with all-cause mortality and for mortality from several chronic conditions, though positively correlated with cancer mortality. Height is also interesting because it reflects the influence of energy intake and composition of diet in early life.

Percent body fat is of interest for multiple reasons. The first is as an alternative measure to body mass index. It is well known that BMI does not distinguish between lean and fat body mass. Thus the percent body fat can help distinguish between those with high BMI from muscle and those with high adiposity. In addition there is interest in variation in body fat, which may be related to energy consumption, hormones, and heredity.

Bioelectrical impedance analysis (BIA) is a relatively simple and non-invasive measure of body fat. It is based on the fact that fat is less conductive than bone and muscle. Measurements are influenced by consumption of food or drink, exercise, conditions that alter fluid and electrolyte balance, the testing environment and ethnicity. Additional research is required to develop predictive equations for different ethnic groups.

Researchers have also been interested in different patterns of fat deposit, in particular on the effects of abdominal or visceral fat, assessed by waist circumference. Waist circumference has been found to contribute beyond BMI or weight to the risk of cardiovascular disease.

For further reading see Section 14.1

3.2 Height

3.2.1 Sample: height

Respondents were excluded from the height measurement if they were pregnant, too stooped to be reliably measured, too unsteady on their feet, chairbound, or if it was painful for the respondent to stand.

3.2.2 Equipment and procedures: height

Standing height was measured with a portable stadiometer. The stadiometer has a footplate, measuring rule, and an adjustable head plate. The respondent was asked to remove their shoes. The respondent was asked to stand with feet flat on the centre of the base plate, feet together and heels against the rod. The respondent was asked to hold their back as straight as possible, preferably against the rod but not leaning on it. The interviewer checked the correct alignment of the respondent's head in the Frankfort plane using a card. This increases the accuracy of the measurement. One measurement, to the nearest millimetre, was taken. Height was recorded on the Measurement Record Card.

3.2.3 Key height variables

Multiple variables report on potential problems with the measurement of height or whether directly measured or estimated. Nurses judge whether these problems reduce reliability. Some key variables are:

w_height	Standing height (cm) including unreliable measurements
w_htok	Whether height measurement is valid
w_ htval	Standing height for those with valid measurements

3.3 Weight and percent body fat

3.3.1 Sample: weight

Participants who were pregnant, unable to stand, or unsteady on their feet were not weighed. In addition, participants with a pacemaker did not have the body fat assessment since the electrical current could affect the pacemaker.

3.3.2 Equipment and procedures: weight and body fat

Weight and percent body fat were measured using a digital floor scale, the Tanita BF 522 scales. Body fat is estimated by sending electrical current through the body through the feet. The estimation is based on the respondent's age, gender and height. We assumed the respondent's body type was "Standard". This approach requires less training and standardisation for measurement.

The scale was placed on a hard level surface. The respondent was asked to remove shoes and socks and stand with both feet flat on the surface of the foot pads. A single measurement was recorded to the nearest .1kg. The scale can also be used in "weight only" mode.

Participants who weighed more than 130kg were asked for their estimated weights because the scales are inaccurate above this level. The interviewer was asked to report circumstances that could make the measurement unreliable, e.g., wearing heavy clothing.

3.3.3 Key weight and body fat variables

w_weight	Weight (kg) including unreliable measurements
w_respwts **w_wtok**	Whether weight measurements attempted or obtained Whether weight measurement is valid
w_ wtval **c_bfpcok**	valid weight (kg) incl. estimated >130kg Whether body fat measurements is valid
w_bfpcval	Percent body fat for valid measurements according to nurse

3.4 Waist circumference

3.4.1 Sample: waist circumference

Participants were excluded from waist measurements if they reported that they had a colostomy or ileostomy, or were pregnant or unable to stand. The nurses reported problems with measurements which may make them unreliable, e.g, not holding still or excessive clothing.

3.4.2 Equipment and procedures

Waist circumference was measured using a tape with an insertion buckle at one end. The waist was defined as the midpoint between the lower rib and the upper margin of the iliac crest. The measurement was taken twice and was recorded to the nearest millimetre. If the two waist measurements differed by more than 3 cm, a third measurement taken.

3.4.3 Key variables waist circumference

w_waist1 – w_waist3 **w_wstokb**	Individual waist measurements in cm Whether waist measurement is valid
w_wstval	valid mean (2 closest measures if 3rd taken) waist circumference (cm)

3.5 Related variables: BMI

Body mass index is the weight in kg divided by height in metres squared. The datasets have two different BMI variables, **w_bmi** (BMI incl. unreliable measurements) and **w_bmival** (valid BMI - incl. estimated>130kg). The first is derived from any non-missing pairs of height and weight. The second is derived from any non-missing pairs of height and weight where no measurement problems were recorded and with weight substituted with the estimated weight for participants weighing an estimated 130kg or more.

3.5.1 Key variables Body Mass Index (BMI)

w_bmiok **w_bmi**	Whether BMI measure is valid Whether BMI measure is valid BMI including unreliable measurements according to nurse
w_bmival **w_bmivg5**	Valid BMI (according to nurse) includes estimated weights more than 130 kg Valid BMI in categories

4. Blood pressure and pulse or resting heart rate

A full description of these measures can be found in the Nurse protocols document.

4.1 Background

Hypertension is an important risk factor for cardiovascular and all-cause mortality. It is associated with multiple chronic conditions including cardiovascular disease, stroke, and diabetes. Heart rate has not been as widely studied, but it has been identified as an independent risk factor for cardiovascular and all-cause mortality. A great deal of research has been done on risk factors for elevated blood pressure. These include age, body weight, alcohol consumption, dietary factors like sodium, and blood pressure levels when younger. Additional behavioural risk factors include physical activity and smoking.

For further reading see Section 14.2

4.2 Blood pressure and pulse or resting heart rate

Understanding Society collects systolic and diastolic blood pressure and pulse. Systolic blood pressure is the peak pressure with diastolic blood pressure being the lowest pressure, linked to the resting phase of the cardiac cycle. Pulse or resting heart rate is measured as the number of beats per minute.

Hypertension or high blood pressure is classified based on elevated values on multiple occasions. The survey situation cannot provide a clinical definition of hypertension as outlined in NICE guidelines but we present classification based on assessments and reported use of medications for hypertension[4].

- Variables have been derived from the blood pressure measurements to categorise participants according to severity and treatment status based on medications (**w_bpmedc**):
    - w_hyper1 hypertensive categories: stage 1 140/90 all taking blood pressure drugs
    - w_hyper2 hypertensive categories: stage 2 160/100 all taking blood pressure drugs

[4] Health Survey for England 2017 Adult Health (digital.nhs.uk)

Additional measurements can be estimated using the blood pressure assessments. These measurements include:

- pulse pressure (**w_ompulval**),
- mean arterial pressure (provided with the data as w_map1/2/3, and **w_ommapval**), or
- variations like isolated high systolic or diastolic blood pressure (**w_omsysval** and **w_omdiaval**).

Systolic blood pressure increases with age. Diastolic blood pressure goes up with age to about age 55 and then plateaus or declines. Blood pressure is similar in men and women.

4.2.1 Sample: blood pressure and pulse or resting heart rate

All participants are eligible except pregnant women.

4.2.2 Equipment: blood pressure and pulse or resting heart rate

The portable blood pressure monitor, the Omron HEM 907, is equipped with three sizes of cuffs: small (17-22 cm), standard (22-32 cm), or large (32-42 cm). This oscillometric device has been validated against mercury sphygmomanometre within the U.S. Health and Nutrition Examination Survey and met the criteria of the Association for the Advancement of Medical Instrumentation for adult age groups.[5] The blood pressure monitor also records the pulse level or number of irregular pulse waves (arrhythmias). The pulse wave having a difference of more than +25% from the average interval of pulse waves is determined as irregular. Irregular pulse waves are displayed up to three times (see figure 6.3.1 in the Nurse protocols document).

