

## User Guide for the General Household Survey (GHS) Time Series Dataset 1972 to 2004 (suitable for pseudo-cohort analysis)

The General Household Survey (GHS) Time Series Dataset 1972 to 2004 was created by the Office for National Statistics (ONS). The dataset allows for analysis over time and for pseudo-cohort analysis. This user guide was created by ONS and the Economic and Social Data Service ([ESDS](#))<sup>i</sup> Government. It draws strongly on documentation for the General Household Survey, the GHS reports *Living in Britain*<sup>ii</sup> and other publications produced by the Office for National Statistics (ONS), such as *The GHSPCD: Introduction and Methodology*<sup>iii</sup> by Zoë Uren (originally the GHS Time Series Dataset was referred to as the GHS Pseudo-Cohort Dataset (GHSPCD)).

It would not be feasible to reproduce a detailed introduction to the GHS and therefore wherever possible links are provided to the original source or to sources for further information on the GHS methodology.

### The contents of the GHS Time Series Dataset User Guide

An introduction to the General Household Survey (GHS)	page 2
Introduction to the GHS Time Series Dataset 1972-2004	4
Introduction to the variables included on the GHS Time Series Dataset	6
Using the GHS Time Series Dataset for pseudo-cohort analysis	94

Variable	Page No	Variable	Page No	Variable	Page No
PSCHEDTY	8	PDVMARDF	43	PHRPSEX	62
PHSERIAL	10	PHHTYPE	44	PDVILO3A	64
PPERSNO	11	PHHTYPF	46	PEDFULL	66
PFSERIAL	12	PHHTYF1	51	DEGREE	69
YHSERIAL	13	PNPERSON	53	PNUMVEH	70
YPSERIAL	13	PN0TO4	53	PGENHLTH	72
YFSERIAL	13	PN5TO15	53	PILLNESS	73
YEAR	13	PN16TO59	53	PLIMITAC	74
PRELTOHR	14	PNAGE60	53	PCUTDOWN	76
PCOUNTRY	16	PNAGE65	53	PLONGILL	78
PWEIGHTN	18	PTEENAGE	53	PLSIRA	79
PSSUPWT	19	PN0TO15	53	PDOCTALK	80
PWEIGHTC	20	PYOUNGAG	53	PNCHATS	81
PTENURE	21	PNMALES	53	POUTPATN	82
PCOB1	23	PNFEMALE	53	PDAYPATN	83
PFCOB1	23	PNADMALE	53	PINPATNT	84
PMCOB1	23	PNADFEMS	53	PCIGSMK1	85
PSEGH	31	PAGE5YR	55	PCIGSMK	88
PHRPSEG3	34	PAGEAD1	57	PNOSMOKE	91
PHPSEG3B	36	PAGEAD2	58	PGIVEUP	92
PHRPSEC8	37	PAGEAD3	59	PFIRSTCI	93
PAGE	38	PDVMARD1	60		
PSEX	39	PDVMARD2	61		
ESTYOB	40	PHRPMAR	62		
CHORT5YR	41	PHRPAGE	62		

## **An introduction to the General Household Survey (GHS)**

The General Household Survey (GHS) is a continuous national survey of people living in private households, conducted on an annual basis by the Office for National Statistics (ONS). The main aim of the survey is to collect data on a range of core topics, covering household, family and individual information. This information is used by government departments and other organizations for planning, policy and monitoring purposes and to present a picture of households, family and people in Great Britain. The GHS is sponsored by ONS, Dept of Health, Office of the Deputy Prime Minister, Dept for Work and Pensions, Dept for Education and Skills, Scottish Executive and other government departments.

The GHS is a multi-purpose sample survey on approximately 9,000 households and about 16,000 adults aged 16 and over. Data are collected on five core topics, namely education, employment, health, housing, and population and family information. Other areas such as leisure, household burglary, smoking and drinking are covered periodically.

The GHS started in 1971 and has been carried out continuously since then, except for breaks in 1997-1998 when the survey was reviewed and 1999-2000 when the survey was redeveloped. Following the 1997 review, the survey was relaunched from April 2000 with a different design. The GHS began reporting on a financial-year basis from 1988, instead of a calendar year, and data are published annually. In addition, since 1994 the GHS has been conducted using face to face computer-assisted personal interviewing (CAPI), with some self-completion.

Following its review, the GHS now consists of two elements: the continuous survey and trailers (modules included at certain intervals). The continuous survey is to remain unchanged for the five-year period April 2000-March 2005, apart from essential changes to take account of, for example, changes in benefits and pensions. The GHS has retained its modular structure and this allows a number of different trailers to be included each year, to a plan agreed by sponsoring government departments.

In the early surveys, the GHS sample of addresses in Scotland was boosted to improve estimates in separate analyses for Scotland. The households included in this supplementary sample answered the household questionnaire only and no information was collected at the individual level. From 1994 onwards the Scottish supplement was discontinued (further details of the Scottish supplement are given in the section on Creating the GHS Time Series Dataset and in Appendix 1).

The sample design of the GHS changed in 1984 when the Postcode Address File (PAF) replaced the Electoral Registers as the sampling frame. The GHS uses a two-stage sample design with postcode sectors, which are similar in size to wards, as the Primary Sampling Units (PSUs). From 1986 to 1989 there was provision for rotation of one third of the PSUs each year. The sampling selection procedure from 1984 until the revised survey design was the same. From the year 2000 the revised survey design introduced new stratifiers. The GHS is now stratified by region, proportion of households with no car, proportion of households in SEG 1-5, or SEG 13, proportion of people who were pensioners. For details of the previous

sampling design and for a comprehensive description of how the GHS was, and is, stratified see the [Living in Britain<sup>iv</sup>](#) reports over the years.

Furthermore, following the review of the GHS in 1999, the GHS has subsequently been weighted. In previous years no weights existed, although data for 1998 has been retrospectively weighted using this method (described above). The 1998 weights are not included on the GHS Time Series Dataset, only weights from 2000 onwards are included.

The weights in 2000-2002 and 2004 use a two-step approach:

1) the data are weighted to compensate for nonresponse in the sample based on known under-coverage in the 1991 Census-linked study of non-response.

2) the sample that has been weighted for non-response, is grossed up that it matches the population distribution in terms of region, age-group and sex (population based weighting) based on totals taken from the LFS household population estimates. A full description of the weighting and grossing scheme is given in the [Living in Britain<sup>v</sup>](#) report for each year.

The weights in 2003 use a three-step approach:

The first approach compensated for the higher probability of selection of addresses in the Avoidance Refusal Training scheme - where second or third sample of addresses were selected in PSUs already drawn from the GHS - design weighting. The next stages were the same as those used in the 2000-2002 and 2004 weights (described above).

For more information detailed on weighting for each year of the GHS see the appendix D of the [Living in Britain<sup>vi</sup>](#) reports.

For further general details on the history of the GHS and information how the survey has changed see the [ONS website<sup>vii</sup>](#) and the [Living in Britain<sup>viii</sup>](#) reports.

## **Introduction to the GHS Time Series Dataset 1972-2004**

The GHS Time Series Dataset 1972-2004 has been developed by ONS to analyse trends in social inequalities. The dataset combines annual data series from the GHS, and contains information on demographics, households, education, employment and health since 1972. At present, researchers are able to access data for each individual year the survey is available. The GHS Time Series Dataset has been created by combining each annual round of data into one dataset.

### **Creating the GHS Time Series Dataset**

In creating the GHS Time Series Dataset there were some issues that had to be addressed these are briefly detailed below and are provided in an expanded form in the [Introduction and Methodology](#)<sup>ix</sup> paper by Uren (2006).

#### *Changes to sampling procedures and sample sizes over time*

The sampling procedure used on the GHS has changed over time, resulting in different sample sizes between survey years. However, the changes to the GHS sample procedures and sample size were relatively small. As a result it was decided by ONS that these changes were likely to have little impact on the reliability of the estimates. Particularly as a representative sample of the population has been achieved for each survey year.

#### *Inconsistencies in variables over time*

In general variables in the GHS have remained fairly consistent over time. However as the GHS has been revised and research interests have changed, some variables have been modified over the past 30 years to reflect this. For example the marital status variable was revised in the 1986 survey to include a category for cohabitation. Similarly some questions were only included on a few survey years, or in more recent rounds of the survey series, which limits analysis over time. Those variables that were only available for a few years, or had substantially changed over time were not included on the final dataset. Variables included in the final dataset were recoded and renamed in the GHS Time Series Dataset to match, as far as possible, those used in the most recent GHS with the exception of a prefix of P.

#### *Scottish Supplement sample*

From 1982, the number of supplementary interviews was increased to account for the reduction in size of the main sample. As a result, for analyses covering the whole of Great Britain the sample needed to be weighted. It was decided to exclude all responses from the Scottish supplement or sample from the final dataset. This was principally because there was no individual level information for these households and that no weights were available in the historic data to adjust for this extra sample. In all years except 1973 it was possible to identify those households that were part of the Scottish supplement, and these records were deleted. For 1973, it was necessary to weight the Scottish data to account for these extra records. For further details on the Scottish supplement sample including how it was removed from the GHS Time Series Dataset please see Appendix 1.

### *Missing values*

Summary: The GHS Time Series Dataset will not distinguish between different types of missing values on the database.

The GHS distinguishes 3 different types of missing values:

- a) *Item non-response*. This is when a respondent is expected to answer a question but does not. This may be because they don't know the answer or refuse to give an answer.
- b) *Does not apply* – this is when the respondent is included in the section of questions, but is not asked that specific question. For example, only smokers are asked the question on the number of cigarettes they smoke. Therefore none responders will be given the "Does not apply" missing code.
- c) *Not asked the section*. The respondent may be excluded from the whole section because of a characteristic. For example, children, proxies and those in the Scottish supplement may be excluded from particular sets of questions.

Ideally it would be useful to distinguish between a) and b) & c) for the GHS Time Series Dataset, however, on the GHS datasets the codes assigned to missing values are not always consistent over the same variable over time nor are they consistent over different variables in each year.

The GHS Time Series Dataset does not distinguish between any of these values since from investigations of missing values by ONS for most variables there is very little item non-response. The large majority of variable have no or very minimal non-response. For most variables non-response among those asked the question is less than 1%. The exception is the income variable where non-response is 11%. Further to this the missing value information on the GHS variables does not always tell us what we want to know - i.e. if there was an item non-response to a first question, in the subsidiary question this answer could appear as a "Does not apply" because the respondent was not routed to the question.

## Introduction to variables included in the GHS Time Series Dataset

The table below provides an overview of the variables that are available on the GHS Time Series Dataset and the years for which they are included on the Dataset. The following pages provide the metadata information for each of the variables, including information on the variables used to create the new GHS Time Series Dataset variables and gives the SPSS syntax which was used to recode the variable.

In order to create the GHS Time Series Dataset there was an initial recoding stage where all the variables for each of the years were renamed to take the name of the new GHS Time Series Dataset variable names. The new variable names are based on the variable names used in the most recent year with a Prefix of a P (for example, PHHTYPF).

The SPSS syntax provided in the metadata information for each variable recodes the old variable values to the new values for consistency over time (the original SPSS recode syntax used to rename the variable is not included). For several of the variables the recode required the creation of or the use of variable which formed an in between step, these variables which formed an in between step are not available on the final Dataset.

