

GENERAL HOUSEHOLD SURVEY 2005: GUIDELINES FOR USING THE DATA

1) GHS questionnaires

The household and individual questionnaires are combined in one document.

2) GHS data files

The data files include households where a response from one or more individuals was missing. On person level files such individuals can be identified by the variable Schedtyp (where Schedtyp = 3). Two data files have been supplied; one at person level and the other at household level.

Person-level file

ghs05_client contains person-level data. There is one record for every individual, including children, living in each responding household. This file contains schedule variables (that is, variables which were on the original questionnaire) and the derived variables which have been programmed. All household-level variables have also been attached to all individuals in each household on the person level file.

The file includes three serial numbers which denote households, families and people. Hserial is a household identifier; each household has a unique value at Hserial, and this value is attached to everyone in the household. Persno is a person identifier; each person within a household has a separate person number.

Fserial is a family identifier. Each family unit has a unique value at Fserial, and this value is attached to everyone in the family unit.

This file can be used to run tables at person level, household level or family level (see below), but for ease of access we have also supplied a household level file.

Household file

Ghs05_client_hhld contains household data. There is one record per responding household. The file contains schedule variables and derived variables. Some schedule variables have been deleted or grouped to maintain confidentiality.

Creating household and family files from the person level file

The variable HRP is the person number of the HRP so to identify households you need to select on HRP=PERSNO.

For families you use FUH=PERSNO (FUH is the person number of the family unit head).

Weight variable: Weight05

The data set is unweighted. Weight05 is the variable you should use to weight the data (see http://www.statistics.gov.uk/downloads/theme_compendia/GHS05/GHS2005_AppxD.pdf).
(included in study documentation.)

This weight applies to both household and individual level data.

3) Missing values

For each variable, there are three possible missing values. For the majority of variables, these are -6, -8, and -9. No values have been declared missing on the files supplied.

-6 means that the person was not eligible to answer a particular section and is therefore not eligible for inclusion in the relevant derived variable. For example, smoking questions are

only asked of adults, so children are coded as -6 for smoking variables. Adults aged 16-59 are asked the family information questions, so children and adults aged 60 and over are coded as -6 for these variables.

-8 (NA) means 'No answer'. This is used where a respondent either does not know the answer or refuses to answer an individual question.

-9 (DNA) means 'Does not apply'. This is used when a household or respondent is eligible to be asked a question, but is routed past it by the flow of the questionnaire. For example, if a respondent is asked whether he or she is eligible to belong to their employer's pension scheme and the answer is 'no', he or she will be coded as -9 on the follow-up question which asked whether he or she belongs to the scheme.

Missing values are defined differently for the income section, as it is important to be able to distinguish those who have refused the whole section. SPSS only allows the use of three missing values, we have combined -6 and -9. For income variables, the three missing values are -7, -8 and -9.

-7 means that a respondent has refused the whole income section.

-8 (NA) has the same meaning as for other variables.

-9 combines -6 and -9, and therefore includes people who are not eligible for the income section, such as children, and respondents who are routed past questions by the flow of the questionnaire.

Some words of warning

If you are using SPSS-NT, then it is very easy to save amended versions of SPSS files when you don't want to. When you try to quit SPSS, the dialogue box asks if you want to save the file. The intuitive reaction is to click on 'yes', but you should only do this if you want to save any amendments you have made. If you have used a 'keep' or 'select' statement, clicking 'yes' will save only the cases or variables you have selected. We recommend that you keep the CDROM or original file safe, just in case you inadvertently change the data files.

Both data files are large, and it takes time to run frequencies and tabulations on the full files. We recommend that you use a 'keep' statement to select relevant variables.

The Social and Vital Statistics division at ONS is currently using version 12.0 of SPSS. You may find you have problems with missing values when trying to create derived variables. The solution we have used is to declare no missing values, then create the new variables, then declare missing values.

4) Excel spreadsheet (Variables – changes)

The list of variables on the Excel spreadsheet denote changes made between the 2004 and the 2005 datasets including:

- Changes to variable names
- Variables (either derived or questionnaire) that have been dropped
- New variables (either derived or questionnaire) that have been introduced

All changes are listed on the Excel spreadsheet.

5) Specifications for derived variables

Also in the zipped file is a specification for derived variables on the data file.

GHS Team

Social and Vital Statistics Division

March 2007

Changes to GHS questionnaire for 2005

In 2005 the EU made it a legal obligation for member states to collect additional statistics on income and living conditions (EU-SILC). This requirement has been integrated into the GHS. EU-SILC covers data on income, poverty, social exclusion and living conditions. These statistics will be used to help plan and monitor European social policy by comparing poverty indicators and changes over time across the European Community.