4.2.3 Procedures: blood pressure and pulse or resting heart rate

Respondents were asked if they had eaten, smoked, drunk alcohol, or been physically active within the past 30 minutes. The assessment was done with the right arm, if possible, with the respondent sitting in a comfortable chair and their arm supported to bring the elbow to the level of the heart. Feet were placed flat on the floor. Before the first reading, the respondents were asked to sit quietly for five minutes. Three readings were taken. The measurements were recorded on the Measurement Record Card.

In addition to the Measurement Record card, the nurse gave the following feedback about blood pressure according to the following values for the lowest of the last two readings.

Classification	Systolic		Diastolic
Normal	<140	and	<90
Mildly raised	140-159	Or	90-99
Raised	160-179	Or	100-114
Considerably raised	180 or more	Or	115 or more

[5]Ostchega Y, Nwankwo T, Sorlie PD, Wolz M, Zipf G. Assessing the validity of the Omron HEM-907XL oscillometric blood pressure measurement device in a National Survey environment. J Clin Hypertens (Greenwich). 2010;12(1):22-8.

These messages, available on the screen, were read:

> **Normal:** "Your blood pressure is normal."
> **Mildly raised:** "Your blood pressure is a bit high today. Blood pressure can vary from day to day and throughout the day so that one high reading does not necessarily mean that you suffer from high blood pressure. 'You are advised to visit your GP within 2 months to have a further blood pressure reading to see whether this is a one-off finding or not."
> **Raised**: "Your blood pressure is a bit high today. Blood pressure can vary from day to day and throughout the day so that one high reading does not necessarily mean that you suffer from high blood pressure. You are advised to visit your GP within 2 weeks to have a further blood pressure reading to see whether this is a one-off finding or not."
> **Considerably raised:** "Your blood pressure is high today. Blood pressure can vary from day to day and throughout the day so that one high reading does not necessarily mean that you suffer from high blood pressure. You are strongly advised to visit your GP within 5 days to have a further blood pressure reading to see whether this is a one-off finding or not."

In addition, fast and slow pulse rates were routinely checked by the Survey Doctor. Nurses were asked to contact the Survey Doctor about considerably raised blood pressure readings. The survey doctor telephoned respondents to advise them to visit their GP within 5 days to have a further blood pressure reading. A letter was also posted with this advice.

5. Grip strength

A full description of these measures can be found in the Nurse protocols document.

5.1 Background: grip strength

Hand grip strength is used in clinical settings and in survey research. While it measures muscular strength in one part of the body, and is used to monitor conditions of the hand such as arthritis, grip strength is also correlated with overall body strength. In surveys, particularly those related to ageing, it is an objective measure of physical capability or the ability to undertake physical tasks of daily living.

Handgrip strength is higher in men than women and is associated with measures of body size, e.g., height and weight. The tracking of strength over time suggests the importance of a reserve of muscle strength in preventing disability.

The picture is not entirely consistent in the relation of grip strength and socioeconomic status. Some studies have found an association with social class, education and wealth.

The accuracy of measurement is influenced by the interviewer's encouraging of the participant to make full effort. In this study, the nurses' demonstrations draw attention to their efforts. The nurses also verbally encourage the participant to exert full force.

For further readings see Section 14.3

5.2 Measuring grip strength

5.2.1 Sample: grip strength

Respondents were excluded if they had swelling or inflammation, severe pain or a recent hand injury, or had surgery on their hands in the last 6 months.

5.2.2 Procedures: grip strength

Grip strength is measured with a gripometer or dynamometer, which consists of a gripping handle with a strain-gauge and an analogue reading scale. The specific model is Smedley's dynamometer.

The nurse prepares the respondent by explaining the purpose of the measurement and that it is very important that to try their hardest for the most accurate reading of their grip strength. Respondents are asked to remove any large rings.

The nurse demonstrates the procedure. If the respondent does not understand, the nurse repeats the demonstration. The dynamometer is adjusted to fit the respondent's hand. The respondent has one practice with their dominant hand. The procedure is done standing with upper arm against the trunk and forearm at a right angle to the upper arm. Following the practice, the respondent does the procedure three times on each hand, squeezing as hard as they can for two seconds. Hands are alternated. The value on the scale was recorded to the nearest whole number. In addition, the results were recorded on the Measurement Record Card.

If the respondent cannot stand, they can do the procedure sitting, keeping their upper arm against the trunk with forearm at right angles to their upper arm. If the device is too heavy to hold, they can use their other hand to support it.

5.2.3 Key variables – grip strength

w_ mmgsdom	dominant hand to be used for grip strength measurements
w_mmgsdval	Valid maximum reading in kg for dominant hand
w_mmgsnval	Valid maximum reading in kg for nondominant hand
w_mmgsd1 – w_mmgsd3	Individual readings for dominant hand in kg
w_mmgsn1 – w_mmgsn3	Individual readings for nondominant hand in kg

A full description of these measures can be found in the Nurse protocols document.

6. Lung Function

6.1 Background: lung function

Tests of lung function are used in conjunction with clinical information to identify and monitor persons with lung obstruction. These are the principal lung function measures obtained in *Understanding Society*:

Forced Vital Capacity	FVC	The total amount of air that can forcibly be blown out after a full inspiration, measured in litres.
Forced Expiratory Volume in 1 Second	FEV1	The amount of air that can be blown out in one second, measured in litres.
FEV1%	FEV1/FVC	The ratio of FEV1 to FVC.
Peak Expiratory Flow	PEF	The speed of air moving out of your lungs at the beginning of expiration, measured in litres per second.

There are different standards for diagnosis and classification of COPD and for how to present the data. Some sources are provided under Further Reading.

Chronic respiratory conditions are the fourth leading cause of mortality in the UK. Chronic obstructive pulmonary disease (COPD) and asthma are the most common chronic lung conditions. About 80% of older persons with COPD have co-morbidities.

Lung function declines with age from a peak in young adulthood, even in persons without lung disease. Poor lung function is associated with mortality, physical function, and cognitive function. Persons in lower social classes have poorer lung function and greater declines in lung function over time even when controlling for smoking. Active and passive smoking are correlated with poorer lung function.

Multiple factors extraneous to lung function affect measurement. Body size (height) influences the size of the lungs. Some of the derived lung function measures incorporate height information. The motivation of the participant is highly influential. The nurses are trained to encourage participants to make a good effort. The new technology has features to support participant motivation. There are also some variables which could be used to identify participants with poor effort. Consideration should also be given to differences in technology and equipment as described below.

For further reading see Section 14.4

6.2 Measuring lung function

The following sections describe the specific measures used during the nurse visit.

6.2.1 Sample for lung function measurements

All participants were eligible except pregnant women, persons who had abdominal or chest surgery (past 3 weeks), those in hospital with a heart complaint (6 weeks), persons who had recent eye surgery (4 weeks), or with a tracheostomy.

6.2.2 Spirometry equipment

The lung function measures were conducted with two different devices: in England and Wales, the electronic NDD Easy On-PC spirometer (explained in the lung function booklet), and in Scotland the Vitalograph Escort (explained in the nurse protocols). The procedures for these two devices are described separately.

There is no evidence about the level of equivalence for the two machines and procedures. We recommend that analysts not combine the measures.