Topic	Variable	Variable name	Period available
Sample information	PSCHEDTY	Full or proxy interview	1972-2004
	PHSERIAL	Serial number - household	1972-2004
	PPERSNO	Person Identifier	1972-2004
	PFSERIAL	Family Identifier	1972-2004
	YHSERIAL	year & household serial number	1972-2004
	YPSERIAL	year & person serial number	1972-2004
	YFSERIAL	year & family serial number	1972-2004
	PRELTOHR	Relationship to HOH/HRP	1972-2004
	PCOUNTRY	UK country	1972-2004
	YEAR	survey year (financial year from 1988 onwards)	1972-2004
	PWEIGHTN	Non-response weight	2000-2004
	PSSUPWT	Scottish supplement weight	1972-2004
	PWEIGHTC	Combined weight	1972-2004
Tenure	PTENURE	Tenure	1972-2004
Demographics	PCOB1	Country of birth	1972-2004
	PFCOB1	Fathers country of birth	1972-2004
	PMCOB1	Mothers country of birth	1972-2004
	PSEGH	Socio-economic group of HOH/HRP	1972-2004
	PHRPSEG3	Socio-economic group of HOH/HRP - 7 categories	1972-2004
	PHPSEG3B	Socio-economic group of HOH/HRP - 2 categories	1972-2004
	PHRPSEC8	8 classes of NS SEC of HRP	2001-2004
	PAGE	AGE	1972-2004
	PSEX	SEX	1972-2004
	ESTYOB	Estimated Year of Birth	1972-2004
	CHORT5YR	5-year estimated birth cohort	1972-2004
	PDVMARDF	Marital Status	1972-2004
	PHHTYPE	Household Type A (Grouped)	1974-2004
	PHHTYPF	Household type F	1984-2004

	PHHTYF1	Household type F (Grouped)	1984-2004
	PNPERSON	No. of persons in the household	1972-2004
	PN0TO4	No. of children aged <5 in household	1972-2004
	PN5TO15	No. of children aged 5 to 15 in household	1972-2004
	PN16TO59	No. of adults aged 16 to 59 in household	1972-2004
	PNAGE60	No. of adults aged 60 and over in household	1972-2004
	PNAGE65	No. of adults aged 65 and over in household	1972-2004
	PTEENAGE	No. of teenagers in household	1972-2004
	PN0TO15	No. of children aged <16 in household	1972-2004
	PYOUNGAG	Youngest age of person in the household	1972-2004
	PNMALES	No. of males in household	1972-2004
	PNFEMALE	No. of females in household	1972-2004
	PNADMALE	No. of adult males in household	1972-2004
	PNADFEMS	No. of adult females in household	1972-2004
	PAGE5YR	AGE OF ADULTS GROUPEd (5 YEARS)	1972-2004
	PAGEAD1	AGE OF ADULTS GROUPEd (6 BANDS)	1972-2004
	PAGEAD2	AGE OF ADULTS GROUPEd (7 BANDS)	1972-2004
	PAGEAD3	AGE OF ADULTS GROUPEd (4 BANDS)	1972-2004
	PDVMARD1	Marital Status (4 Groups)	1972-2004
	PDVMARD2	Marital Status (3 Groups)	1972-2004
	PHRPSEX	sex of HRP/HOH	1972-2004
	PHRPMAR	marital status of HRP/HOH	1972-2004
	PHRPAGE	age of HRP/HOH	1972-2004
<b>Labour market</b>	PDVILO3A	DV for ILO in employment - 3 categories	1972-2004
<b>Education</b>	PEDFULL	Education Level - full form	1972-2004
	DEGREE	Highest qualification a degree or higher	1972-2004
<b>Car ownership</b>	PNUMVEH	Number of cars	1972-2004
<b>Health</b>	PGENHLTH	Health on the whole in last 12 months	1977-2004
	PILLNESS	Any longstanding illness or disability	1973-2004
	PLIMITAC	If longstanding illness limits activity	1973-2004
	PCUTDOWN	Illness/Injury reduce activity	1972-76, 1979-2004
<b>Health Services</b>	PDOCTALK	Consulted doctor last 2 wks (exc hospital)	1972-2004
	PNCHATS	Number of GP consultations in the last 2 weeks	1972-2004
	POUTPATN	Hospital outpatient attend - last 3 mths	1972-2004
	PDAYPATN	In hospital as day patient in last year	1992-2004
	PINPATNT	Hospital in patient last year	1972-2004
	PLONGILL	LLSI OR NON-LLSI	1972-2004
	PLSIRA	IF LLSI or restricted activity	1972-2004
<b>Cigarette smoking</b>	PCIGSMK1	smoking status (ever smoked)	1972-2004
	PCIGSMK	Number of cigarettes smoked per day	1972-2004
	PNOSMOKE	Go without smoking for whole day	1992, 1994, 1996-2004
	PGIVEUP	Like to give up smoking altogether	1992, 1994, 1996-2004
	PFIRSTCI	When first cigarette of day	1992, 1994, 1996-2004

## PSCHEDTY 'Full or proxy interview'

- 1 Full interview
- 2 Proxy
- 3 Eligible but NC or Refusal
- 4 Child
- 5 Missing
- 6 No information available for this year

*Names of the original GHS variables used to create PSCHEDTY:*  
 PROXY, COMPLETE and SCHEDTYP

### Notes:

The variable PROXY in 1972 used to create PSCHEDTY does not have a category for "Eligible but NC or refusal"

Information on schedule type is not available for 1973-1976 and 1979. Therefore in these years, a variable has been created, so that records for children can easily be distinguished. (3= 'Child' and 6= 'Information not available'). Please note that from the recode 'Child' is computed straight from the 'age' variable (PAGE).

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS					New codes		
	1972	1977, 1980, 1978, 1982	1983	1996	1998		2001	2002
	proxy	Complete	SCHEDTYP	SCHEDTYP	SCHEDTYP	SCHEDTYP	SCHEDTYP	PSCHEDTY
person was interviewed	0							1
info from other HH member	1							2
Fully Co-Op"		1						1
ELIGIBLE;PROXY COMP"		2						2
ELIGIBLE;NON-CONTACT"		3						3
ELIGIBLE;REFUSAL"		4						3
CHILD UNDER 16"		9						4
CHILD			-9		-6		0	4
Full interview			1		1		1	1
Proxy			2		2		2	2
Eligible but NC or Refusal			3		3		3	3

### SPSS syntax to recode the variable PSCHEDTY

```

*de-select missing values*.
MISSING VALUES pschedty ().
EXE.
*select missing values for page*.
MISSING VALUES page (-5).
EXE.

***in 1972 page <16 will select those aged 0-16***.
***however, those with a value of -5 will not be selected as these have been declared as missing**.
DO IF (year=1972).
+   DO IF (page<16).
+       COMPUTE pschedty=4.
+   ELSE.
+       RECODE pschedty (0=1)(1=2)(ELSE=5).
+   END IF.

ELSE IF (year>=1973 and year<=1976).
+   DO IF (page<16).
+       COMPUTE pschedty=4.
+   ELSE.
+       COMPUTE pschedty=6.
+   END IF.

ELSE IF (year=1979).

```

```

+ DO IF (page<16).
+ COMPUTE pschedty=4.
+ ELSE.
+ COMPUTE pschedty=6.
+ END IF.

ELSE IF (year>=1977 and year<=1978).
+ DO IF (page<16).
+ COMPUTE pschedty=4.
+ ELSE.
+ RECODE pschedty (1 thru 3=Copy)(4=3)(ELSE=5).
+ END IF.

ELSE IF (year>=1980 and year<=1982).
+ DO IF (page<16).
+ COMPUTE pschedty=4.
+ ELSE.
+ RECODE pschedty (1 thru 3=Copy)(4=3)(ELSE=5).
+ END IF.

ELSE IF (year>=1983 and year<=2004).
+ DO IF (page<16).
+ COMPUTE pschedty=4.
+ ELSE.
+ RECODE pschedty (1 thru 3=Copy)(ELSE=5).
+ END IF.

END IF.

*de-select missing values*.
MISSING VALUES page (.).
EXE.

VARIABLE LABELS pschedty 'Full or proxy interview'.
VALUE LABELS pschedty 1 'Full interview' 2 'Proxy' 3 'Eligible but NC or Refusal' 4 'Child' 5 'Missing' 6
'No information available for this year'.
EXECUTE .

```

**PHSERIAL** 'Serial number - household'

**Definition of household (from *Living in Britain 2002*)**

A household is defined as: A single person or a group of people who have the address as their only or main residence and who either share one meal a day or share the living accommodation.

A group of people is not counted as a household solely on the basis of a shared kitchen or bathroom.

A person is in general regarded as living at the address if he or she (or the informant) considers the address to be his or her main residence. There are, however, certain rules, which take priority over this criterion.

- (a) Children aged 16 or over who live away from home for purposes of either work or study and come home only for holidays are not included at the parental address under circumstances.
- (b) Children of any age away from home in a temporary job and children under 16 at boarding school are always included in the parental household.
- (c) Anyone who has been away from the address continuously for six months or longer is excluded.
- (d) Anyone who has been living continuously at the address for six months or longer is included even if he or she has his or her main residence elsewhere.
- (e) Addresses used only as second homes are never counted as a main residence.

Each household (as defined above), in each survey year has been assigned a number.

*Names of the original GHS variables used to create PHSERIAL:*

1972-1982: HOUSENO

1983-1998: HSERNO

2000-2004: HSERIAL.

No changes were made to this variable other than the name change.

Notes:

The same identifiers may be used in different years. Therefore another variable has been created which combines year & household serial number, so that each household as a unique identifier.

Other similar variables on the GHS Time Series Dataset:

PPERSNO	Person identifier
PFSERIAL	Family identifier
YHSERIAL	year & household serial number
YPSERIAL	Year, household & person serial number
YFSERIAL	Year, household & family serial number

*SPSS syntax to recode the variable PHSERIAL*

```
*de-select missing values*.
MISSING VALUES phserial (.).
EXE.

RECODE
  phserial
  (0 thru HIGHEST=Copy) (ELSE=-5).
VARIABLE LABELS phserial 'Serial number - household'.
VALUE LABELS phserial -5 'Missing'.
EXE.
```

### **PPERSNO 'Person identifier'**

Each person living in the household numbered from 1 upwards, regardless of whether they belong in different family units. For example:

Household Identifier	Family Identifier	Person Identifier	Family member
1009191	10091911	1	Adult (Father)
1009191	10091911	2	Adult (Mother)
1009191	10091911	3	Child aged 14
1009191	10091911	4	Child aged 16
1009191	10091912	5	Foster child aged 18

*Names of the original GHS variables used to create PPERSNO:*

1972-1982: PERNO

1983-2004: PERSNO

No changes were made to this variable other than the name change.

#### Notes:

The person identifier is not unique for different years and therefore a new variable has been created with combined this variable with the variable YEAR 'year'.

Other similar variables on the GHS Time Series Dataset:

PHSERIAL	Household serial number
PFSERIAL	Family identifier
YHSERIAL	Year & household serial number
YPSERIAL	Year, household & person serial number
YFSERIAL	Year, household & family serial number

#### *SPSS syntax to recode the variable PPERSNO*

```
*de-select missing values*.
MISSING VALUES ppersno (.).
EXE.

RECODE
  ppersno
  (0 thru HIGHEST=Copy) (ELSE=-5).
VARIABLE LABELS ppersno 'Person Identifier'.
VALUE LABELS ppersno -5 'Missing'.
EXECUTE .
```

## **PFSERIAL 'Family Identifier'**

### **Definition of family unit (from [Living in Britain 2002](#))<sup>x</sup>**

A GHS family unit is defined as:

1. A married or opposite sex cohabiting couple on their own; or
2. A married or opposite sex cohabiting couple, or a lone parent, and their never-married children (who may be adult), provided these children have not children of their own.

Persons who cannot be allocated to a family as defined above are said to be persons not in the family i.e. as 'non-family units'.

In general, GHS family units cannot span more than two generations i.e. grandparents and grandchildren cannot belong to the same family unit. The exception to this is where it is established that the grandparents are responsible for looking after the grandchildren (e.g. when the parents are abroad). Adopted and stepchildren belong to the same family unit as their adoptive/step-parents. Foster-children, however, are not part of their foster-parents' family (since they are not related to their foster-parents) and are counted as separate non-family units.

*Names of the original GHS variables used to create PFSERIAL:*

1972-1996/97: FAMUNIT

1998/99 onwards: FSERIAL

Pfserial is a one digit variable in 1972-1996, and is the household serial number plus one digit in 1998-2004.

### **Notes:**

For the GHS Time Series Dataset another variable YFSERIAL was created, where the first 4 digits represent the Year, the next 8 digits represent the 'household serial number' and the final digit represent the 'family serial number'. Please see the variable YFSERIAL for further details.

Other similar variables on the GHS Time Series Dataset:

PHSERIAL	Household serial number
PPERSNO	Person identifier
YHSERIAL	Year & household serial number
YPSERIAL	Year, household & person serial number
YFSERIAL	Year, household & family serial number

*SPSS syntax to recode the variable PFSERIAL*

```
*de-select missing values*.
MISSING VALUES pfserial ().
EXE.

RECODE
  pfserial
  (0 thru HIGHEST=Copy) (ELSE=-5) .
VARIABLE LABELS pfserial 'Family Identifier'.
VALUE LABELS pfserial -5 'Missing'.
EXECUTE .
```

## ***The Unique Identifiers***

### **YHSERIAL 'Year & Household Serial Number'**

The first 4 digits represent the Year

The next 8 digits represent PHSERIAL the 'household serial number'

See PHSERIAL for more information about the household serial number.

### **YPSERIAL 'Year & Person Serial Number'**

The first 4 digits represent the Year

The next 8 digits represent PHSERIAL the 'household serial number'

The last 2 digits are PERSNO 'the person number'

See PERSNO for more information about the person identifier.

