

User guide for end user license dataset 2019-2020

1 Contents

2	Introduction.....	2
2.1	Background.....	2
2.2	Legal Basis.....	2
3	Methodology.....	3
3.1	Sample Design.....	3
3.2	Questionnaire Design.....	3
4	Data Collection.....	4
5	Notes about particular variables.....	4
5.1.1	Table 1- Routing discrepancies between the model questionnaire and the survey by variable.....	4
6	Weighting.....	5
6.1	Calculation of Weights.....	5
6.2	Using Weights.....	6
7	Comparability to EHIS wave 2.....	6
7.1.1	Table 2- Differences between variables in EHIS wave 2 and wave 3.....	6
8	EHIS datasets.....	7
8.1	Eurostat Dataset.....	7
8.2	End User Licence Dataset.....	7

2 Introduction

2.1 Background

The European Health Interview Survey (EHIS) is a major European Union (EU) reference source for comparative statistics on health status, health determinants and use of health care services.

The first wave of EHIS (EHIS wave 1) was launched under an informal agreement and implemented in 17 Member States (the United Kingdom (UK) did not take part in wave 1) and in Switzerland and Turkey between 2006 to 2009.

The second wave of EHIS (EHIS wave 2) was completed under a European Parliament and Council regulation in all 28 EU Member States, Iceland, Norway, and Turkey between 2013 and 2015.

This third wave of EHIS (EHIS wave 3) was planned for 2019 and was completed by all member states in accordance with the European Parliament and Council regulation. The UK was required to complete EHIS wave 3 (the UK was a member state in 2019), and the Office for National Statistics (ONS) collected the data for the four UK Health partner organisations.

EHIS wave 3 was set up to be comparable with EHIS wave 2. Therefore, the EHIS wave 3 questionnaire is largely based on the questions asked in wave 2, with some replacements to meet emerging EU needs.

2.2 Legal Basis

The legal framework for developing EHIS was the Regulation (EC) No 1338/2008 of the European Parliament and of the Council of 16 December 2008 on Community statistics on public health and health and safety at work. This framework regulation specifies in its annexes the use of population surveys such as EHIS to collect statistics every five years on health status, access and use of healthcare and health determinants.

Detailed specification of the data and metadata to be provided is pursuant to the Commission Implementing Regulation (EU) No. 2018/255 of 19 February 2018 implementing Regulation (EC) No 1338/2008 of the European Parliament and of the Council on Community statistics on public health and health and safety at work, as regards statistics based on EHIS and Commission Implementing Decision (EU) 2018/257 of 19 February 2018 granting derogations to certain Member States with respect to the transmission of statistics pursuant to Regulation (EC) No 1338/2008 of the European Parliament and of the Council, as regards statistics based on EHIS.

Article 3 of the Commission Regulation (EU) No 2018/255 stipulates "To achieve a high level of harmonisation of the survey results across countries, the Commission (Eurostat) shall, in close cooperation with Member States, propose methodological and practical recommendations and guidelines on sampling and the implementation of the survey. These recommendations and guidelines will be set out in a 'European Health Interview Survey Manual', which will include a model questionnaire."

3 Methodology

3.1 Sample Design

Households that had previously taken part in the Labour Force Survey (LFS) were selected to take part in EHIS as a wave 6 follow up study. Sample frames were constructed of households from wave 5 of the main LFS, wave 4 of the local LFS boost (Annual Population Survey- APS) and households aged 75+ from other waves of the main LFS, where 1 or more household member had not objected to recall, and telephone numbers were available.

Systematic sampling was used to select households by arranging the sampling frame with respect to local authorities, postcode, household counts and household identification number information. Samples were drawn separately for GB and Northern Ireland and then combined to get the required total sample size.

Information on the sample design of the LFS can be found in [Volume 1 of the LFS User Guide](#).

3.2 Questionnaire Design

The EHIS wave 3 questionnaire used within the UK was largely based upon the model questionnaire provided as an appendix in the [Eurostat European Health Interview Survey \(EHIS wave 3\) Methodological Manual](#).

The model questionnaire is comprised of the following modules:

- European Health Status Module (EHSM): allows measurement of the health status of the population in general and not only in relation to specific health problems.
- European Health Care module (EHCM): allows measurement of the use of health care services and the unmet needs for health care.
- European Health Determinants Module (EHDM): allows measurement of some aspects in lifestyles or health-related behaviours that may have a positive or negative impact on someone's health.
- Core Social variables Module: allows measurement of a range of demographic and socioeconomic information from respondents.