6.2.3 Lung function procedures in England and Wales

The NDD Easy On-PC has some distinctive features among portable spirometers. It uses ultrasonic technology such that the energy in the breath is changed to ultrasound and measured. The device links directly to the nurse laptops and records a wide range of parameters. In addition it provides quality control. With each blow, the machine gives feedback to the nurse, identifying specific reasons that a blow was classified as inadequate. The machine also provides an overall quality assessment or grade for a given participant.

Nurses conducted a calibration check, using a three litre calibration syringe, each day of work. However, the manufacturer says that calibration is not needed.

With a sample mouthpiece and nose clips, the nurse demonstrates how to use the mouthpiece—not holding it by the teeth but firmly wrapping their lips around the mouthpiece so that air does not escape. The nurse demonstrates the manoeuvre, beginning with a full inhalation, a big start on exhalation and maintaining the exhalation as long as possible. The blow should ideally last at least 6 seconds, and not be interrupted by coughing, glottis closure laughing or leakage of air. The procedure is done seated in an upright position. The nurse also points out the computer animation to the respondent. This is a little boy blowing up a balloon that pops as the participant completes their exhalation.

The nurse gave participants a new nose clip and mouthpiece to help them become familiar with how they feel. The procedure was done with the participant seated, with feet on the floor.

The nurse input age, sex, and ethnicity for the prediction equations. In the manoeuvre, the participant breathes in as deeply as possible and then immediately blasts out the air as hard and as fast as they can. They keep blowing until there was no more air in their lungs. The nurse encourages the participant to “keep going, keep going, keep going...” to get the maximum and fastest expiration possible.

When the blow has been completed, the programme rates it as either acceptable or unacceptable. For not acceptable attempts, the programme gives feedback about the reason, which the nurse uses to instruct the participant in how to make their next blow acceptable.

Trials are ended when the participant produces a quality grade of A was achieved, if the participant was unwilling or too tired to continue, if the nurse had concerns about the participant's safety. Grades of B were also acceptable but if the nurse thought the participant was capable of achieving an A grade, participants were asked to make another blow.

Here is some information about the quality grades.

Table 1 Quality criteria for spirometry sessions

Grade	Number of acceptable forced expiratory manoeuvres	Additional criteria
Grade A	At least three	Two highest FVC and FEV1 within 100ml
Grade B	At least three	Two highest FVC and FEV1 within 150ml
Grade C	At least two	Two highest FVC and FEV1 within 200ml
Grade D	Only one or	Or best two FEV1 or FVC were not within 200ml
Grade F	None	N/A

6.2.4 Lung function procedures used in Scotland

The Vitalograph Escort is a portable spirometer that uses Fleisch air flow technology. It does not have the convenience of automatic storing and recording of results in the lap top. The quality judgements are made by the nurse and are more subjective. The Vitalograph spirometer is calibrated with a one litre syringe each day that the nurse conducts assessments.

The nurse demonstrates the manoeuvre which includes deep inspiration, a sharp blast on exhale and continued exhalation as long as possible. The procedure is done standing unless chairbound. The nurse encourages saying "Keep going, keep going, keep going."

The nurse requests that the respondent repeat the manoeuvre five times. In each trial the nurse evaluates the blow as technically satisfactory or not satisfactory. The following may result in a technically unsatisfactory blow. They include a bad start, laughing or coughing, leakage of air around the mouthpiece, or blocking the flow of air by teeth, tongue or hands.

The data file **w_lfscot_ns** has the lung function measurements recorded in Scotland. Values are recorded for each trial for FVC, FEV, PEF and fev1/fvc ratio. There are also indicators of whether the trial was satisfactory and reasons for not having a reading.

6.2.5 Key variables – lung function

There are multiple choices for lung function measures. First, there are summary measures of the major lung function measures, the highest technically satisfactory measures for FVC, FEV1, PEF, and the ratio of FEV1/FVC. They are produced for both England/Wales and Scotland. However, the

England/Wales values are selected by the spirometry computer program. The ratio of FEV1/FVC may be based on two different manoeuvres. The values for Scotland are based on the maximum with a satisfactory technique over the multiple readings. The meaning of satisfactory is therefore not the same in England/Wales vs. Scotland.

w_htfvc	Highest technically satisfactory value for FVC
w_htfev	Highest technically satisfactory value for FEV
w_htpef	Highest technically satisfactory value for peak flow
w_htfevfvc	fev1/fvc ratio (based on highest technically satisfactory values)

We also have a number of 'predicted' lung function variables in the dataset. These were calculated as a percentage of the expected values for each individual based on their characteristics, according to the Stanojevic 2009 all-ages equations [6]

w_fvcpred	Percentage predicted of FVC
w_fev1pred	Percentage predicted of FEV1
w_fev1fvcp	Percentage predicted of FEV1/FVC ratio
w_fev1pred_sc	Percentage predicted of FEV1 (Scotland)
w_fvcpred_sc	Percentage predicted of FVC1 (Scotland)
w_fev1fvcp_sc	Percentage predicted of FEV1/FVC ratio (Scotland)

7. Blood sample

Blood samples were taken from 13,130 adults who had given written consent at the nurse visit. The consent form booklet can be found online https://www.understandingsociety.ac.uk/documentation/health-assessment/fieldwork-documents. The participants received a copy of the signed consent form which contained information about how to withdraw their consent to blood storage, if they decided to do so after the nurse had left the household. Respondents aged 16 and 17 years old were asked to consent to their own participation. However, nurses were advised to check with parents (when present), as a matter of courtesy, before taking a blood sample from this age group.

Several biological datasets have been derived from the blood sample and include: Biomarkers, Proteomics, Polygenic scores and Epigenetic ageing algorithms (clocks) datasets. These are available

[6] Stanojevic S, Wade A, Cole TJ et al. Asthma UK Spirometry Collaborative Group. Spirometry Centile Charts for Young Caucasian Children: The Asthma UK Collaborative Initiative. Am J Resp Critical Care Med. 2009;180:547-52

in this nurse assessment dataset (see Table 2), but their descriptions are available in separate user guides.

7.1 Biomarkers

- *Cholesterol and triglycerides (measures of fat in the blood)*
- *Apolipoproteins (markers of lipid metabolism and cardiovascular disease risk)*
- *Glycated haemoglobin (an indicator of diabetes)*
- *High sensitivity-C-reactive protein, fibrinogen and CMV seropositivity (measures of inflammation and the immune system– the body's way of responding to damage and infection)*
- *Haemoglobin and ferritin (measures of anaemia)*
- *Complement Factor H (Regulator of the complement system; associated with immune response and inflammatory disease)*
- *Liver function tests, albumin, creatinine (Liver and kidney function)*
- *Testosterone, SHBG, DHEAs, Insulin-like growth factor-1 (Hormones that balance and build up the body)*

A detailed User Guide and Glossary for the biomarker data in *Understanding Society* is available .

7.2 Proteomics

184 proteins were measured in 6,180 adults. These proteins focus on cardiometabolic health (e.g. cardiovascular disease, diabetes) and neurological processes (neurodegenerative disease, cognitive function).

A detailed User Guide for the proteomics data in *Understanding Society* is available .