### **YFSERIAL 'Year & Family Serial Number'**

The first 4 digits represent the Year

The next 8 digits represent PHSERIAL the 'household serial number'

The final digit represents PFSERIAL the 'family serial number'

See the variable PFSERIAL for more information about the Family ID.

#### **SPSS syntax to recode the 3 variables: YHSERIAL, YPSERIAL & YFSERIAL**

**\*\*Syntax to create the unique identifiers\*\*.**

**\*\*The first 4 digits are the year. The next 8 digits are the household serial\*\*.**

**\*\*For the 'year & person serial number' the last 2 digits are the person number\*\*.**

**\*\*For the 'year & family serial number' the last digit is the family number\*\*.**

Compute YHserial = year \* 100000000 + phserial.

EXE.

Compute YPserial = YHserial \* 100 + ppersno.

EXE.

**\*\*pfserial is a one digit variable in 1972-1996, and is the household serial number plus one digit in 1998-2004\*\*.**

DO IF (year >= 1972 and year <= 1996).

Compute YFserial = YHserial \* 10 + pfserial.

ELSE IF (year >= 1998 and year <= 2004).

Compute YFserial = year \* 100000000 + pfserial.

END IF.

EXE.

FORMAT YHserial (F12.0).

FORMAT YPserial (F14.0).

FORMAT YFserial (F13.0).

VARIABLE LABELS

YHserial 'year & household serial number'

YPserial 'year & person serial number'

YFserial 'year & family serial number'

EXECUTE .

### **YEAR 'Survey year (financial year from 1988 onwards)'**

The variable YEAR was added to GHS Time Series Dataset at an earlier stage in the creation of the dataset (before all the GHS data was merged).

## PRELTOHR 'Relationship to HOH/HRP'

0	Head of household
1	Spouse
2	Cohabitee
3	Son/daughter (inc adopted)
4	Step-son/daughter
5	Foster child
6	Son-in-law/daughter-in-law
7	Parent/guardian
8	Step-parent
9	Foster parent
10	Parent-in-law
11	Brother/sister (incl. adopted)
12	Step-brother/sister
13	Foster brother/sister
14	Brother/sister-in-law
15	Grand-child
16	Grand-parent
17	Other relative
18	Other non-relative
-5	Missing

*Names of the original GHS variables used to create PRELTOHR:*  
HOHRELN, RELTOHOH and RELTOHRP

### Notes:

For years 1996 and 2001-2004 no value labels were included.

In the original GHS datasets the HOH/HRP is normally always coded 0, however in the years 1994 to 1996 HOH was coded as -9. This has been coded to 0 so it is comparable with other years. This may mean that some missing categories may be miscoded as Head of household, although this will only affect on a few cases.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS							New codes
	1972	1983	1984	1986	1994	1998	2000	
	1982	1985	1993	1996	2004	PRELTOHR		
HEAD OF HOUSEHOLD	0		0	0	-9	0	0	0
SPOUSE OF HOH	1	1	1	1	1	1	1	1
CHILD,NOT FOSTER	2	2	2	2	3	3	3	3
STEP-SON DAUGHTER					4	4	4	4
FOSTER CHILD					5	5	5	5
SON,D-IN LAW	3	3	3	3	6	6	6	6
PARENT	4	4	4	4	7	7	7	4
STEP PARENT					8	8	8	8
FOSTER PARENT					9	9	9	9
PARENT-IN-LAW	5	5	5	5	10	10	10	10
BROTHER,SISTER	6	6	6	6	11	11	11	11
STEP-SIBLING					12	12	12	12
FOSTER-SIBLING					13	13	13	13
SIBLING-IN-LAW					14	14	14	14
GRANDCHILD	7	7	7	7	15	15	15	15
GRANDPARENT					16	16	16	16
OTHER RELATIVES	8	8	8	8	17	17	17	17
NON RELATIVES	9	9	9	9	18	18	18	18
CO-HABITEE OF HOH				11	2	2	2	2
Missing values	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	-5

*SPSS syntax to recode the variable PRELTOHR*

```
*de-select missing values*.  
MISSING VALUES preltohr ().  
EXE.
```

```
DO IF (year<=1993).  
+ RECODE preltohr (0 thru  
1=Copy)(2=3)(3=6)(4=7)(5=10)(6=11)(7=15)(8=17)(9=18)(11=2)(ELSE=-5).  
ELSE IF (year=1994).  
+ RECODE preltohr (0 thru 18=Copy)(-9=0)(ELSE=-5).  
ELSE IF (year=1995).  
+ RECODE preltohr (0 thru 18=Copy)(-9=0)(ELSE=-5).  
ELSE IF (year=1996).  
+ RECODE preltohr (0 thru 18=Copy)(-9=0)(ELSE=-5).  
ELSE IF (year>=1997).  
+ RECODE preltohr (0 thru 18=Copy)(ELSE=-5).  
END IF.
```

```
VARIABLE LABELS preltohr 'Relationship to HOH/HRP'.
```

```
VALUE LABELS preltohr
```

0	'Head of household'
1	'Spouse'
2	'Cohabitee'
3	'Son/daughter (inc adopted)'
4	'Step-son/daughter'
5	'Foster child'
6	'Son-in-law/daughter-in-law'
7	'Parent/guardian'
8	'Step-parent'
9	'Foster parent'
10	'Parent-in-law'
11	'Brother/sister (incl. adopted)'
12	'Step-brother/sister'
13	'Foster brother/sister'
14	'Brother/sister-in-law'
15	'Grand-child'
16	'Grand-parent'
17	'Other relative'
18	'Other non-relative'
-5	'Missing'.

```
EXECUTE .
```

## PCOUNTRY 'UK Country'

- 1 England
- 2 Wales
- 3 Scotland
- 5 Missing

*Names of the original GHS variables used to create PCOUNTRY:*  
 REGION, REG8417, REG17, COUNTRY1 and COUNTRY

### Notes:

The Scottish supplement was removed at a stage prior to PCOUNTRY being derived. However, in 1973 the variable REGION does not include a value for Scottish supplement and therefore could not be removed from the dataset (see PSSUPTW for more information for year 1973).

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS							New codes
	1972	1978	1984		1990	1998	2000	
	1977	1983	1988	1989	1996		2003	<b>PCOUNTRY</b>
	REGION	REGION	REGION	REG8417	REG17	COUNTRY1	COUNTRY	
NORTH	1	1						1
NORTH MET			1					1
NORTH N-MET			2					1
YORKS, HUMBERSIDE	2	2						1
YORKS & HUMB MET			3					1
YORK & HUM N-MET			4					1
NORTH WEST	3	3						1
NORTH WEST MET			5					1
NTH WEST N-MET			6					1
EAST MIDLANDS	4	4	7					1
WEST MIDLANDS	5	5						1
WEST MIDS MET			8					1
WEST MIDS N-MET			9					1
EAST ANGLIA	6	6	10					1
LONDON	7	7						1
GR LONDON INNER			11					1
GR LONDON OUTER			12					1
SOUTH EAST	8							1
SOUTH EAST OMA		8	13					1
SOUTH EAST REM		9	14					1
SOUTH WEST	9	10	15					1
WALES	10	11						2
WALES I			16					2
WALES II			17					2
SCOTLAND	11	12						3
MAIN SCOT I			18					3
MAIN SCOT II			19					3
M SCOT 3 GLSGOW			20					3
M SCOT 3 OTHER			21					3
MAIN SCOT IV			22					3
England				1	1	1	1	1
Wales				2	2	2	2	2
Scotland						3	3	3
Scotland Main				3	3			3

Missing values	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	-5
----------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	----

*SPSS syntax to recode the variable PCOUNTRY*

```
*de-select missing values*.
MISSING VALUES pcountry (.).
EXE.
```

```
DO IF (year>=1998).
+   RECODE pcountry (1 thru 3=Copy) (ELSE=-5) .
ELSE IF (year>=1989 and year<=1996).
+   RECODE pcountry (1 thru 3=Copy) (ELSE=-5).
ELSE IF (year>=1984 and year<=1988).
+   RECODE pcountry (1 thru 15=1) (16 thru 17=2) (18 thru 22=3) (ELSE=-5).
ELSE IF (year>=1980 and year<=1983).
+   RECODE pcountry (1 thru 10=1) (11=2) (12=3) (ELSE=-5).
ELSE IF (year>=1978 and year<=1979).
+   RECODE pcountry (1 thru 10=1) (11=2) (12=3) (ELSE=-5).
ELSE IF (year>=1972 and year<=1977).
+   RECODE pcountry (1 thru 9=1) (10=2) (11=3) (ELSE=-5).
END IF.
```

```
VARIABLE LABELS pcountry 'UK country'.
```

```
VALUE LABELS pcountry
  1      'England'
  2      'Wales'
  3      'Scotland'
 -5      'Missing'.
```

```
EXECUTE .
```

**PWEIGHTN 'Non-response weight'**

*Names of the original GHS variables used to create PWEIGHTN:*  
 WEIGHT00, WEIGHT01, WEIGHT02, WEIGHT03 and WEIGHT04

**Notes:**

Non-response weights have been attached to the GHS Time Series Dataset for the years where these exist (from 2000 onwards). For remaining years the weight has been set to 1.

For more information see on weighting used in the GHS see the section 'An introduction to the General Household Survey (GHS)' or any of the [Living in Britain<sup>xi</sup>](#) reports over the years.

In addition, in 2001 the weighting variable used in Living in Britain is different from the weight on the data archive. The weights used in Living in Britain report are derived from the 2000 LFS household population estimates whereas the weight on the data archive uses the 2001-based intermediate population estimates from LFS. See [Appendix D<sup>xii</sup>](#) from the 2002 LIB report. The 2001 LIB and subsequent reports show data weighted using the 2000 LFS pop estimates. The archive has the 'published' weight for 2001 and a revised weight for 2002 which was derived from 2001-based intermediate population estimates.

There are no missing values on the weight variables.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS					New codes
	2000	2001	2002	2003	2004	
		WEIGHT00	WEIGHT01	WEIGHT02	WEIGHT03	WEIGHT04
NUMERIC	√	√	√	√	√	√

```

SPSS syntax to recode the variable PWEIGHTN

**The weight values contain no missing information, therefore -5s were removed from the syntax.
*****
*de-select missing values*.
MISSING VALUES pweightn ().
EXE.

DO IF (year<=1998).
COMPUTE
    pweightn=1.
END IF.
EXE.

DO IF (year>=2000).
RECODE
    pweightn
    (LOWEST thru HIGHEST=Copy).
END IF.
EXE.

VARIABLE LABELS pweightn 'Non-response weight'.
EXECUTE.
    
```

## PSSUPWT 'Scottish supplement weight'

*Names of the original GHS variables used to create PSSUPWT:*  
YEAR and the GHS time series variable PCOUNTRY

### Notes:

This weight was created for the year 1973 where the Scottish supplement can not be identified and exclude. Data for Scotland has been downweighted in the year.

For more information on the Scottish supplement see Appendix 1 on 'Further information on the Scottish supplement'.

### *SPSS syntax to recode the variable PSSUPWT*

```
**Data for Scotland has been downweighted in years where Scottish Supplement data cannot be
identified for exclusion.
**a weight of 1 has been applied for all other years.
**The weight values contain no missing information, therefore -5s were removed from the syntax.
*****.
```

```
DO IF (year=1973).
+   DO IF (pcountry=3).
+       COMPUTE pssupwt=0.5.
+   ELSE.
+       COMPUTE pssupwt=1.
+   END IF.
```

```
ELSE.
+   COMPUTE pssupwt=1.
END IF.
```

```
VARIABLE LABELS pssupwt 'Scottish supplement weight'.
EXECUTE.
```

## PWEIGHTC 'combined weight'

*Names of the GHS time series variables used to create PWEIGHTC:*  
PWEIGHTN and PSSUPWT

### Notes:

PWEIGHTC: Combines PWEIGHTN and PSSUPWT.

Firstly PWEIGHTN (applicable from 2000 onwards) is scaled (i.e. make the average weight = 1). In order to scale the weights a scaling factor is created which is:

=number of cases (unweighted base)/sum of the weights (weighted base)

The original weight is then times by this scaling factor - to create the new variable PWEIGHTC.

For 1973 there is the weight PSSUPWT is added to account for the Scottish sample.

For more information see the metadata sheets for PWEIGHTN and PSSUWT.