The inclusion of questions for EU-SILC has led to large scale changes to the 2005 data set. Some of the questions asked in previous years have been removed to make room for the new questions. This has resulted in changes to both schedule and derived variables summaries of which are included below and full details of which can be found on the accompanying excel spreadsheet.

Household Questionnaire

Accommodation Type

EU-SILC questions on accommodation replaced some of the previously asked questions. This has also led to the removal of variables derived from the old questions. For details of specific variable changes see the Excel sheet "changes 2004 to 2005.xls"

Consumer Durables

This section has seen the largest change for EU-SILC with 56 variables added and 68 removed.

For details of specific variable changes see the Excel sheet "changes 2004 to 2005.xls"

Financial situation

This is a new block for EU-SILC.

Housing costs

This is a new block for EU-SILC.

Individual Questionnaire

Education

Quals The list of qualifications has been updated. It now includes a full list of Scottish qualifications:

- Highers (replace SCE/NQ highers)
- Advanced Highers or Certificate of Sixth Year Studies (CSYS)
- Intermediate 2 National Qualifications (NQs)
- Intermediate 1 NQs
- Standard Grade or O Grade
- SGA

- Access level NQs.

Degree level has been extended to include foundation degrees.

EdExcel (the awarding body for BTEC), vocational A-levels/AS levels/GCSEs, OCR, Key Skills/Basic Skills and Entry Level Qualifications have also been added.

Teach The answer categories 'secondary' and 'primary' education have been replaced by five categories of 'Key Stages' (1-4) and 'Foundation Stage'.

GCSE, NumOL and EngMath have been extended to include the Scottish qualifications, Intermediate 2 at A-C and Intermediate 1 at A-C.

EngMath The interviewer instruction to exclude English Literature at this question was causing some confusion. As the case of a respondent having a grade A-C in English literature and not English Language is unlikely, it has been removed.

SGA is a new question asked if the highest qualification is an SGA (Scotland), and it determines the level of the SGA (Advanced Higher, Higher or Intermediate).

GNVQ The answer categories have been extended to give the option of either 'Part One' or 'Full' Intermediate/Foundation levels.

Appren has an extra answer category (3) 'Yes, has completed one apprenticeship and is now doing a further one.'

Enroll and Course now also include further education (FE) as well as 6th form colleges.

EducPres and Edtyp are now asked of all adults.

Employment

24 new variables added for EU-SILC and 6 2004 variables removed. For details of specific variable changes see the Excel sheet "changes 2004 to 2005.xls"

Health

Variables: consuln1, docwher1, forpern1, gp1, icd1, matnhst1, mtnight1, mtprvst1, nhs1, nhstrea1, presc1, prvstay1, whsbhlf1, seenurse

Renamed as: consulno, docwhere, forperno, gp, icd, matnhstr, mtnights, mtprvsty, nhs, nhstreat, presc, prvstay, whsbhlf, SeeChn

For details of EU-SILC variable changes see the Excel sheet "changes 2004 to 2005.xls"

Income

Variables: othrgpa1 to othrgpa3, othsrcm1 to othsrcm4, reglrpm1 to reglrpm3

Renamed as: othpay1 to othpay3 and othpay4 added, othsrc1 to othsrc4 and othsrc5 added, reglrp3 to reglrp3

For details of EU-SILC variable changes see the Excel sheet "changes 2004 to 2005.xls"

Smoking

agesmk1, agesmk2,, agesmk5 - new smoking age bands
tar05, tar05g1, tar05g2, tarlevel - new tar levels

Drinking

Average weekly alcohol consumption

Questions to establish average weekly alcohol consumption were included in the 2005 survey, and were previously last asked in 2002. They have been included on the GHS in their current form periodically since 1986. The measure was developed in response to earlier medical guidelines suggesting maximum recommended weekly amounts of alcohol. Its use continues to provide a consistent measure of alcohol consumption through which trends can be monitored. Respondents are asked how often over the last year they have drunk normal strength beer, strong beer, wine, spirits, fortified wines and alcopops, and how much they have usually drunk on any one day. This information is combined to give an estimate of the respondent's weekly alcohol consumption (averaged over a year) in units of alcohol.

Family Information

Variables: arewed, chldtyp1, chlvmn1 to chlvmn7, chlivr1

Renamed: wherewed, chldtype, chlvmn to chlvm7, chlivr

Pensions

Variables: occlpen1, perspen1 to perspen4, seprssh1 to seprssh3

Renamed: occpeng1, perspe1 to perspe4, seprss1 to seprss3

Trailers

Trailer on Views of your Local Area removed.

Trailer on Dental Health removed.

New trailer on social mobility added.