7.3 Genetics and Epigenetics

A genome-wide scan (Illumina human core exome array) has been conducted on DNA samples from approximately 10,000 people. DNA methylation profiling has been conducted on DNA samples from approximately 3,650 individuals: 1,425 from the British Household Panel Survey component and another 2,230 from the General Population Sample. Using the Illumina methylation EPIC beadchip we have integrated over 850,000 DNA methylation sites across the genome. Genomic and epigenomic data are available to researchers by application only. Applications for genetic or epigenetic data only can be made to the European genome-phenome archive (EGA: www.ebi.ac.uk): genome wide scan dataset, accession number egas00001001232; DNA methylation egas0001002836. Where genomic or epigenomic data are linked to the main or other dataset applications should be made via the Understanding Society omics application portal https://www.understandingsociety.ac.uk/documentation/omics/access-the-omics-data/

Polygenic scores and Epigenetic ageing algorithms (clocks) have been derived from the genomic and epigenomic datasets, respectively. Polygenic scores are available in the Special Licence version only SN7587. The user guides for the proteomics and the epigenetic ageing algorithms (clocks) are available under SN7251.

8. Prescription medications

Generally, the nurse coded prescription medications while the respondent was resting before the blood pressure readings. The nurse asked, "Are you taking any other medicines, pills, ointments or injections prescribed for you by a doctor (**w_medcnjd**)?" The nurse was instructed to ask to see the containers for the medicines.

Drugs are coded using their British National Formulary (BNF) classification codes - down to the third level of classification. The nurse records the brand name or generic name and codes it. Coded values are recorded in a six-digit format with leading zeroes. For example, simvastatin, is coded Cardiovascular/Lipid lowering drug/statin or 021201. There are also follow-up questions about whether the drug is one commonly prescribed for cardiovascular disease or hypertension.

A short booklet with the most commonly used drugs in alphabetical order is also available to the nurse. Detailed data on medications are available under a special licence agreement **w_ytake** and **w_medbi** and related variable **w_medbia**. A set of derived variables grouping prescribed medications according to BHF chapters are available under the end user licence agreement (**w_medtyp1-w_medtyp13**).

9. Data structure

9.1 File and variable naming conventions

The data files related to the Wave 2 Nurse Health Assessment begin with "**b_**" for Wave 2 and a suffix of **_ns** for "nurse". The files for Wave 3 begin with "C". We express wave in general as **w_**. Similarly, variable names begin with either "**b_**" or "**c_**". There is also a cross-wave file **xindresp_ns** which has data from both waves. Only a few variables in this file have the wave prefix, e.g., the wave-specific household identifier.

Each individual has a public identifier, not related to survey operations, which is consistent across waves. Because the nurse data was collected following the main survey interview for the wave, a few households had split by that time. They can be identified with the variable **b_splitnum**. Their household id is the same as in the main survey interview. As a consequence, there are duplicate household identifiers. See the section on Data management tasks for a programming example related to this issue.

Table 2 summarises the data files in this release.

Table 2. Data files released

File name	Description
w_callrec_ns	Call records (sampling process)
w_hhsamp_ns	Sample and household level information for each eligible household issued to the nurses (sampling process)

w_indsamp_ns	Individual eligibility and enumeration information for individuals issued to the nurse (sampling process)
w_indresp_ns	Individual responses and measurements
w_issue_ns	Information about each household issued (sampling process)
w_lfscot_ns	Lung function readings from Scotland[7]
w_ptimings_ns	Individual level timings for nurse health assessment
xindresp_ns	Individual responses and measurements for Waves 2 & 3 combined
xivdata_ns	Characteristics of nurse interviewers
xlabblood_ns	Cross wave blood-based biomarker data
xepigen_clocks_ns	Cross wave epigenetic ageing algorithms (clocks)
xpolygen_ns	Cross wave polygenic score data
xproteo_ns	Cross wave Proteomics data
xproteo_lod_ns	Limit of detection file to accompany the proteomic file

9.2 Missing values

These values have not been specified as system missings in statistical packages that you may be using such as Stata, SAS, R or SPSS. However, these statistical packages have commands to assign values to missing for many variables simultaneously.

9.3 Survey variables copied over to indresp_ns files

We have provided selected health and social variables collected in waves 2 or 3 for analytic convenience. The online documentation system has more information for each variable. (See https://www.understandingsociety.ac.uk/documentation/health-assessment/dataset-documentation/). Table 3 summarises the variables copied over from the interview, classified as

[7] see section on lung function procedures in Scotland

socio-demographic or health status and health behaviours. Note that some of the health variables were measured only in Wave 2.

Table 3. Variables from main survey interview

Variable name	Description
Socio-demographic variables	
w_hhsize	number of people in household
w_hhtype_dv	Composition of household, LFS-version
w_jbstat	current economic activity
w_mlstat	Legal marital status
w_marstat_dv	De facto marital status
w_hiqual_dv	highest qualification
w_jbnssec8_dv	Current job: Eight Class NS-SEC
w_jbnssec5_dv	Current job: Five Class NS-SEC
w_jbnssec3_dv	Current job: Three Class NS-SEC
w_jlnssec8_dv	Last job: Eight Class NS-SEC
w_jlnssec5_dv	Last job: Five Class NS-SEC
w_jlnssec3_dv	Last job: Three Class NS-SEC
w_urban_dv	Urban or rural area
w_ieqmoecd_dv	Modified OECD equivalence scale
w_fihhmnnet1_dv	Total household net income – no deductions
w_fihhmngrs1_dv	Gross household income: month before interview, hb adj
Health conditions and health behaviours	
w_sf1	General or self-rated health
b_wkfruit	days each week eat fruit
b_fruvege	number of servings of fruit/vegetables eaten
b_smever	ever smoked cigarettes
b_smnow	smoke cigarettes now
b_ncigs	usual number of cigarettes smoked per day
b_smcigs	ever smoked cigarettes regularly
b_smncigs	number of cigarettes smoked in the past
w_health	long standing illness
w_scghqa – w_scghql	12 items of the General Health Questionnaire (GHQ)
w_scghq1_dv	Subjective wellbeing (GHQ): Likert or continuous measure

w_scghq2_dv	Subjective wellbeing (GHQ): Caseness or categorical measure

Understanding Society offers many other domains. A general view of the content of *Understanding Society* can be seen in the long term content plan: https://www.understandingsociety.ac.uk/documentation/mainstage/long-term-content-plan.

9.4 Data issues

We request that researchers using the data notify us about errors, inconsistencies, and other problems with the data identified during their use of the data. We make use of this information in improving the data. Please raise any issues relating to data or data analysis with our user support service https://iserredex.essex.ac.uk/support/projects/support

10. Data management tasks

This section describes multiple data management tasks using Stata which are important for the analysis of the nurse health assessment data.

10.1 Merging individual level nurse health assessment data with data from the main survey

The following Stata code would create a dataset with the BMI (derived from measured height and weight) from the Nurse Health Assessment component and combine this with the respondents' Wave 1 BMI (derived from self-reported height and weight). The missing value codes can be recoded using the **mvdecode** command.

```
use pidp b_hidp b_bmival using b_indresp_ns, clear
merge 1:1 pidp using a_indresp, keep(3) keepusing(a_bmi_dv) nogen
mvdecode b_bmival a_bmi_dv, mv(-9/-1)
```

The **correlate** command can be used to produce the correlation coefficient between the two variables to see how well they do/don't match.

```
correlate b_bmival a_bmi_dv
```

10.2 Merging household level nurse health assessment data with data from the main survey

To merge in household level data from Wave 2, it is necessary to specify a many-to-one relationship to accommodate the fact that the Nurse Health Assessment took place between waves and that a small proportion of Wave 2 households had split up in the intermittent period. These households still have the same Wave 2 household IDs (**b_hidp**) and can be identified by this and the household split identifier, **b_splitnum**. That is, it takes two variables to uniquely identify households.

```
use b_hhsamp_ns, clear
merge m:1 b_hidp using b_hhresp, keep(1 3) keepusing(b_ncars) nogen
```