### SPSS syntax to recode the variable PWEIGHTC

```
***the following gets the number of cases (unweighted base) and the sum of the weights (weighted base).
```

```
***this only needs to be run once, this has been included for info.
```

```
*SORT CASES BY year .
```

```
*SPLIT FILE.
```

```
*LAYERED BY year.
```

```
*DESCRIPTIVES.
```

```
* VARIABLES=pweightn.
```

```
* /STATISTICS=SUM .
```

```
*****In order to scale the weights we need to create a scaling factor = number of cases (unweighted base) /sum of the weights (weighted base).
```

```
**2000/1 = 19266 / 57109168 = 0.0003373539.
```

```
**2001/2 = 21180 / 57259597 = 0.0003698943.
```

```
**2002/3 = 20149 / 56570394 = 0.0003561757.
```

```
**2003/4 = 24489 / 56720724 = 0.0004317470.
```

```
**2004/5 = 20421 / 56873454 = 0.0003590603
```

```
*****.
```

```
DO IF (year=2000).
```

```
+ COMPUTE pweightc = pweightn * 0.0003373539.
```

```
ELSE IF (year=2001).
```

```
+ COMPUTE pweightc = pweightn * 0.0003698943.
```

```
ELSE IF (year=2002).
```

```
+ COMPUTE pweightc = pweightn * 0.0003561757.
```

```
ELSE IF (year=2003).
```

```
+ COMPUTE pweightc = pweightn * 0.0004317470.
```

```
ELSE IF (year=2004).
```

```
+ COMPUTE pweightc = pweightn * 0.0003590603.
```

```
ELSE.
```

```
+ COMPUTE pweightc = pssupwt.
```

```
END IF.
```

```
EXECUTE .
```

```
VARIABLE LABELS pweightc "combined weight".
```

## PTENURE 'Tenure'

- 1 Owns outright
- 2 Buying with a mortgage
- 3 Private Rented
- 4 Social Rented
- 5 Missing

*Names of the original GHS variables used to create PTENURE:*  
OWNMORT, RENTFROM, TENURE and TENURE96

### Notes:

From 1972-1982 tenure is derived from two different variables:

- 1) OWNMORT – to determine whether the property is owned outright or with a mortgage
- 2) RENTFROM – to determine whether property is socially or privately rented.

These 2 variables were originally renamed in a previous recode to POWNMORT and PRENTFRO, while details are provided relating to these 2 variables in the SPSS recode syntax (and also in the Excel DataMap) they are not included on the final GHS Time Series Dataset. From 1983 onwards TENURE (and TENURE96) are used.

Caravan/House boats are assigned to 'Owns outright' rather than owned with a mortgage. The category 'Rents DK landlord/UN' has been assigned to 'Private rents'. The category 'NA TO MORTGAGE' (in 1983) have been added to 'Buying with a mortgage'. All 3 assumptions given above were based on preliminary analysis and investigations carried out by ONS.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS								New codes
	1972		1983	1985	1991	1996	2000	2001	
	OWNMORT	RENTFROM	1982	1984	1990	1995	1998	2004	PTENURE
Owns-DK Type			1	1	1	1			2
Owned outright	1		2	2	2	2	1	1	1
Buying on mortgage	2		3	3	3	3	2	2	2
Renting with job			4	4	4	4			3
Rent with busin			5	5	5	5			3
Rent LA			6	6	6	6	3	3	4
Rent HA			7	7			4	4	4
Rents pte org HA					7	7			4
Rents pte org UNF			8	8	8	8			3
Rents PTE IND UNF			9	9	9	9			3
PRIV Furnished			10	10	10	10			3
Rents DK Landlord/UN			11	11	11	11			3
Caravan/House boat			12						1
Caravan			-	12	12	12			1
Rents pte furn							6	5	3
Rents pte unfurn NK							5	6	3
Squats rents DK llrd							7	7	3
Local Authority		0							4
New town corporation		1							4
Property company		2							3
Housing association		3							4
Scottish SHA		4							4
Employer-company		5							3

Other organisation		6							3
Relative		7							3
Employer- individual		8							3
Other individual		9							3
Missing values	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	All other Missings	-5

*SPSS syntax to recode the variable PTENURE*

\*\*For 1972-82 ptenure needs to be derived from the variables rentfro ownmort\*\*.

\*de-select missing values\*.

RECODE prentfro pownmort ptenure (SYSMIS = -5) (ELSE = COPY).

EXE.

MISSING VALUES prentfro pownmort ptenure ().

EXE.

DO IF (year >= 1972 and year <= 1982).

```
+ DO IF (pownmort = 1).
+   COMPUTE ptenure = 1.
+ ELSE IF (pownmort = 2).
+   COMPUTE ptenure = 2.
+ ELSE IF (prentfro = 2).
+   COMPUTE ptenure = 3.
+ ELSE IF (prentfro >= 5).
+   COMPUTE ptenure = 3.
+ ELSE IF (prentfro = 0).
+   COMPUTE ptenure = 4.
+ ELSE IF (prentfro = 1).
+   COMPUTE ptenure = 4.
+ ELSE IF (prentfro = 3).
+   COMPUTE ptenure = 4.
+ ELSE IF (prentfro = 4).
+   COMPUTE ptenure = 4.
+ ELSE.
+   COMPUTE ptenure = -5.
+ END IF.
```

END IF.

EXE.

DO IF (year >= 1983 and year <= 1998).

```
+ RECODE
+   ptenure
+     (1=2) (2=1) (3=2) (4=3) (5=3) (6=4) (7=4) (8=3) (9=3) (10=3)
+     (11=3)(12=1)(ELSE=-5).
+ ELSE IF (year >= 2000).
+   RECODE
+     ptenure
+       (1=1) (2=2) (3=4) (4=Copy) (5 thru 7 =3) (ELSE=-5).
```

END IF.

VARIABLE LABELS ptenure 'Tenure'.

VALUE LABELS ptenure

```
1 'Owns outright'
2 'Buying with a mortgage'
3 'Private Rented'
4 'Social Rented'
-5 'Missing' .
```

EXECUTE.

**PCOB1 'Country of birth'**

**PFCOB1 'Father's Country of birth'**

**PMCOB1 'Mother's Country of birth'**

- 0 UK
- 1 England
- 2 Scotland
- 3 Wales
- 4 N Ireland
- 5 GB other
- 6 Eire
- 7 Europe
- 9 Old Commonwealth
- 10 India
- 11 East Africa, new Com
- 12 Bangladesh
- 13 Caribbean Com
- 14 Mediterranean Com
- 15 Remainder new Com
- 16 Pakistan
- 17 Rest of world
- 5 Missing

**Notes:**

The 3 variables relating to country of birth have been recoded for the time series GHS dataset following the work done on deriving variables to be consistent over time by ESDS Government for further details see the [ESDS Government website](#)<sup>xiii</sup>.

From 1972-78, there are a high number of missing values. This is because children do not have country of birth recorded in these years. In 1972 all children aged under 15 are excluded, for the other years all children under 16 do not have country of birth recorded. From 1979, country of birth is recorded for all persons.

PCOB1 'Country of birth', PFCOB1 'Father's country of birth' and PMCOB1 'Mother's country of birth' are coded the same for each year. Therefore the recode syntax is identical for each variable.

The GHS Time Series Dataset Excel DataMap file includes tables of the old and new codes for the dataset, however, due to the large number of categories in the years 1972-77 it was not deemed practicable to include many large tables in this user guide. Information for old and new codes are given in the following tables for PCOB1 'Country of birth', however they are the same for PFCOB1 'Father's country of birth' and PMCOB1 'Mother's country of birth.'

1972-77	WHBORN	New codes (as outlined by DV work by ESDS)	78-80	WHBORN	New codes (as outlined by DV work by ESDS)
0	OUTSIDE UK-DK WHERE	-5	1	ENGLAND	1
1	England	1	2	SCOTLAND	2
2	Scotland	2	3	WALES	3
3	Wales	3	4	NORTHERN IRELAND	4
4	Northern Ireland	4	5	UK,GB OR BRIT.ISLES	5
5	Australia	9	6	IRISH REPUBLIC	6

6	Australian dependency	9	7	OTHER EUROPE-EEC	7
7	New Zealand	9	8	GREECE,SPN,PORTGAL	7
8	NewZealand dependency	9	9	OTH.EUROPE-NON EEC	7
9	Canada	9	10	USSR	7
10	India	10	11	OLD COMMONWEALTH	9
11	Indian dependency	11	12	INDIA-NEW C'WLTH	10
12	Pakistan	16	13	BANGLADESH-N.C'WLTH	12
13	Ceylon	15	14	E.AFRICAN-NEW C'WLTH	11
14	Brunei	15	15	WEST INDIES-N.C'WLTH	13
15	Hong Kong	15	16	MEDITERRANEAN C'WLTH	14
16	Malaysia	15	17	REST OF NEW C'WLTH	15
17	Singapore	15	18	PAKISTAN	16
18	Islands: Fiji etc	15	19	RHODESIA	15
19	the Gambia	15	20	REST OF WORLD	17
20	Ghana	15	21	OUTSIDE UK-DK WHERE	-5
21	Nigeria	15	22	M NA,REFUSAL,DK	-5
22	Sierra Leone	15			
23	Kenya	11			
24	Malawi	11			
25	Tanzania	11			
26	Uganda	11			
27	Zambia	11			
28	Botswana	15			
29	Lesotho	15			
30	Rhodesia	15			
31	Swaziland	15			
33	Islands: Ascension etc	15			
34	Cyprus	14			
35	Gibraltar	14			
36	Malta	14			
37	Barbados	13			
38	British Honduras	13			
39	Guyana	13			
40	Jamaica	13			
41	Trinidad&Tobago	13			
42	Antigua,Dominica,etc	13			
43	W.Indies: other	13			
44	Falkland Islands	15			
45	Austria	7			
46	Belgium	7			
47	Denmark	7			
48	Finland	7			
49	France	7			
50	West Germany	7			
51	Greece	7			
52	Iceland	7			
53	Irish Republic	6			
54	Italy	7			
55	Netherlands	7			
56	Norway	7			
57	Portugal	7			
58	Spain	7			

81-82	WHBORN	New codes (as outlined by DV work by ESDS)
1	ENGLAND	1
2	SCOTLAND	2
3	WALES	3
4	N.IRELAND	4
5	GB,UK,BRITISH ISLES	5
6	IRISH REPUBLIC	6
7	EEC COUNTRIES	7
8	OTHER EUROPE	7
9	OLD COMMONWEALTH	9
10	IND,BANGLA,SRI LANKA	10
11	EAST.AFRICA N.C'WLTH	11
12	OTHR.AFRICA N.C'WLTH	15
13	CARIBBEAN N.C'WLTH	13
14	MEDITERRAN. N.C'WLTH	14
15	ASIAN&OTHER N.C'WLTH	15
16	PAKISTAN	16
17	REST OF WORLD	17
18	OUTSIDE UK-DK WHERE	-5
19	M NA, REFUSAL, DK	-5
20	DOES NOT APPLY & MD	-5

1983	COB	New codes (as outlined by DV work by ESDS)
1	ENGLAND	1
2	SCOTLAND	2
3	WALES	3
4	N.IRELAND	4
5	GB,UK,BRITISH ISLES	5
6	EIRE	6

59	Sweden	7	7	EEC COUNTRIES	7
60	Switzerland	7	8	OTHER EUROPE	7
61	Turkey	7	9	OLD COMMONWEALTH	9
62	Yugoslavia	7	10	INDIA, BANGLADESH, SRI LANKA	10
63	other W.Europe	7	11	EAST AFRICA - NEW C'WLTH	11
64	Albania	7	12	REST OF AFRICA - NEW C'WLTH	15
65	Bulgaria	7	13	CARIBBEAN C'WLTH	13
66	Czechoslovakia	7	14	MEDITERRANEAN C'WLTH	14
67	East Germany	7	15	REMAINDER NEW C'WLTH	15
68	Hungary	7	16	PAKISTAN	16
69	Poland	7	17	AFRICA	17
70	Romania	7	18	AMERICA	17
71	USSR	7	19	ASIA - MIDDLE EAST	17
72	French Somaliland	17	20	REST OF ASIA & OCEANIA	17
73	Algeria	17	21	OUTSIDE UK OR EIRE	-5
74	Angola	17			
75	Burundi	17			
76	Cameroon republic	17			
77	Central African Rep	17			
78	Chad	17			
79	Congo	17			
80	Zaire	17			
81	Dahomey	17			
82	Equatorial Guinea	17			
83	Ethiopia	17			
84	Gabon	17			
85	Guinea, Rep of	17			
86	Portuguese Guinea	17			
87	Ivory Coast	17			
88	Liberia	17			
89	Madagascar	17			
90	Mali	17			
91	Mauritania	17			
92	Morocco	17			
93	Mozambique	17			
94	Niger	17			
95	Rwanda	17			
96	Spanish Sahara	17			
97	Senegal	17			
98	Somalia	17			
99	South Africa	17			
100	Sudan	17			
101	Togo	17			
102	Tunisia	17			
103	UAR Egypt	17			
104	Upper Volta	17			
105	other Africa	17			
106	Bahrain	17			
107	Iran	17			
108	Iraq	17			
109	Israel	17			
110	Jordan	17			