Two questions in the health modules were modified in the final questionnaire so differ from the model questionnaire, AL3 and AL5. In these questions related to alcohol consumption, 'drinks' was replaced with 'units'. Interviewer guidance for these questions provided information on what is classified as a 'unit' in the UK to help respondents answer the questions accurately. This change was within guidance provided by the methodological manual to ensure the question was interpreted and answered consistently in relation to the UK's official definition of a unit of alcohol.

Following the guidance in the methodological manual, if an interview was being conducted by proxy, the respondent was only asked a specific subset of the health module questions.

A series of standard questions used by ONS were included in the survey to gather the data required to derive the core social variables and to supplement the health questions in the model questionnaire.

4 Data Collection

All data were collected by ONS. Data for England, Scotland and Wales were collected between July 2019 and March 2020 ONS. Data for Northern Ireland were collected between October 2019 and December 2019.

EHIS wave 3 data collection was carried out by ONS Telephone Interviewers who work in a centralised Telephone Unit in Titchfield, Hampshire. They conducted interviews using Computer Assisted Telephone Interviews (CATI) in Blaise computer software.

A total of 16,269 individual interviews were achieved from 12,344 households. This resulted in an overall household response rate of 41.5%.

5 Notes about particular variables

During data processing routing issues with the following variables were found: SK3, SK4, DH5, DH6, PL2, PL4 and PL5. This meant that some respondents who should have been asked these questions were not asked them. During processing, these were recoded to -1 'Not Stated' resulting in high non-response rates for these variables, particularly DH5, DH6 and SK3. Table 1 shows the differences between the model questionnaire routing and the survey routing for each variable.

5.1.1 Table 1- Routing discrepancies between the model questionnaire and the survey by variable

Variable	Model questionnaire routing	Survey routing	Issue
SK3	ASK IF SK1 = 2, 3 or -1	ASK IF SmokEver = 1 AND SK1 = 2 or 3	Those who answered refusal or don't know (-1) to SK1 were not asked SK3 and those who answered 2 to SmokEver were not asked SK3 when they should have been resulting in high levels of not stated.
SK4	ASK IF SK1 = 1 OR (SK1 = 2, 3 or -1 AND SK3 = 1)	ASK IF SK1 = 1 OR (SK1 = 2, 3 AND SK3 = 1)	Those who answered refusal or don't know (-1) to SK1 and 1 to SK3 were not asked SK4 when they should have been.
DH5	ASK ALL	ASK IF DH3 = 1	Only those who answered 1 to DH3 were asked DH5 rather than everyone, resulting in high levels of not stated.
DH6	ASK ALL	ASK IF DH3 = 2	Only those who answered 2 to DH3 were asked DH6 rather than everyone, resulting in high levels of not stated.

PL2	ASK IF PL1 = 1, 2 or -1	ASK IF PL1 = 1 or 2	Those who answered refusal or don't know (-1) to PL1 were not asked PL2 when they should have been.
PL4	ASK IF PL3 = 1, 2 or -1	ASK IF PL3 = 1 or 2	Those who answered refusal or don't know (-1) to PL3 were not asked PL4 when they should have been.
PL5	ASK IF PL3 = 1, 2 or -1	ASK IF PL3 = 1 or 2	Those who answered refusal or don't know (-1) to PL3 were not asked PL5 when they should have been.

6 Weighting

6.1 Calculation of Weights

Since the EHIS sample was taken from the final wave of the LFS/APS, weighting was designed to account for bias caused from individuals dropping out of the LFS/APS after wave 1 (attrition) and not consenting to follow-up after the last wave. The weighting strategy used also considers differential non-response to EHIS itself. Bias from all sources was reduced by using an explicit non-response adjustment and adding controls related to variables associated with health in the calibration stage, in addition to ensuring that the final weights sum to the population totals provided by ONS geography, age and sex and accounting for selection probabilities.

Therefore, the EHIS wave 3 weighting method consisted of the following steps –

1. Design weights were calculated using the inverse of the address-level probability of selection. One household per address was selected and every individual aged 16 and over was included in the weight.
2. The design weights were adjusted for unit non-response to EHIS.
3. The adjusted design weights were then calibrated to population controls.
4. The final weights summed to specified population totals but remained as close as possible to the adjusted design weights.