10.3 Merging household level data from the nurse health assessment with individual data from the nurse health assessment

To merge household level data to individual level data within the Nurse Health Assessment datasets, it is necessary to include the variable, **b_splitnum**, as a key variable.

```
use pidp b_hidp b_splitnum b_nuroutc using b_indsamp_ns, clear
merge m:1 b_hidp b_splitnum using b_hhsamp_ns, keep(1 3)
keepus(b_nsoutcome) nogen
tab b_nuroutc b_nsoutcome
```

10.4 Merging nurse characteristics to individual level nurse health assessment data

The nurse characteristics, also referred to as nurse paradata, are held on a file, **xivdata,** that is common to both the Wave 2 and 3 nurse health assessments. It is therefore necessary to rename the nurse unique identifier variables from **b_nsintnum** to **nsintnum** in **b_indsamp_ns** and from **c_nsintnum** to **nsintnum in c_indsamp_ns** to merge the data in Stata;

```
use pidp b_nsex b_nsintnum using b_indsamp_ns, clear
rename b_nsintnum nsintnum
merge m:1 nsintnum using xivdata_ns,keep(1 3) nogen
tab b_nsex sex,m

use pidp c_nsex c_nsintnum using c_indsamp_ns, clear
rename c_nsintnum nsintnum
merge m:1 nsintnum using xivdata_ns,keep(1 3) nogen
tab c_nsex sex, m
```

11. Analysis guidance

11.1 Accounting for complex sample design in analysis

11.1.1 Sample design variables and analysis

The sample design involves stratification, clustering and weighting. These design features affect standard errors and should therefore be taken into account in analysis. Appropriate variables are provided to allow the analyst to do this. The weighting variables are described in a separate section. Here we describe the stratification and clustering variables.

11.1.2 Example using Stata

In Stata, to obtain estimates that correctly take into account the sample design the user need only specify the design variables using the svyset command, for example:

```
svyset b_psu [pweight= b_indnsus_xw], strata(b_strata)
```

Then, any compatible command simply needs to be prefixed with svy, for example:

```
svy: logistic depvar  variable1  variable2  variable3
```

11.2 Weighting adjustments for the nurse visit release

Any analysis which relies on the data collected by the nurses should use appropriate weights to adjust for unequal selection probabilities and differential nonresponse. Weighted analysis will adjust for the different sampling fraction across countries and for response rate differences between subgroups of the sample. Considering the complexity of the study design, weights should be selected carefully following the advice provided below. It is important to conduct weighted analyses if your aim is to generalize your results to the population. For advanced users who want to model nonresponse in their own way, we provide a W2 nurse issue weight (**b_indnsus_li**) which reflects the conditions of issue into the nurse dataset the GPS sample and a W2 combined nurse issue weight (**b_indnsub_li**) for the combined GPS+BHPS samples (see Technical Description for more details)

11.2.1 Selecting the correct weight for your analysis

The choice of the weight should reflect the instrument (nurse visit vs. information extracted from blood) and the time to which the analysis refers. This release includes the weights related to nurse visit only.

Table 4 Description of weight variables for the Nurse Health Assessment

Analysis level	**Wave(s)**	**Sample**	**Data source**	**Analysis Weight**
Cross-sectional weights				
Individual	2	GPS	W2 nurse visit and/or W2 nurse visit with W2 full interview	b_indnsus_xw
Individual	3	BHPS	W3 nurse visit and/or W3 nurse visit with W3 full interview	c_indnsbh_xw
Individual	3	BHPS+GPS	W3 nurse visit and/or W3 nurse visit with W2 and/or W3 full interviews	c_indnsub_xw
Longitudinal weights				
Individual	1&2	GPS	W2 nurse visit with W1 and W2 interviews	b_indnsus_lw
Individual	1991 through W3	BHPS	W3 nurse visit with 1991-W3 longitudinal BHPS full interview	c_indns91_lw

Individual	2001 through w3	BHPS	W3 nurse visit with 2001-W3 longitudinal BHPS full interview	c_indns01_lw
Individual	2&3	BHPS+GPS	W3 nurse visit with W2-W3 longitudinal BHPS+GPS full interview	c_indnsub_lw
Individual	1991 through *n*	BHPS	W3 nurse visit with 1991-W$_n$ longitudinal BHPS full interview	*n*_indns91_lw
Individual	2001 through *n*	BHPS	W3 nurse visit with 2001-W$_n$ longitudinal BHPS full interview	*n*_indns01_lw
Individual	2 through *n*	BHPS+GPS	W2 nurse visit with W2-W$_n$ interviews	*n*_indnsub_lw
For advanced users only				
Individual		GPS	W2 nurse visit inclusion weight for GPS	b_indnsus_li
Individual		BHPS+GPS	W2 nurse visit inclusion weight for BHPS+GPS	b_indnsub_li

All weights follow a naming convention, designed to help users select the correct weight. The name of each weight reflects the wave for which the weight is calculated, the level of analysis, the data source and its nature (inclusion weight, cross-sectional analysis weight or longitudinal analysis weight). The rules are described in the "Weighting guidance" section in the documentation for the main survey https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/weighting-guidance .

Some analyses may be cross-sectional: using nurse data alone or nurse data in combination with the questionnaire data. If your analysis uses nurse data alone the best weight is **c_indnsub_xw** – this is a combined weight for the BHPS and GPS samples and therefore has the largest effective sample size. This weight is defined for all sample members who responded to the nurse visit. The other two cross-sectional weights (**b_indnsus_xw** and **c_indnsbh_xw**) are for analyses using nurse data with questionnaire data from specific waves (Wave 2 or Wave 3).

The selection of longitudinal weights depends on which waves are to be analysed.

If using this data with nurse data:	Select this weight:
BHPS 1991 wave through wave 3 mainstage	c_indns91_lw
BHPS 2001 wave through wave 3 mainstage	c_indns01_lw

Wave 1 and 2 mainstage	b_indnsus_lw
Wave 2 and 3 mainstage	c_indnsub_lw

Looking ahead to future waves, if your analysis uses the nurse dataset with longitudinal information from BHPS wave 1991 through any wave (*n*), select the "lw" (longitudinal) version of the 1991 BHPS nurse visit weight (***n*_indns91_lw**). Similarly, for analysis from BHPS 2001 wave through any wave (*n*), the weight is ***n*_indns01_lw**. For analysis from Wave 2 through any wave (n), the weight is ***n*_indnsub_lw**. These weights are defined for sample members who participated in the nurse visit and responded to the full adult interview at all waves between the starting wave and wave *n* (inclusive). The weights will be available on ***n*_indresp** as part of the data release for wave *n*.

The nurse visit weights provided should be used for analysis of nurse visit data alone, or the nurse data together with data from the household-level, proxy or main adult interview, or self-completion questionnaire.

11.2.2 Not using weights

An unweighted analysis does not reflect the population structure correctly unless the assumptions below are true. It is suggested that researchers publishing or presenting unweighted estimates make these assumptions explicit.

If no weighting is used, an analysis which includes nurse visit information assumes:

1. that all estimates of interest are the same in Scotland and Wales as in England;
2. that people who live at an address with more than three dwellings or more than three households are the same as those who don't;
3. that people who gave a main adult interview at Wave 1 are the same with respect to your estimates as those who did not; that people who continued to respond to main adult interview up and including Wave *n* are the same as those who did not; and that people who responded to each particular instrument used in the analysis (nurse visit, blood information) are the same as those who did not. There is a working paper which reports on some of the known differences in response and attrition.[8]

We therefore strongly suggest using weighted analyses at all times when analysing nurse visit data.