1984-1997	COB	New codes (as outlined by DV work by ESDS)
1	ENGLAND	1
2	SCOTLAND	2
3	WALES	3
4	NORTHERN IRELAND	4
5	GB,UK,OTHER	5
6	IRISH REPUBLIC	6
7	EEC COUNTRIES	7
8	OTHER EUROPE	7
9	OLD COMMONWEALTH	9
10	INDIA	10
11	E AFRICAN CMWLTH	11
12	REST AFRICA CW	15
13	CARRIB CMWLTH	13
14	MED CMWLTH	14
15	FAR EAST CMWLTH	15
16	REMAINDER NEW	15
17	PAKISTAN	16
18	BANGLADESH	12
19	REST -AFRICA	17
20	REST- AMERICA	17
21	REST-MIDDLE EAST	17
22	REST-ASIA&OCEANI	17
23	UN SPEC&NOT DK	-5
24	OTHER (1994 ONWARDS)	17

1998-2004	COB1	New codes (as outlined by DV work by ESDS)
1	UNITED KINGDOM	1
5	CHANNEL IS, IoM	5
6	EIRE	6

111	Kuwait	17	7	EU EUROPE	7
112	Lebanon	17	8	OTHER EUROPE	7
113	Libya	17	9	OLD COMMONWLTH	9
114	Oman	17	10	INDIA	10
115	Qatar	17	11	E AFRICA NEW CW	11
116	Saudi Arabia	17	12	REST AF NEW CW	15
117	South Yemen	17	13	CARIB COMMWLTH	13
118	Syria	17	14	MEDIT COMMWLTH	14
119	UnitedArabEmirates	17	15	FAR EAST COMMWLTH	15
120	the Yemen	17	16	OTHER COMMWLTH	15
121	Afghanistan	17	17	PAKISTAN	16
122	Bhutan	17	18	BANGLADESH	12
123	Burma	17	19	REST - AFRICA	17
124	Cambodia=KhmerRepa	17	20	REST - AMERICA	17
125	China	17	21	REST - MID EAST	17
126	Taiwan	17	22	REST-ASIA&OCEAN	17
127	Indonesia	17	23	OTHER	17
128	Japan	17	-8	NA	-5
129	North Korea	17			
130	South Korea	17			
131	Laos	17			
132	Macau	17			
133	Maldives	17			
134	Mongolia	17			
135	Nepal	17			
136	Philippines	17			
137	Thailand	17			
138	North Vietnam	17			
139	South Vietnam	17			
140	Pacific islands	17			
141	USA	17			
142	Argentina	17			
143	Bolivia	17			
144	Brazil	17			
145	Chile	17			
146	Colombia	17			
147	Costa Rica	17			
148	Cuba	17			
149	Dominican Republic	17			
150	Ecuador	17			
151	Guatemala	17			
152	French Guiana	17			
153	Surinam	17			
154	Haiti	17			
155	Honduras	17			
156	Mexico	17			
157	Nicaragua	17			
158	Panama	17			
159	Paraguay	17			
160	Peru	17			
161	Puerto Rico	17			
162	El Salvador	17			

163	Uruguay	17	
164	Venezuela	17	
165	other America	17	
166	Germany: E or W	7	
167	Ireland: N or S	6	
168	Korea N or S	17	
169	Vietnam N or S	17	
170	UK unspecified	5	
171	Bangladesh	12	
172	W.Indies n.e.s	13	
999	MA\PA in this household		

*SPSS syntax to recode the 3 variables: PCOB1, PFCOB1 & PMCOB1*

```

*de-select missing values*.
MISSING VALUES pcob1 (.).
EXE.

DO IF (year<=1977).
+   RECODE
      pcob1
      (1 thru 4=Copy) (5 thru 9=9) (10 thru 11=Copy) (12=16) (171=12) (13
thru 22,28 thru 33,44=15) (23 thru 27 =11) (34,35,36=14)(37 thru 43,172=13) (170=5)(45 thru
52,54 thru 71,166=7) (53,167=6)(72 thru 165=17) (ELSE=-5).
END IF.
EXE.

DO IF (year>=1978 and year<=1980).
+   RECODE
      pcob1
      (1 thru 7=Copy)(8,9,10=7) (11=9) (12=10) (13=12) (14=11) (15=13)
(16=14)(17,19=15) (18=16)(20=17) (ELSE=-5).
END IF.
EXE.

DO IF (year>=1981 and year<=1982).
+   RECODE
      pcob1
      (1 thru 7=Copy)(8=7) (9 thru 11= Copy) (12=15) (13 thru
17=Copy)(ELSE=-5).
END IF.
EXE.

DO IF (year=1983).
+   RECODE
      pcob1
      (1 thru 7=Copy)(8=7) (9 thru 11= Copy) (12=15) (13 thru 16=Copy)(17
thru 21=17)(ELSE=-5).
END IF.
EXE.

DO IF (year>=1984 and year<=1997).
+   RECODE
      pcob1
      (1 thru 7=Copy)(8=7) (9 thru 11= Copy) (12,16=15) (13 thru
15=Copy)(18=12)(17=16)(19 thru 22,24=17) (ELSE=-5).
END IF.
EXE.

DO IF (year>=1998 and year<=2004).
+   RECODE
      pcob1
      (1=0)(2 thru 7=Copy)(8=7) (9 thru 11=Copy)(12,16=15) (13 thru
15=Copy) (18=12) (17=16) (19 thru 23=17)(ELSE=-5).
VARIABLE LABELS pcob1 'Country of birth'.
VALUE LABELS pcob1
      0 'UK'

```

```

1 'England'
2 'Scotland'
3 'Wales'
4 'N Ireland'
5 'GB other'
6 'Eire'
7 'Europe'
9 'Old Commonwealth'
10 'India'
11 'East Africa,new Com'
12 'Bangladesh'
13 'Caribbean Com'
14 'Mediterranean Com'
15 'Remainder new Com'
16 'Pakistan'
17 'Rest of world'
-5 'Missing'.
END IF.
EXE.

*de-select missing values*.
MISSING VALUES pfcob1 (.).
EXE.

DO IF (year<=1977).
+ RECODE
    pfcob1
        (1 thru 4=Copy) (5 thru 9=9) (10 thru 11=Copy) (12=16) (171=12) (13
thru 22,28 thru 33,44=15) (23 thru 27 =11) (34,35,36=14)(37 thru 43,172=13) (170=5)(45 thru
52,54 thru 71,166=7) (53,167=6)(72 thru 165=17) (ELSE=-5).
END IF.
EXE.

DO IF (year>=1978 and year<=1980).
+ RECODE
    pfcob1
        (1 thru 7=Copy)(8,9,10=7) (11=9) (12=10) (13=12) (14=11) (15=13)
(16=14)(17,19=15) (18=16)(20=17) (ELSE=-5).
END IF.
EXE.

DO IF (year>=1981 and year<=1982).
+ RECODE
    pfcob1
        (1 thru 7=Copy)(8=7) (9 thru 11= Copy) (12=15) (13 thru
17=Copy)(ELSE=-5).
END IF.
EXE.

DO IF (year=1983).
+ RECODE
    pfcob1
        (1 thru 7=Copy)(8=7) (9 thru 11= Copy) (12=15) (13 thru 16=Copy)(17
thru 21=17)(ELSE=-5).
END IF.
EXE.

DO IF (year>=1984 and year<=1997).
+ RECODE
    pfcob1
        (1 thru 7=Copy)(8=7) (9 thru 11= Copy) (12,16=15) (13 thru
15=Copy)(18=12)(17=16)(19 thru 22,24=17) (ELSE=-5).
END IF.
EXE.

DO IF (year>=1998 and year<=2004).
+ RECODE
    pfcob1
        (1=0)(2 thru 7=Copy)(8=7) (9 thru 11=Copy)(12,16=15) (13 thru

```

```

15=Copy) (18=12) (17=16) (19 thru 23=17)(ELSE=-5).
VARIABLE LABELS pfcob1 'Fathers country of birth'.
VALUE LABELS pfcob1
    0 'UK'
    1 'England'
    2 'Scotland'
    3 'Wales'
    4 'N Ireland'
    5 'GB other'
    6 'Eire'
    7 'Europe'
    9 'Old Commonwealth'
    10 'India'
    11 'East Africa,new Com'
    12 'Bangladesh'
    13 'Caribbean Com'
    14 'Mediterranean Com'
    15 'Remainder new Com'
    16 'Pakistan'
    17 'Rest of world'
    -5 'Missing'.
END IF.
EXE.

*de-select missing values*.
MISSING VALUES pfcob1 (.).
EXE.

DO IF (year<=1977).
+   RECODE
        pfcob1
            (1 thru 4=Copy) (5 thru 9=9) (10 thru 11=Copy) (12=16) (17=12) (13
thru 22,28 thru 33,44=15) (23 thru 27 =11) (34,35,36=14)(37 thru 43,172=13) (170=5)(45 thru
52,54 thru 71,166=7) (53,167=6)(72 thru 165=17) (ELSE=-5).
END IF.
EXE.

DO IF (year>=1978 and year<=1980).
+   RECODE
        pfcob1
            (1 thru 7=Copy)(8,9,10=7) (11=9) (12=10) (13=12) (14=11) (15=13)
(16=14)(17,19=15) (18=16)(20=17) (ELSE=-5).
END IF.
EXE.

DO IF (year>=1981 and year<=1982).
+   RECODE
        pfcob1
            (1 thru 7=Copy)(8=7) (9 thru 11= Copy) (12=15) (13 thru
17=Copy)(ELSE=-5).
END IF.
EXE.

DO IF (year=1983).
+   RECODE
        pfcob1
            (1 thru 7=Copy)(8=7) (9 thru 11= Copy) (12=15) (13 thru 16=Copy)(17
thru 21=17)(ELSE=-5).
END IF.
EXE.

DO IF (year>=1984 and year<=1997).
+   RECODE
        pfcob1
            (1 thru 7=Copy)(8=7) (9 thru 11= Copy) (12,16=15) (13 thru
15=Copy)(18=12)(17=16)(19 thru 22,24=17) (ELSE=-5).
END IF.
EXE.

```

```

DO IF (year>=1998 and year<=2004).
+   RECODE
      pmcob1
      (1=0)(2 thru 7=Copy)(8=7) (9 thru 11=Copy)(12,16=15) (13 thru
15=Copy) (18=12) (17=16) (19 thru 23=17)(ELSE=-5).
VARIABLE LABELS pmcob1 'Mothers country of birth'.
VALUE LABELS pmcob1
  0 'UK'
  1 'England'
  2 'Scotland'
  3 'Wales'
  4 'N Ireland'
  5 'GB other'
  6 'Eire'
  7 'Europe'
  9 'Old Commonwealth'
  10 'India'
  11 'East Africa,new Com'
  12 'Bangladesh'
  13 'Caribbean Com'
  14 'Mediterranean Com'
  15 'Remainder new Com'
  16 'Pakistan'
  17 'Rest of world'
  -5 'Missing'.
END IF.
EXE.

```

### PSEGH 'Socio-economic group of HOH/HRP'

- 1 EMPLOYERS: LARGE
- 2 MANAGERS: LARGE
- 3 EMPLOYERS:SMALL
- 4 MANAGERS:SMALL
- 5 PROF-SELF EMP
- 6 PROF-EMPLOYEE
- 7 INTERMED NON-MAN ANC
- 8 INTERMED NON-MAN FOREMN
- 9 JUNIOR NON-MAN
- 10 PERSONAL SERVICE
- 11 MANUAL: FOREMN/SV
- 12 SKILLED MANUAL
- 13 SEMI-SKILLED MANUAL
- 14 UNSKILLED MANUAL
- 15 OWN ACC NON-PROF
- 16 FARMERS: EMP&MGRS
- 17 FARMERS: OWN ACC
- 18 AGRIC WORKERS
- 19 ARMED FORCES
- 20 FULL TIME STUDENT
- 21 NEVER WORKED
- 22 NO USUAL OCCUPATION
- 5 Missing

*Names of the original GHS variables used to create PSEGH:*

HOHSEG, SEGHOH, SEGGHOH, HOHSEGE, HOHSEG, HRPSEG and HRPNWSEG

#### Notes:

In common with other government surveys, in 2000, the GHS replaced the Head of Household (HOH) with the Household Reference Person (HRP) for this purpose. [Appendix A of the 2000 Living in Britain](#)<sup>xiv</sup> explains the change. This suggests that in 14% of households the HRP will be different to the HOH.