The additional adjustments for unit non-response and external data of step 2 were carried out separately and in sequence in relation to two factors:

- Quintiles of the Index of Multiple Deprivation by Urban indicator from Census 2011
- Household tenure, using the distribution from the Annual Population Survey (APS) April 2019 – March 2020

In the final weighting stage, the adjusted design weights were calibrated with respect to sex, age groups, geography (NUTS1- The Nomenclature of Territorial Units for Statistics 1), economic activity and a variable derived from HS1 by collapsing categories. The population totals by age, sex and geography are based on ONS population projections. The population totals by economic activity and HS1 collapsed categories are estimates from APS April 2019 – March 2020.

6.2 Using Weights

There are two weight variables included in the dataset, WGT and WGT_SPEC. Both WGT and WGT_SPEC were calculated using the steps outlined above however, WGT is calculated for all cases in the dataset whereas WGT_SPEC is calculated only for cases with non-proxy responses so that they are properly weighted for analysis.

Therefore, WGT_SPEC is appropriate for questions that do not allow proxy, whereas WGT should be used as the weight for variables where proxy was allowed.

7 Comparability to EHIS wave 2

Whilst the EHIS wave 3 questionnaire was largely the same as the wave 2 questionnaire, there were some revisions. Table 2 lists the questions that are different between the two waves, and what the differences are.

7.1.1 Table 2- Differences between variables in EHIS wave 2 and wave 3

Variable	Comparability to wave 2
HS3	Slight revision of question, split to HS3A and HS3B to create HS3 in wave 3, wave 2 asked as one question.
CD2	New for wave 3
CD1p	New for wave 3
AW1	Small change in denomination and answer categories for MAINSTAT that is used for filtering for this question
PL2	Filter modification, PL2 ask of respondents who are not blind in PL1
PL4	Filter modification, PL4 ask of respondents who are not deaf in PL3
PL5	Filter modification, PL4 ask of respondents who are not deaf in PL3
PL8	New for wave 3
PL9	New for wave 3
PC-	Age filter changed from 65+ to 55+
HA-	Age filter changed from 65+ to 55+
HO1	HO1 and HO2 merged into one variable, HO1 for wave 3
HO2	HO3 and HO4 merged into one variable, HO2 for wave 3
AM6	AM6A- slight revision of question, definition broader
PE9	New for wave 3
DH-	DH5 asks separately about juices rather than including them in DH1-4 These variables were called FV in wave 2
DH5	New for wave 3
DH6	New for wave 3
SK1	Slight revision, excluding electronic cigarettes/similar added for Wave 3
SK2	Strong revision, can compare SK2A and SK2B to SK3 in wave 2, but asks for daily smoking of cigarettes exclusively so can only compare on cigarettes
SK3	New for wave 3
SK4	New for wave 3
SK5	Matches to SK4 in wave 2, revision to answer options
SK6	New for wave 3

8 EHIS datasets

8.1 Eurostat Dataset

A quality assured version of the UK EHIS wave 3 dataset will be deposited with Eurostat when the ONS has a data sharing agreement with the EU.

8.2 End User Licence Dataset

To permit the wider use of this dataset several treatments have been applied to permit its deposit with the UK Data Service as an End User Licence dataset.

These treatments include:

- Recoding the Age variable to two-year age bands for those aged 16 & 17 years and 18 & 19 years, then five-year age bands for those aged 20-24 years through to 80-84 years before top coding those aged 85 and over.
- Recoding the occupation variable (JOBISCO) to the 1-digit level to present 10 occupations.
- Recoding the economic sector variable (LOCNACE) to present 9 sectors
- Top coding the household size variable (HHNBERS) at 8 and over
- Top coding the household age band count variable (HHNBERS_0_13) at 4 and over.
- Recoding HATLEVEL to combine ISCED 1 Primary education and ISCED 2 Lower secondary education categories.
- Recoding MAISNTAT to combine compulsory military or civilian service with other category.
- Recoding the existing Region variable from sub-regional NUTS Level 2 to the regional NUTS Level 1.
- Recoding the absence from work variable (AW2) to bands of 1 to 7 days, 8 to 14 days, 15 to 30 days, 31 to 180 days, and more than 181 days.
- Recoding the wearing glasses or contact lenses and difficulty seeing variables (PL1 & PL3) to yes/no responses only.
- Calculating a BMI variables and banding to less than 18.5, 18.5 to less than 25, 25 to less than 30, 30 to less than 40, and 40 or more and deleting the original height and weight variables (BM1 and BM2).
- Recoding the alcohol variable AL5 to combine 0 units a day and 1 unit a day into one category of 0 to 1 unit a day.

9 Documentation

Please see the EHIS_Wave_3_Dataset_Variables for a list of the variables in the dataset.

10 Contact Details

Email adhoc.social.surveys@ons.gov.uk