[8]Lynn, P., Burton, J., Kaminska, O., Knies, G., & Nandi, A. (2012). An initial look at non-response and attrition in Understanding Society. *Understanding Society working paper series*. Retrieved from https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2012-02.

11.3 Technical Details of computation of weights

This section describes the computation of weights for the nurse visit data. They are ordered as weights for the GPS sample, for the BHPS sample and for combined GPS and BHPS samples. Each weight has been scaled to have a mean of one amongst cases eligible to receive the weight.

11.3.1 Nurse visit weights for the GPS sample

In this section we describe in turn how the weights were derived for:

- W2 nurse visit inclusion weight;
- W2 nurse visit longitudinal weight for Waves 1 and 2;
- W2 nurse visit cross-sectional weight for Wave 2;

W2 nurse visit inclusion weight

The W2 nurse visit inclusion weight consists of three components: the GPS design weight, adjustment for adult main interview nonresponse in Waves 1 and 2, and a selection scaling factor.

The GPS design weight (**a_psnengp_xd**) was created for the UKHLS-GPS sample only. The adjustment reflected nonequal selection probabilities for the addresses with more than three dwellings and dwellings with more than three households.

Next, we created an adjustment for adult main interview nonresponse in Waves 1 and 2. This reflects the fact that the nurse visit was conditional on participation in the Wave 2 main adult interview. The nonresponse correction calculated for the longitudinal (Waves 1 and 2) adult main interview weight (**b_indinus_lw**) was applied (refer to mainstage documentation for the detailed description of its calculation). This was done by separating **b_indinus_lw** into the design weight (**b_psnenus_xd**) and the nonresponse correction. The latter corrects for household-level nonresponse at Wave 1, within household nonresponse to adult main interview at Wave 1, and the conditional nonresponse to adult main interview at Wave 2. This nonresponse correction was multiplied by the GPS design weight (**a_psnengp_xd**) to create a longitudinal adult main interview weight for the GPS subsample of UKHLS.

The selection scaling factor reflects eligibility and selection probability for the issue into the W2 nurse visit. Not everyone was eligible for inclusion in the nurse visit at Wave 2. Noneligible persons (residents of Northern Ireland, the dead, those who moved abroad, women pregnant at the time of the nurse fieldwork, and those who completed the main adult interview at Wave 2 in a language other than English) received a value of 0 for the selection scaling factor. Respondents from Scotland and Wales had a chance of selection that was 1.127/2 times that of respondents from England. Thus respondents from England received the value of 1.127 and Scottish and Welsh respondents – the value of 2.

To create the weight for inclusion into the W2 nurse visit (**b_indnsus_li**) the GPS design weight (**a_psnengp_xd)** was multiplied by the nonresponse correction and the selection scaling factor described above. Given that the inclusion weight is for advanced users, no truncation was implemented. The weight was scaled to the mean of 1.

W2 nurse visit longitudinal weight for Waves 1 and 2

The W2 nurse visit longitudinal weight for Waves 1 and 2 (**b_indnsus_lw**) relies on the inclusion weight (**b_indnsus_li**) and additionally corrects for nonresponse to the nurse visit at W2. Nonresponse correction to the nurse visit is the reverse of the predicted propensity to respond, estimated using backward stepwise logistic regression. The predictors include information from the mainstage Wave 2 household grid and household questionnaire, and the individual adult main questionnaire. The nonresponse correction is multiplied by the inclusion weight to create the W2 nurse visit longitudinal weight for Waves 1 and 2. This weight was truncated (influencing 27 outlying weights) and scaled to a mean of 1.

W2 nurse visit cross-sectional weight for Wave 2

This cross-sectional weight for the Wave 2 nurse visit was calculated via the weight-share method. For the households who had OSMs with an inclusion weight of 0 (mainly due to not providing an adult main interview in Wave 1) or TSMs and at least one OSM with a non-zero **b_indnsus_lw** weight, the weight for all responding adults to the nurse visit was calculated as an average of non-zero **b_indnsus_lw** weights within the household. For all other households the cross-sectional weight is equal to the **b_indnsus_lw** weight. The cross-sectional weight (**b_indnsus_xw**) was then scaled to a mean of 1.

11.3.2 Nurse visit weights for BHPS sample

There are multiple weights to be used with the BHPS sample component. This section describes their calculation.

W3 nurse visit longitudinal weight for BHPS wave 1991 or 2001 through Wave 3

The W3 nurse visit longitudinal weight for BHPS wave 1991/2001 through Wave 3 (**c_indns91_lw** and **c_indns01_lw**) represents respondents who continuously responded to the full interview since 1991/2001 up to and including Wave 3 and also participated in the nurse visit. To create these weights we start with the longitudinal weights for full interviews (**c_indin91_lw** and **c_indin01_lw**) and add the correction for conditional nonresponse to nurse visit. The details of these weights are described in the documentation for the mainstage data (https://www.understandingsociety.ac.uk/documentation/mainstage). The nonresponse correction to the nurse visit is the reverse of the predicted propensity to respond, estimated using backward stepwise logistic regression. The predictors include information from the mainstage Wave 3 household grid and household questionnaire, and the individual adult main questionnaire. The nonresponse correction is multiplied by the longitudinal weights for full interviews (**c_indin91_lw** and **c_indin01_lw**) to create the W3 nurse visit longitudinal weight for waves 1991/2001 through Wave 3 (**c_indns91_lw** and **c_indns01_lw**).

W3 nurse visit cross-sectional weight for BHPS

In order to provide a non-zero weight to as many nurse visit respondents as possible the cross-sectional nurse visit weight for BHPS starts with the Wave 2 enumeration weight (**b_psnenbh_lw**) instead of the cross-sectional Wave 3 full interview weight (**c_indinbh_xw**). We first model the response to the full interview in Wave 3 conditional on enumeration at Wave 2 (**b_psnenbh_lw**). The model is a logistic stepwise regression with predictors from the household questionnaire and

household grid at Wave 2. The reversed predicted response propensities are then multiplied by **b_psnenbh_lw**. The resulting weight is then weight-shared to TSMs and OSMs with zero weights, such that we have an equivalent of a Wave 3 full interview cross-sectional weight. The next step corrects for nonresponse to the nurse visit conditional on a full response to Wave 3. The correction uses logistic stepwise regression with predictors from the Wave 3 household grid and questionnaire, and the individual full interview. The predicted response propensity is reversed and multiplied by the equivalent of Wave 3 full interview cross-sectional weight to create a nurse visit cross-sectional weight for the BHPS (**c_indnsbh_xw**). This weight can be used for analysis of the BHPS sample nurse visit alone, or for the nurse visit together with information from the W3 mainstage full interview.

BHPS nurse visit longitudinal weights after Wave 3

Two longitudinal weights of the form ***n*_indns91_lw** and ***n*_indns01_lw** are provided for longitudinal analyses incorporating information from the nurse visit at Wave 3 and information from mainstage waves starting at wave 1991 or wave 2001 and continuing after Wave 3. These weights are released in ***n*_indresp** dataset as part of the data release for the n^{th} wave. Each such weight is based on W3 nurse visit longitudinal weights (**c_indns91_lw** or **c_indns01_lw**), and additionally corrects for nonresponse to the adult main questionnaire in each wave starting with Wave 3 and continuing up to Wave *n*. The additional nonresponse correction is calculated as part of the mainstage adult main interview nonresponse correction. The details can be found in the section on weighting in the main documentation (https://www.understandingsociety.ac.uk/documentation/mainstage).