From April 2001, NS-SEC ([National Statistics Socio-economic Classification](#))<sup>xv</sup> replaced Socio-economic Group. The operational categories of NS-SEC can be aggregated to produce an approximated version of the previous Socio-economic Group. These approximations have been shown to achieve an overall continuity level of 87%. Some tables on smoking have used this approximation (due to DH having a target based on SEG). Thus there is a variable from 2001 called HRPNWSEG which gives the approximated version of SEG for the HRP. As the approximation does not give a 100% continuity level it is not directly comparable with HRPSEG.

It would seem that in the 1972 GHS the questions used to create this variable were not asked of retired people in quarter 1+2. (One of the category codes in HOHSEG in 1972 "is not asked of retired in qtr 1+2", although no-one has actually been given this code – however there are a larger number of system missing in this year). Therefore a researcher may wish to exclude this year when carrying out analysis, or restrict analysis to exclude all people of retirement age.

There is no data for 1983. Although a variable called SEGHOH exists on the 1983 GHS dataset all values are missing (coded as missing values for this year).

In addition, the following assumptions have been made:

In 2000 GHS there are 2 variables: HRPSEG and HRPNWSEG. The variable HRPSEG was used for as it was assumed that the HRPNWSEG was created as part of the comparability testing process.

The label "EMPLOYERS:LARGE" used up to 1982 is the same as "EMPLOYERS- 25 PLUS" used from 1984 onwards.

The label "EMPLOYERS:SMALL" used up to 1982 is the same as "EMPLOYERS- UNDER 25" used from 1984 onwards.

The label "ANCILLARY,ARTISTS" used up to 1982 is the same as "INTERMED NON-MAN ANC" used from 1984 onwards.

Missing categories change over time. FULL TIME STUDENT and NEVER WORKED are separately identified in some years but not others. It has been decided to carry over information for these categories for the years in which they exist. With the exception of 2000, FULL TIME STUDENT is available for the years 1984 onwards and NEVER WORKED is available for the years 1985 onwards.

Value labels/names for the codes on the original GHS dataset	codes from each year of the GHS											New codes <b>PSEGH</b>
	1972	1983	1984	1985	1989	1990	1991	1992	1998	2000	2001	
	1982 HOHSEG	SEGHOH	SEGHOH	1988 SEGGHOH	HOHSEGE	HOHSEGE	HOHSEGE	1996 HOHSEGE	HOHSEG	HRPSEG	2004 HRPNWSEG	
EMPLOYERS:LARGE	1		1	1	1	1	1	1	1	1.1	1	1
MANAGERS :LARGE	2		2	2	2	2	2	2	2	1.2	2	2
EMPLOYERS:SMALL	3		3	3	3	3	3	3	3	2.1	3	3
MANAGERS :SMALL	4		4	4	4	4	4	4	4	2.2	4	4
PROF-SELF EMP	5		5	5	5	5	5	5	5	3	5	5
PROF-EMPLOYEE	6		6	6	6	6	6	6	6	4	6	6
ANCILLARY,ARTISTS (label in 1972-1982) INTERMED NON-MAN ANC (label from 1984)	7		7	7	7	7	7	7	7	5.1	7	7
INTERMED NON-MAN FOREMN	8		8	8	8	8	8	8	8	5.2	8	8
JUNIOR NON-MAN	9		9	9	9	9	9	9	9	6	9	9
PERSONAL SERVICE	10		10	10	10	10	10	10	10	7	10	10
MANUAL:FOREMN/SV	11		11	11	11	11	11	11	11	8	11	11
SKILL MAN	12		12	12	12	12	12	12	12	9	12	12
SEMI-SKILL MANPS	13		13	13	13	13	13	13	13	10	13	13
UNSKILL MANUAL	14		14	14	14	14	14	14	14	11	14	14
OWN ACC NON-PROF	15		15	15	15	15	15	15	15	12	15	15
FARMERS:EMP&MGRS	16		16	16	16	16	16	16	16	13	16	16
FAMERS:OWN ACC	17		17	17	17	17	17	17	17	14	17	17
AGRIC WORKERS	18		18	18	18	18	18	18	18	15	18	18
ARMED FORCES	19		19	19	19	19	19	19	19	16	19	19
FULL TIME STUDENT			20	20	20	20	20	20	20		20	20
NEVER WORKED				21	21	21	21	21	21		21	21
NO USUAL OCCUPATION	0											22
Missing values	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	-5

*SPSS syntax to recode the variable PSEGH*

\*de-select missing values\*.  
MISSING VALUES psegh ().  
EXE.

```

DO IF (year=2000).
+   RECODE psegh (1.1 = 1)(1.2 = 2)(2.1 = 3)(2.2 = 4)(3 = 5)(4 = 6)(5.1 = 7)(5.2 = 8)(6 =
9)(7 = 10)(8 = 11)(9 = 12)(10 = 13)(11 = 14)(12 = 15)(13 = 16)(14 = 17)(15 = 18)(16 = 19)
(ELSE = -5).
ELSE IF (year>=2001).
+   RECODE psegh (1 THRU 21=COPY)(ELSE = -5).
ELSE IF (year<=1998).
+   RECODE psegh (1 THRU 21=COPY)(0=22)(ELSE = -5).
END IF.

VARIABLE LABELS psegh 'Socio-economic group of HOH/HRP'.
VALUE LABELS
    psegh 1 'EMPLOYERS:LARGE' 2 'MANAGERS :LARGE' 3 'EMPLOYERS:SMALL' 4 'MANAGERS
:SMALL' 5 'PROF-SELF EMP' 6 'PROF-EMPLOYEE' 7 'INTERMED NON-MAN ANC'
    8 'INTERMED NON-MAN FOREMN' 9 'JUNIOR NON-MAN' 10 'PERSONAL SERVICE' 11
'MANUAL:FOREMN/SV' 12 'SKILLED MANUAL' 13 'SEMI-SKILLED MANUAL'
    14 'UNSKILLED MANUAL' 15 'OWN ACC NON-PROF' 16
'FARMERS:EMP&MGRS' 17 'FARMERS:OWN ACC' 18 'AGRIC WORKERS' 19 'ARMED FORCES'
20 'FULL TIME STUDENT'
    21 'NEVER WORKED' 22 'NO USUAL OCCUPATION' -5 'Missing'.
EXECUTE.

```

### PHRSEG3 'Socio-economic group of HOH/HRP - 7 categories'

- 1 Professional
- 2 Employer-manager
- 3 Intermed non-man
- 4 Junior non-man
- 5 Skilled man
- 6 Smi-skill manps
- 7 Unskill manual
- 9 Armed forces
- 8 FT student, Never worked, No usual occupation
- 5 Missing

#### Notes

This variable is a re-code of the variable PSEGH. For a detailed description of the original values on the GHS data for each year please see the metadata information for PSEGH.

Value labels/names for the codes on for the original variable	codes from each year	
	1972	New codes
	2004	<b>PHRSEG3</b>
	PSEGH	
EMPLOYERS: LARGE	1	2
MANAGERS: LARGE	2	2
EMPLOYERS:SMALL	3	2
MANAGERS:SMALL	4	2
PROF-SELF EMP	5	1
PROF-EMPLOYEE	6	1
INTERMED NON-MAN ANC	7	3
INTERMED NON-MAN FOREMN	8	3
JUNIOR NON-MAN	9	4
PERSONAL SERVICE	10	6
MANUAL: FOREMN/SV	11	5
SKILLED MANUAL	12	5
SEMI-SKILLED MANUAL	13	6
UNSKILLED MANUAL	14	7
OWN ACC NON-PROF	15	5
FARMERS: EMP&MGRS	16	2
FARMERS: OWN ACC	17	5
AGRIC WORKERS	18	6
ARMED FORCES	19	-9
FULL TIME STUDENT	20	-8
NEVER WORKED	21	-8
NO USUAL OCCUPATION	22	-8
Missing	-5	-5

#### SPSS syntax to recode the variable PHRSEG3

\*de-select missing values\*.  
MISSING VALUES psegH (.).  
EXE.

```
RECODE psegH
  ( 5,6 = 1 )
  ( 1,2,3,4,16 = 2 )
  ( 7,8 = 3 )
  ( 9 = 4 )
  ( 11,12,15,17 = 5 )
```

```
( 10,13,18 = 6 )
( 14 = 7 )
( 19 = -9 )
( 20,21,22 = -8)
(ELSE=-5) INTO phrpseg3.
EXECUTE.

VARIABLE LABELS phrpseg3 'Socio-economic group of HOH/HRP - 7 categories'.
VALUE LABELS phrpseg3
  1 'Professional'
  2 'Employer-manager'
  3 'Intermed non-man'
  4 'Junior non-man'
  5 'Skilled man'
  6 'Smi-skill manps'
  7 'Unskill manual'
  -9 'Armed forces'
  -8 'FT student, Never worked, No usual occupation'
  -5 'Missing'.
```

### PHPSEG3B 'Socio-economic group of HOH/HRP - 2 categories'

- 1 Non-manual
- 2 Manual
- 9 Armed forces
- 8 FT student, Never worked, No usual occupation
- 5 Missing

#### Notes:

This variable recodes the variable PHPSEG3, which was derived from PSEGH. For a detailed description of the original values on the GHS data for each year please see the metadata information for the variable PSEGH.

Value labels/names for the codes on for the original variable	codes from each year		New codes
	1972	2004	
	PHRPSEG3		
Professional	1		1
Employer-manager	2		1
Intermed non-man	3		1
Junior non-man	4		1
Skilled man	5		2
Smi-skill manps	6		2
Unskill manual	7		2
Armed forces	-9		-9
FT student, Nev worked, No usual occ	-8		-8
Missing	-5		-5

#### SPSS syntax to recode the variable PHPSEG3B

```
*de-select missing values*.
MISSING VALUES phrpseg3 (.).
EXE.
```

```
RECODE phrpseg3
(1 thru 4=1)(5 thru 7=2)(ELSE=copy) INTO phpseg3b.
EXECUTE.
```

```
VARIABLE LABELS phpseg3b 'Socio-economic group of HOH/HRP - 2 categories'.
VALUE LABELS phpseg3b
1 'Non-manual'
2 'Manual'
-9 'Armed forces'
-8 'FT student, Nev worked, No usual occ'
-5 'Missing'.
```

## PHRPSEC8 '8 classes of NS SEC of HRP'

NS-SEC is the [National Statistics Socio-economic Classification](#)<sup>xvi</sup>

- 1.1 Large employers and higher managerial occs
- 1.2 Higher professional occs
- 2 Lower Managerial and professional occs
- 3 Intermediate occupations
- 4 Small employers and own account workers
- 5 Lower supervisory and technical occupations
- 6 Semi-routine occupations
- 7 Routine occupations
- 8 Never worked and long term unemployed
- 5 Missing

Name of the original GHS variables used to create PHRPSEC8: HRPSEC8

### Notes:

This variable is only available from 2001 onwards

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS		New codes
	2001	2004	
	HRPSEC8		PHRPSEC8
Large employers and higher managerial occs	1.1		1.1
Higher professional occs	1.2		1.2
Lower Managerial and professional occs	2		2
Intermediate occupations	3		3
Small employers and own account workers	4		4
Lower supervisory and technical occupations	5		5
Semi-routine occupations	6		6
Routine occupations	7		7
Never worked and long term unemployed	8		8
Missing values	All other missings		-5

### SPSS syntax to recode the variable PHRPSEC8

```
*de-select missing values*.
MISSING VALUES phrpsec8 (.).
EXE.
```

```
RECODE phrpsec8 (0 thru HIGHEST=COPY) (ELSE=-5).
VARIABLE LABELS phrpsec8 '8 classes of NS SEC of HRP'.
VALUE LABELS phrpsec8
```

- 1.1 'Large employers and higher managerial occupations'
- 1.2 'Higher professional occ'
- 2 'Lower Managerial and professional occ'
- 3 'Intermediate occupations'
- 4 'Small employers and own account workers'
- 5 'Lower supervisory and technical occupations'
- 6 'Semi-routine occupations'
- 7 'Routine occupations'
- 8 'Never worked and long term unemployed'
- 5 'Missing'.

```
EXECUTE.
```

## PAGE 'Age'

99 99 or older  
-5 Missing

Name of the original GHS variables used to create PAGE: AGE

### Notes:

The code "99" represents those aged 99 or older. This code was already used in the GHS data for 1972-82. For consistency this category (Aged 99+) has been created for survey years 1983 onwards, where actual ages were recorded for those aged 100 or older (except 2001/2).

In 2001 the oldest ages are categorised into 85+ in the original data. In other words the code "85" refers to those aged 85 or older.

In the original GHS data in 1972 there was a code -9 "Age unknown, but 15+". In 1973 there was a code -1 "But 16 or over", -9 "NA But under 16" These have all been recoded as missing for the GHS Time Series Dataset.

Value labels/names for the codes on the original GHS dataset	codes from each year of the GHS				New codes
	1972	1973 1982	1983 2004		
	AGE	AGE	AGE	PAGE	
NUMERIC	√	√	√	√	
BABY UNDER 1 YR. OLD	0	0		0	
99,100 OR MORE	99	99		99	
AGE UNKNOWN, BUT 15+	-9M			-5	
BUT 16 OR OVER		-1M		-5	
N.A. BUT UNDER 16		-9M		-5	

Numeric with the exception of 2001/2 (see notes).

No value labels found.

### SPSS syntax to recode the variable PAGE

```
*de-select missing values*.  
MISSING VALUES page ().  
EXE.
```

```
RECODE page (0 thru 98=Copy) (99 thru Highest = 99) (ELSE=-5).  
EXECUTE.
```

```
VARIABLE LABELS page 'AGE'.  
VALUE LABELS page 99 '99 or older' -5 'Missing'.
```

## **PSEX 'Sex'**

- 1 Male
- 2 Female
- 5 Missing

*Name of the original GHS variables used to create PSEX: SEX*

Value labels/names for the codes on the original GHS dataset	codes from each year of the GHS		New codes
	1972	2004	
	SEX		<b>PSEX</b>
MALE			<b>1</b>
FEMALE			<b>2</b>
Missing values	All other missings		<b>-5</b>

### SPSS syntax to recode the variable PSEX

```
*de-select missing values*.
MISSING VALUES psex (.).
EXE.
```

```
RECODE
  psex
  (1=1) (2=2) (ELSE=-5) .
VARIABLE LABELS psex 'SEX'.
VALUE LABELS psex 1 'Male' 2 'Female' -5 'Missing'.
EXECUTE .
```

## **ESTYOB 'Estimated Year of Birth'**

-5 Missing

This variable was specifically derived for the GHS Time Series Dataset from the variable PAGE, for more information on the original variable and coding see the metadata notes on the variable PAGE.

From the original GHS variable AGE in 2001 the oldest ages are categorised into 85+. In other words for 2001 there anyone aged 85+ will be recoded as having an estimated year of birth as 1916.

*SPSS syntax to recode the variable ESTYOB*

```
DO IF page = -5.  
+   COMPUTE estyob = -5.  
ELSE.  
+   COMPUTE estyob = (year - page).  
END IF.  
EXE.  
  
VARIABLE LABELS estyob 'Estimated Year of Birth'.  
VALUE LABELS estyob  
-5 'Missing'  
EXECUTE.
```

## CHORT5YR '5-year estimated birth cohort'

This variable was specifically derived for the GHS Time Series Dataset from the variable ESTYOB, which was originally derived from the variable PAGE. For more details see the metadata information both variables.

From the original GHS variable AGE in 2001 the oldest ages are categorised into 85+. In other words for 2001 there anyone aged 85+ will be recoded as being part of the 1916-1920 estimated birth cohort.

*SPSS syntax to recode the variable CHORT5YR*

```
*de-select missing values*.
MISSING VALUES estyob (.).
EXE.
RECODE estyob
  (1871 thru 1875=1871)
  (1876 thru 1880=1876)
  (1881 thru 1885=1881)
  (1886 thru 1890=1886)
  (1891 thru 1895=1891)
  (1896 thru 1900=1896)
  (1901 thru 1905=1901)
  (1906 thru 1910=1906)
  (1911 thru 1915=1911)
  (1916 thru 1920=1916)
  (1921 thru 1925=1921)
  (1926 thru 1930=1926)
  (1931 thru 1935=1931)
  (1936 thru 1940=1936)
  (1941 thru 1945=1941)
  (1946 thru 1950=1946)
  (1951 thru 1955=1951)
  (1956 thru 1960=1956)
  (1961 thru 1965=1961)
  (1966 thru 1970=1966)
  (1971 thru 1975=1971)
  (1976 thru 1980=1976)
  (1981 thru 1985=1981)
  (1986 thru 1990=1986)
  (1991 thru 1995=1991)
  (1996 thru 2000=1996)
  (2001 thru 2005=2001)
  (ELSE=-5) INTO chort5yr.
EXECUTE.

VARIABLE LABELS chort5yr '5-year estimated birth cohort'.
VALUE LABELS chort5yr
  1871 '1871-1875'
  1876 '1876-1880'
  1881 '1881-1885'
  1886 '1886-1890'
  1891 '1891-1895'
  1896 '1896-1900'
  1901 '1901-1905'
  1906 '1906-1910'
  1911 '1911-1915'
  1916 '1916-1920'
  1921 '1921-1925'
  1926 '1926-1930'
  1931 '1931-1935'
  1936 '1936-1940'
  1941 '1941-1945'
  1946 '1946-1950'
  1951 '1951-1955'
  1956 '1956-1960'
  1961 '1961-1965'
  1966 '1966-1970'
  1971 '1971-1975'
  1976 '1976-1980'
  1981 '1981-1985'
```

1986 '1986-1990'  
1991 '1991-1995'  
1996 '1996-2000'  
2001 '2001-2005'  
-5 'Missing'.

MISSING VALUES chort5yr (-5).

## PDVMARDF 'Marital status'

- 1 Married
- 2 Cohabiting
- 3 Single
- 4 Widowed
- 5 Divorced
- 6 Separated
- 7 Same sex couple
- 5 Missing

*Names of the original GHS variables used to create PDVMARDF:*  
MARITAL, MARSTAT and DVMARDF

### Notes:

Separate identification of cohabiting persons from 1986 onwards. Prior to this cohabiting persons who described themselves as married were classified as such. Those who described themselves by their legal marital status would have been classified into those categories.

In cases where less than 16yrs is recorded, these are recoded into 'single'.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS							New codes
	1972 1978	1979	1980 1982	1983 1985	1986 1992	1993 1996	1998 2004	PDVMARDF
	MARITAL	MARITAL	MARITAL	MARSTAT	MARSTAT	MARSTAT	DVMARDF	
MARRIED	1	1	1	1	1	1	1	1
COHABITING					2	2	2	2
SINGLE-16 & OVER	2	2	2	2	3	3	3	3
WIDOWED	3	3	3	3	4	4	4	4
DIVORCED	4	4	4	4	5	5	5	5
SEPARATED	5	5	5	5	6	6	6	6
SAME SEX COUPLE						7	7	7
SINGLE-UNDER 16	6	0						3
Missing values	All other Missings	All other missings	All other missings	All other missings	All other missings	All other missings	All other Missings	-5

### SPSS syntax to recode the variable PDVMARDF

```
*de-select missing values*.
MISSING VALUES pdvmardf (.).
EXE.

DO IF (year<=1978).
+ RECODE pdvmardf (1=Copy) (2=3) (3=4) (4=5) (5=6) (6=3) (ELSE=-5).
ELSE IF (year=1979).
+ RECODE pdvmardf (0=3) (1=Copy) (2=3) (3=4) (4=5) (5=6) (ELSE=-5).
ELSE IF (year>=1980 and year<=1985).
+ RECODE pdvmardf (1=Copy) (2=3) (3=4) (4=5) (5=6) (ELSE=-5).
ELSE IF (year>=1986).
+ RECODE pdvmardf (1 thru 7=Copy) (ELSE=-5).
END IF.
EXE.

VARIABLE LABELS pdvmardf 'Marital Status'.
VALUE LABELS pdvmardf
  1 'Married'
  2 'Cohabiting'
  3 'Single'
  4 'Widowed'
  5 'Divorced'
  6 'Separated'
  7 'Same sex couple'
 -5 'Missing'.
EXECUTE
```

## **PHHTYPE 'Household type A (Grouped)'**

- 1 1 adult aged 16-59
- 2 2 adults aged 16-59
- 3 Youngest person aged 0-4
- 4 Youngest person aged 5-15
- 5 3 or more adults
- 6 1 adult aged 60+
- 7 2 adults, 1 or both aged 60+
- 5 Missing

*Names of the original GHS variables used to create PHHTYPE:*  
HHTYPE and HHTYPA

### **Notes:**

The GHS Time Series Dataset includes a grouped version of HHTYPA, which is consistent with the GHS variable HHTYPE (1974-1982), to facilitate analysing trends.

HHTYPA was introduced in 1983, as a more detailed version of HHTYPE. HHTYPA includes information about the number of adults living in a household where children are present.

Value labels/names for the codes on the original GHS dataset	codes from each year of the GHS		New codes
	1974	1983	
	1982	2004	<b>PHHTYPA</b>
	HHTYPE	HHTYPA	
1 adult under 60	1		1
2 adults under 60	2		2
child lt 5 in hhold	3		3
with children 5-15	4		4
3 or more adults	5		5
1 adult 60 or over	6		6
2 adults, 1 or 2 60+	7		7
1 adult 16-59		1	1
2 ads; both 16-59		2	2
1,2 ads, 1,2 ch 0-4		3	3
1,2 ads, 1,2 ch 5-15		4	4
1,2 ads, 3+ch 0-4		5	3
3+ ads, 2+ ch 0-4		6	3
1,2 ads, 3+ch 5-15		7	4
3+ ads, 2+ ch 5-15		8	4
3+ ads, 1 ch 0-4		9	3
3+ ads, 1 ch 5-15		10	4
3+ ads, no children		11	5
2 ads, 1 or both 60+		12	7
1 adult 60-99		13	6
Missing values	All other missings	All other missings	-5

### **SPSS syntax to recode the variable PHHTYPE**

```
*de-select missing values*.
MISSING VALUES phhtype (.).
EXE.
```

```
DO IF (year<=1982).
+ RECODE phhtype (1 thru 7= Copy) (ELSE=-5).
END IF.
EXE.
```

```
DO IF (year>=1983).
+ RECODE phhtype (1 thru 2 =Copy) (3,5,6,9=3) (4,7,8,10=4) (11=5) (13=6) (12=7)
```

```
(ELSE=-5).  
END IF.
```

```
VARIABLE LABELS phhtype 'Household Type A (Grouped)'.  
VALUE LABELS phhtype 1 '1 adult aged 16-59' 2 '2 adults aged 16-59' 3 'Youngest person aged  
0-4' 4 'Youngest person aged 5-15' 5 '3 or more adults' 6 '1 adult aged 60+' 7 '2 adults, 1 or  
both aged 60+' -5 'Missing'.  
EXECUTE .
```

### PHHTYPF 'Household type F'

1	1 PERS,PENSIONER	42	COHAB CPL, 2 DCH, M16-29
2	1 PERS, NOT PENS	43	COHAB CPL, 2 DCH, M30-44
3	2+ PERS,1+ IS PENS	44	COHAB CPL, 2 DCH, M45-64
4	2+ PERS,NO PENS	45	COHAB CPL, 2 DCH, M 65+
5	MARR CPL, 0 CH, W16-29	46	COHAB CPL, 3 DCH, M16-29
6	MARR CPL, 0 CH, W30-44	47	COHAB CPL, 3 DCH, M30-44
7	MARR CPL, 0 CH, W45-64	48	COHAB CPL, 3 DCH, M45-64
8	MARR CPL, 0 CH, W 65+	49	COHAB CPL, 3 DCH, M 65+
9	MARR CPL, 1 DCH, M16-29	50	COHAB CPL, 4+DCH, M16-29
10	MARR CPL, 1 DCH, M30-44	51	COHAB CPL, 4+DCH, M30-44
11	MARR CPL, 1 DCH, M45-64	52	COHAB CPL, 4+DCH, M45-64
12	MARR CPL, 1 DCH, M 65+	53	COHAB CPL, 4+DCH, M 65+
13	MARR CPL, 2 DCH, M16-29	54	COHAB CPL, NONDC, M16-29
14	MARR CPL, 2 DCH, M30-44	55	COHAB CPL, NONDC, M30-44
15	MARR CPL, 2 DCH, M45-64	56	COHAB CPL, NONDC, M45-64
16	MARR CPL, 2 DCH, M 65+	57	COHAB CPL, NONDC, M 65+
17	MARR CPL, 3 DCH, M16-29	58	CPL,0 CH,W16:29
18	MARR CPL, 3 DCH, M30-44	59	CPL,0 CH,W30:44
19	MARR CPL, 3 DCH, M45-64	60	CPL,0 CH,W45:64
20	MARR CPL, 3 DCH, M 65+	61	CPL,0 CH,W 65+
21	MARR CPL, 4+DCH, M16-29	62	CPL,1 DCH,M16:29
22	MARR CPL, 4+DCH, M30-44	63	CPL,1 DCH,M30:44
23	MARR CPL, 4+DCH, M45-64	64	CPL,1 DCH,M45:64
24	MARR CPL, 4+DCH, M 65+	65	CPL,1 DCH,M 65+
25	MARR CPL, NONDC, M16-29	66	CPL,2 DCH,M16:29
26	MARR CPL, NONDC, M30-44	67	CPL,2 DCH,M30:44
27	MARR CPL, NONDC, M45-64	68	CPL,2 DCH,M45:64
28	MARR CPL, NONDC, M 65+	69	CPL,2 DCH,M 65+
29	LONE P, 1+DEP CH	70	CPL,3 DCH,M16:29
30	LONE P, CH NONDEP	71	CPL,3 DCH,M30:44
31	2+ FAMILIES	72	CPL,3 DCH,M45:64
32	UNCLASSIFIABLE	73	CPL,3 DCH,M 65+
33	SAME SEX COHAB	74	CPL,4+DCH,M16:29
34	COHAB CPL, 0 CH, W16-29	75	CPL,4+DCH,M30:44
35	COHAB CPL, 0 CH, W30-44	76	CPL,4+DCH,M45:64
36	COHAB CPL, 0 CH, W45-64	77	CPL,4+DCH,M 65+
37	COHAB CPL, 0 CH, W 65+	78	CPL,NONDC,M16:29
38	COHAB CPL, 1 DCH, M16-29	79	CPL,NONDC,M30:44
39	COHAB CPL, 1 DCH, M30-44	80	CPL,NONDC,M45:64
40	COHAB CPL, 1 DCH, M45-64	81	CPL,NONDC,M 65+
41	COHAB CPL, 1 DCH, M 65+	-5	Missing

*Name of the original GHS variables used to create PHHTYPF: HHTYPF*

#### Notes:

This variable is only available from 1984 onwards.