11.3.3 Nurse visit weights for BHPS+GPS samples

Analysts may use weights combining the BHPS and GPS sample components. This section describes the calculation of multiple weights which apply to the BHPS and GPS sample components.

BHPS+GPS nurse visit inclusion weight

The BHPS+GPS nurse visit inclusion weight (**b_indnsub_li**) is present on Wave 2 **b_indall** dataset as part of the mainstage release. Unlike many other weights, the BHPS+GPS nurse visit inclusion weight should be used with care and full understanding of sample design for nurse visit. This weight combines the BHPS and GPS samples at the Wave 2 enumeration stage, excluding the probability of selection through the EMB and Northern Ireland samples. But, importantly, it does not include the design correction for the fact that only half of the GPS respondents from Scotland and Wales and around 88.7% of GPS respondents from England were issued to nurse data. The inclusion weight does not incorporate this correction because the above selection depended on giving a full interview in Wave 2. These differential conditional selection probabilities that are part of the design are dealt with appropriately in the nurse response weight (e.g., **c_indnsub_lw**). We suggest that analysts who wish to make their own nonresponse correction start with this weight (**c_indnsub_lw**). Alternatively, additional correction for design needs to be implemented if one uses **b_indnsub_li** with full understanding that selection probability into the nurse visit depended on full response to Wave 2 for the GPS sample.

The BHPS+GPS nurse visit inclusion weight (**b_indnsub_li**) follows the same procedure as the BHPS+UKHLS mainstage inclusion weight (**b_indenub_li**) with a few important exceptions. Both weights relate to respondents who were enumerated at Wave 2 of the mainstage. This means that

an analyst needs to take into account all the nonresponse that has happened between the W2 enumeration and the waves and instruments of interest (in addition to the conditional correction mentioned above).

The details on the BHPS+UKHLS mainstage inclusion weight (**b_indenub_li**) are described in the weighting part of the documentation (https://www.understandingsociety.ac.uk/documentation/mainstage.) A few important differences between **b_indenub_li** and **b_indnsub_li** weights reflect the different design for nurse data compared to the mainstage survey. First, the EMBoost selection probabilities were set to 0 (the EMBoost part of the sample was not issued to a nurse visit). Second, the selection probability was set to 0 for those living in Northern Ireland in 2001 and/or 2009 to reflect that NI subsamples were not issued to a nurse visit. Third, the household response rates and conditional individual response rates between Wave 1 and Wave2 were recalculated so that they represent GPS response rates (instead of combined GPS+EMB response rates). The remaining parts of the procedure were the same as those for **b_indenub_li**.

W3 BHPS+GPS nurse visit longitudinal weight

A W3 BHPS+GPS nurse visit longitudinal weight (**c_indnsub_lw**) is created to be a base weight for the future longitudinal weights (***n*_indnsub_lw**). Because of this we condition it on full response in Wave 2 and Wave 3 in addition to a response to nurse visit. This weight is therefore appropriate for nurse data use in the context of a longitudinal Wave 2-Wave 3 analysis.

Three steps are implemented to create **c_indnsub_lw** weight: 1) nonresponse correction for Wave 2 and Wave 3 full interviews conditional on Wave 2 enumeration; 2) correction for subselection in the GPS part of the sample; 3) nonresponse correction for the nurse visit. This is elaborated in the next paragraphs.

Theoretically we treat the sample at the Wave 2 enumeration stage as one sample with appropriate weights that take into account the selection and response probabilities to be in this sample. We then correct for conditional individual Wave 2 and Wave 3 responses to full interviews for this sample. We now 'pretend' that the sample was split into two parts (BHPS and GPS), and the GPS part was subsampled. We therefore correct for subselection in the GPS, and delete GPS cases that were not issued to nurse visit. In details, the nonresponse correction for combined Wave 2 and Wave 3 full interviews is the reverse of the predicted response propensity conditional on enumeration at Wave 2. The stepwise logistic model uses predictors from the Wave 2 household grid and household questionnaire. This nonresponse correction is then multiplied by the nurse visit inclusion weight (**b_indnsub_li**). For the GPS part of the sample, the design correction is then applied, by multiplying the resulting weight by 2 for the GPS Scotland and Wales samples, and by 1.127 for the GPS England sample. The GPS respondents who have a nonzero weight but were not issued to nurse visit are then deleted. Next, the conditional nurse visit response is predicted using stepwise logistic regression with predictors from the Wave 3 household grid and questionnaire, and the Wave 3 individual full interview. The reverse predicted nurse visit response probability is then multiplied by the above weight to create the Wave 3 combined BHPS+GPS nurse visit longitudinal weight (**c_indnsub_lw**).

BHPS+GPS nurse visit longitudinal weights after Wave 3

Longitudinal weights of the form ***n*_indnsub_lw** is provided for the longitudinal analysis incorporating information from BHPS and GPS nurse visit and information from mainstage waves starting at Wave 2 and continuing after Wave 3. These weights are released in ***n*_indresp** dataset as part of the data release for the n^{th} wave. Each such weight is based on combined BHPS+GPS nurse visit longitudinal weight (**c_indnsub_lw**), and additionally corrects for nonresponse to adult main questionnaire in each wave starting with Wave 3 and up to Wave *n*. The additional nonresponse correction is calculated as part of the mainstage adult main interview nonresponse correction.

BHPS+GPS nurse visit cross-sectional weight

The combined BHPS+GPS cross-sectional weight for nurse visit was calculated via the weight-share method. The weight-share was implemented in two steps, separately for BHPS and GPS samples. This is because these samples had nurse visits conducted in different waves. For households with no TSM (or OSMs with a zero longitudinal weight) the cross-sectional weight is simply equal to the longitudinal weight (**c_indnsub_lw**). For households with a responding TSM (or OSM with zero longitudinal weight) the cross-sectional weight is equal to the household mean of the **c_indnsub_lw** weights. The combined BHPS+GPS nurse visit cross-sectional weight (**c_indnsub_xw**) should be referred for use with combined nurse data.

Each weight has been scaled to have a mean of one amongst cases eligible to receive the weight.

12. Data access

Access to the data is by application through the UK Data Service (UKDS) and two versions are available: End User Licence (SN 7251) and Special Licence (SN 7587). As described elsewhere in this guide some sensitive variables are only available in the Special Licence version. The list of variables only available in the Special License version is available here: https://doc.ukdataservice.ac.uk/doc/7587/mrdoc/pdf/7587_ukhls_ns_sl_documentation.pdf

SN 7251 can be accessed here: https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=7251

and SN 7587 here: https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=7587&type=Data%20catalogue.

Application for the Special Licence version requires the completion of a Special Licence form and subsequent approval from the data owners so please allow additional time for access to this particular dataset over the EUL version. Both versions of the data can be applied for by researchers within and without the UK.

It is not unusual to use the Nurse data in conjunction with the *Understanding Society* mainstage data. The End User Licence version (SN 6614) of the data can be found at the UKDS here: https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6614.

Before applying for the data from the UKDS it is strongly recommended that researchers ensure that the dataset contains the data that they require including whether they require the EUL or Special Licence version. As well as information provided in this guide further information can be found at: https://www.understandingsociety.ac.uk/documentation/health-assessment and full details on the data files and the variables they contain (including whether they are Special Licence variables only) can be found here: https://www.understandingsociety.ac.uk/documentation/health-assessment/dataset-documentation/. Details on the *Understanding Society* mainstage data can be found here: https://www.understandingsociety.ac.uk/documentation/mainstage.

13. Citation

13.1 Citing the data

If you use Understanding Society data you must acknowledge this.

Please visit https://www.understandingsociety.ac.uk/documentation/citation for the citation for the latest version of the data. Please cite each dataset(s) that you use.

All works which use or refer to these materials should acknowledge these sources by means of bibliographic citation. To ensure that such source attributions are captured for bibliographic indexes, citations must appear in footnotes or in the reference section of publications. In addition to the main survey and nurse data, analysts should also cite specific papers included in the polygenic scores and epigenetic ageing algorithms datasets, which document how these measures were created.

13.2 Citing this User Guide

When citing this User Guide, you can use the citation of this particular version quoted below. Note that where an online version is available on the Understanding Society website it is always the most up to date.

Institute for Social and Economic Research. (2025), *Understanding Society: Waves 2 and 3 Nurse Health Assessment, 2010-2012, User Guide, Version 4, June 2025*, Colchester: University of Essex.

14. Further reading

14.1 FURTHER READING: ANTHROPOMETRIC MEASURES

• Adult height and the risk of cause-specific death and vascular morbidity in 1 million people: individual participant meta-analysis. International Journal of Epidemiology. 2012;41(5):1419-33.

• Flint AJ, Rexrode KM, Hu FB, Glynn RJ, Caspard H, Manson JE, et al. Body mass index, waist circumference, and risk of coronary heart disease: a prospective study among men and women. Obesity Research and Clinical Practice. 2010;4(3):e171-e81.

• Gunnell D, Okasha M, Smith GD, Oliver SE, Sandhu J, Holly JM. Height, leg length, and cancer risk: a systematic review. Epidemiological Review. 2001;23(2):313-42.

• Heim N, Snijder MB, Heymans MW, Deeg DJ, Seidell JC, Visser M. Optimal cutoff values for high-risk waist circumference in older adults based on related health outcomes. American Journal of Epidemiology. 2011;174(4):479-89.

• Hirani V. Adult anthropometric measures, overweight and obesity. In: Craig R, Hirani V, editors. Health Survey for England, 2010. London: NHS Information Centre; 2011.

• Hotchkiss JW, Davies CA, Gray L, Bromley C, Capewell S, Leyland A. Trends in cardiovascular disease biomarkers and their socioeconomic patterning among adults in the Scottish population 1995 to 2009: cross-sectional surveys. BMJ Open. 2012;2(3).

• Musaad S, Haynes EN. Biomarkers of obesity and subsequent cardiovascular events. Epidemiological Review. 2007;29:98-114.

• Sternfeld B, Ngo L, Satariano WA, Tager IB. Associations of body composition with physical performance and self-reported functional limitation in elderly men and women. American Journal of Epidemiology. 2002;156(2):110-21.

Ulijaszek S. Anthropometric measures. In: Margetts BM, Nelson M, editors. Design Concepts in Nutritional Epidemiology: Oxford University Press; 1997.

• Visscher TL, Seidell JC. The public health impact of obesity. Annual Review of Public Health. 2001;22:355-75.

• Willett WC. Anthropometric measures and body composition. In: Willett W, editor. Nutritional Epidemiology: Oxford University Press; 1998.

14.2 FURTHER READINGS ON BLOOD PRESSURE, PULSE AND HYPERTENSION

• Bryan S, Saint-Pierre Larose M, Campbell N, Clarke J, Tremblay MS. Resting blood pressure and heart rate measurement in the Canadian Health Measures Survey, cycle 1. Health Reports. 2010;21(1):71-8.

• Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. Lancet. 2005;365(9455):217-23.

• Knott C, Mindell J. Chap. 3 Hypertension. In: Craig R, Mindell J, editors. Health Survey for England 2011 Vol 1, Health, Social Care and Lifestyles. London: National Health Service Information Centre; 2011.

• Krause T, Lovibond K, Caulfield M, McCormack T, Williams B. Management of hypertension: summary of NICE guidance. BMJ. 2011;343:d4891

• Marmot M, Elliott P. Coronary heart disease epidemiology: from aetiology to public health. 2nd ed. New York: Oxford University Press; 2005.

• Tardif JC. Heart rate as a treatable cardiovascular risk factor. British Medical Bulletin. 2009;90:71-84.

• Willett W, Stampfer M. Diet and coronary heart disease. In: Willett W, editor. Nutritional Epidemiology. 3rd ed. New York: Oxford University Press; 2012.

14.3 FURTHER READING ABOUT GRIP STRENGTH

• Cesari M, Penninx BW, Pahor M, Lauretani F, Corsi AM, Rhys Williams G, et al. Inflammatory markers and physical performance in older persons: the InCHIANTI study. Journals of Gerontology A Biolological Sciences Medical Sciences. 2004;59(3):242-8.

• Cohen DD, Voss C, Taylor MJ, Stasinopoulos DM, Delextrat A, Sandercock GR. Handgrip strength in English schoolchildren. Acta Paediatrica. 2010;99(7):1065-72.

• Cooper R, Kuh D, Hardy R. Objectively measured physical capability levels and mortality: systematic review and meta-analysis. BMJ. 2010;341:c4467.

• Cooper R, Kuh D, Cooper C, Gale CR, Lawlor DA, Matthews F, et al. Objective measures of physical capability and subsequent health: a systematic review. Age and Ageing. 2011;40(1):14-23.

• Frederiksen H, Hjelmborg J, Mortensen J, McGue M, Vaupel JW, Christensen K. Age trajectories of grip strength: cross-sectional and longitudinal data among 8,342 Danes aged 46 to 102. Annals of Epidemiology. 2006;16(7):554-62.

• Gjonca E, Tabassum F, Breeze E. Socioeconomic differences in physical disability at older age. Journal of Epidemiology and Community Health. 2009;63(11):928-35.

• Mohd Hairi F, Mackenbach JP, Andersen-Ranberg K, Avendano M. Does socio-economic status predict grip strength in older Europeans? Results from

14.4 FURTHER READINGS ON LUNG FUNCTION

• Akgun KM, Crothers K, Pisani M. Epidemiology and management of common pulmonary diseases in older persons. Journals of Gerontology A Biological Sciences Medical Sciences. 2012;67(3):276-91. Understanding Society: Waves 2 and 3 Nurse Health Assessment, 2010- 2012 32

• Bartley M, Kelly Y, Sacker A. Early life financial adversity and respiratory function in midlife: a prospective birth cohort study. American Journal of Epidemiology. 2012;175(1):33-42.

• Gruffydd-Jones K, Loveridge C. The 2010 NICE COPD Guidelines: how do they compare with the GOLD guidelines? Primary Care Respiratory Journal. 2011;20(2):199-204.

• Mindell J, Chaudhury M, Aresu M, Jarvis D. Chapter 3. Lung function in adults. In: Craig R, Hirani V, editors. Health Survey for England 2010: Volume 1 Respiratory Health. London: The Health and Social Care Information Centre; 2011. p. 1-31.

• Ramsay SE, Whincup PH, Lennon LT, Morris RW, Wannamethee SG. Longitudinal associations of socioeconomic position in childhood and adulthood with decline in lung function over 20 years: results from a population-based cohort of British men. Thorax. 2011;66(12):1058-64.

• Sabia S, Shipley M, Elbaz A, Marmot M, Kivimaki M, Kauffmann F, et al. Why does lung function predict mortality? Results from the Whitehall II Cohort Study. American Journal of Epidemiology. 2010;172(12):1415-23.

• Singh-Manoux A, Dugravot A, Kauffmann F, Elbaz A, Ankri J, Nabi H, et al. Association of lung function with physical, mental and cognitive function in early old age. Age (Dordr). 2011;33(3):385-92.