In 1984 there are no data labels for the data in this year but it is assumed that it is consistent with 1985.

In 1993 same sex cohabiting was introduced on the GHS and were therefore included in the household type variable. In the past cohabiting had not been separately distinguished, and had been grouped with married couples (although the 'marital status' variable has included a

cohabiting category since 1986). In order to keep the detail between couple households (prior to 1993), as well as, separate married and cohabiting households, PHHTYPF has been recoded (to include 81 categories) so that each of the household types are preserved through time.

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				New codes  <b>PHHTYPF</b>
	1984	1993	1998		
		1992	1996	2003	
		HHTYPF	HHTYPF	HHTYPF	
1 PERS,PENSIONER		1		1	1
1 PERS, NOT PENS		2		2	2
2+ PERS,1+ IS PENS		3		3	3
2+ PERS,NO PENS		4		4	4
MARR CPL, 0 CH, W16-29				5	5
MARR CPL, 0 CH, W30-44				6	6
MARR CPL, 0 CH, W45-64				7	7
MARR CPL, 0 CH, W 65+				8	8
MARR CPL, 1 DCH, M16-29				9	9
MARR CPL, 1 DCH, M30-44				10	10
MARR CPL, 1 DCH, M45-64				11	11
MARR CPL, 1 DCH, M 65+				12	12
MARR CPL, 2 DCH, M16-29				13	13
MARR CPL, 2 DCH, M30-44				14	14
MARR CPL, 2 DCH, M45-64				15	15
MARR CPL, 2 DCH, M 65+				16	16
MARR CPL, 3 DCH, M16-29				17	17
MARR CPL, 3 DCH, M30-44				18	18
MARR CPL, 3 DCH, M45-64				19	19
MARR CPL, 3 DCH, M 65+				20	20
MARR CPL, 4+DCH, M16-29				21	21
MARR CPL, 4+DCH, M30-44				22	22
MARR CPL, 4+DCH, M45-64				23	23
MARR CPL, 4+DCH, M 65+				24	24
MARR CPL, NONDC, M16-29				25	25
MARR CPL, NONDC, M30-44				26	26
MARR CPL, NONDC, M45-64				27	27
MARR CPL, NONDC, M 65+				28	28
LONE P, 1+DEP CH	29	29		29	29
LONE P, CH NONDEP	30	30		30	30
2+ FAMILIES	31	31		31	31
UNCLASSIFIABLE	32	32		32	32
SAME SEX COHAB		33		33	33
COHAB CPL, 0 CH, W16-29				34	34
COHAB CPL, 0 CH, W30-44				35	35
COHAB CPL, 0 CH, W45-64				36	36
COHAB CPL, 0 CH, W 65+				37	37
COHAB CPL, 1 DCH, M16-29				38	38
COHAB CPL, 1 DCH, M30-44				39	39
COHAB CPL, 1 DCH, M45-64				40	40
COHAB CPL, 1 DCH, M 65+				41	41
COHAB CPL, 2 DCH, M16-29				42	42
COHAB CPL, 2 DCH, M30-44				43	43
COHAB CPL, 2 DCH, M45-64				44	44
COHAB CPL, 2 DCH, M 65+				45	45
COHAB CPL, 3 DCH, M16-29				46	46
COHAB CPL, 3 DCH, M30-44				47	47
COHAB CPL, 3 DCH, M45-64				48	48
COHAB CPL, 3 DCH, M 65+				49	49
COHAB CPL, 4+DCH, M16-29				50	50

COHAB CPL, 4+DCH, M30-44			51	51
COHAB CPL, 4+DCH, M45-64			52	52
COHAB CPL, 4+DCH, M 65+			53	53
COHAB CPL, NONDC, M16-29			54	54
COHAB CPL, NONDC, M30-44			55	55
COHAB CPL, NONDC, M45-64			56	56
COHAB CPL, NONDC, M 65+			57	57
CPL,0 CH,W16:29	5	5		58
CPL,0 CH,W30:44	6	6		59
CPL,0 CH,W45:64	7	7		60
CPL,0 CH,W 65+	8	8		61
CPL,1 DCH,M16:29	9	9		62
CPL,1 DCH,M30:44	10	10		63
CPL,1 DCH,M45:64	11	11		64
CPL,1 DCH,M 65+	12	12		65
CPL,2 DCH,M16:29	13	13		66
CPL,2 DCH,M30:44	14	14		67
CPL,2 DCH,M45:64	15	15		68
CPL,2 DCH,M 65+	16	16		69
CPL,3 DCH,M16:29	17	17		70
CPL,3 DCH,M30:44	18	18		71
CPL,3 DCH,M45:64	19	19		72
CPL,3 DCH,M 65+	20	20		73
CPL,4+DCH,M16:29	21	21		74
CPL,4+DCH,M30:44	22	22		75
CPL,4+DCH,M45:64	23	23		76
CPL,4+DCH,M 65+	24	24		77
CPL,NONDC,M16:29	25	25		78
CPL,NONDC,M30:44	26	26		79
CPL,NONDC,M45:64	27	27		80
CPL,NONDC,M 65+	28	28		81
Missing values	All other missings	All other missings	All other missings	-5

*SPSS syntax to recode the variable PHHTYPE*

```

*de-select missing values*.
MISSING VALUES phhtypf (.).
EXE.

DO IF (year>=1984 and year<=1996).
+   RECODE
      phhtypf
      (1 thru
4=Copy)(5=58)(6=59)(7=60)(8=61)(9=62)(10=63)(11=64)(12=65)(13=66)(14=67)(15=68)
      (16=69)(17=70)(18=71)(19=72)(20=73)(21=74)(22=75)(23=76)(24=77)(25=78)(26=79)(27=80)
      (28=81)(29 thru 33=Copy)(ELSE=-5).
ELSE IF (year>=1998).
+   RECODE
      phhtypf (1 thru 57=Copy) (ELSE=-5) .
END IF.

VARIABLE LABELS phhtypf 'Household type F'.
VALUE LABELS
  phhtypf
    1 '1 PERS,PENSIONER'
    2 '1 PERS, NOT PENS'
    3 '2+ PERS,1+ IS PENS'
    4 '2+ PERS,NO PENS'
    5 'MARR CPL, 0 CH, W16-29'
    6 'MARR CPL, 0 CH, W30-44'
    7 'MARR CPL, 0 CH, W45-64'
    8 'MARR CPL, 0 CH, W 65+'
    9 'MARR CPL, 1 DCH, M16-29'
    10 'MARR CPL, 1 DCH, M30-44'

```

11 'MARR CPL, 1 DCH, M45-64'  
 12 'MARR CPL, 1 DCH, M 65+'  
 13 'MARR CPL, 2 DCH, M16-29'  
 14 'MARR CPL, 2 DCH, M30-44'  
 15 'MARR CPL, 2 DCH, M45-64'  
 16 'MARR CPL, 2 DCH, M 65+'  
 17 'MARR CPL, 3 DCH, M16-29'  
 18 'MARR CPL, 3 DCH, M30-44'  
 19 'MARR CPL, 3 DCH, M45-64'  
 20 'MARR CPL, 3 DCH, M 65+'  
 21 'MARR CPL, 4+DCH, M16-29'  
 22 'MARR CPL, 4+DCH, M30-44'  
 23 'MARR CPL, 4+DCH, M45-64'  
 24 'MARR CPL, 4+DCH, M 65+'  
 25 'MARR CPL, NONDC, M16-29'  
 26 'MARR CPL, NONDC, M30-44'  
 27 'MARR CPL, NONDC, M45-64'  
 28 'MARR CPL, NONDC, M 65+'  
 29 'LONE P, 1+DEP CH'  
 30 'LONE P, CH NONDEP'  
 31 '2+ FAMILIES'  
 32 'UNCLASSIFIABLE'  
 33 'SAME SEX COHAB'  
 34 'COHAB CPL, 0 CH, W16-29'  
 35 'COHAB CPL, 0 CH, W30-44'  
 36 'COHAB CPL, 0 CH, W45-64'  
 37 'COHAB CPL, 0 CH, W 65+'  
 38 'COHAB CPL, 1 DCH, M16-29'  
 39 'COHAB CPL, 1 DCH, M30-44'  
 40 'COHAB CPL, 1 DCH, M45-64'  
 41 'COHAB CPL, 1 DCH, M 65+'  
 42 'COHAB CPL, 2 DCH, M16-29'  
 43 'COHAB CPL, 2 DCH, M30-44'  
 44 'COHAB CPL, 2 DCH, M45-64'  
 45 'COHAB CPL, 2 DCH, M 65+'  
 46 'COHAB CPL, 3 DCH, M16-29'  
 47 'COHAB CPL, 3 DCH, M30-44'  
 48 'COHAB CPL, 3 DCH, M45-64'  
 49 'COHAB CPL, 3 DCH, M 65+'  
 50 'COHAB CPL, 4+DCH, M16-29'  
 51 'COHAB CPL, 4+DCH, M30-44'  
 52 'COHAB CPL, 4+DCH, M45-64'  
 53 'COHAB CPL, 4+DCH, M 65+'  
 54 'COHAB CPL, NONDC, M16-29'  
 55 'COHAB CPL, NONDC, M30-44'  
 56 'COHAB CPL, NONDC, M45-64'  
 57 'COHAB CPL, NONDC, M 65+'  
 58 'CPL,0 CH,W16:29'  
 59 'CPL,0 CH,W30:44'  
 60 'CPL,0 CH,W45:64'  
 61 'CPL,0 CH,W 65+'  
 62 'CPL,1 DCH,M16:29'  
 63 'CPL,1 DCH,M30:44'  
 64 'CPL,1 DCH,M45:64'  
 65 'CPL,1 DCH,M 65+'  
 66 'CPL,2 DCH,M16:29'  
 67 'CPL,2 DCH,M30:44'  
 68 'CPL,2 DCH,M45:64'  
 69 'CPL,2 DCH,M 65+'  
 70 'CPL,3 DCH,M16:29'  
 71 'CPL,3 DCH,M30:44'  
 72 'CPL,3 DCH,M45:64'  
 73 'CPL,3 DCH,M 65+'  
 74 'CPL,4+DCH,M16:29'  
 75 'CPL,4+DCH,M30:44'  
 76 'CPL,4+DCH,M45:64'  
 77 'CPL,4+DCH,M 65+'  
 78 'CPL,NONDC,M16:29'  
 79 'CPL,NONDC,M30:44'  
 80 'CPL,NONDC,M45:64'  
 81 'CPL,NONDC,M 65+'

```
EXECUTE .
```

```
-5 'Missing'.
```

## **PHHTYF1 'Household type F (Grouped)'**

- 1 1 PERSON ONLY
- 2 2+ UNRELATED ADULTS
- 3 MARRIED COUPLE, DEP CHILDREN
- 4 MARRIED COUPLE, INDEP CHILDREN
- 5 MARRIED COUPLE, NO CHILDREN
- 6 LONE PARENT, DEP CHLDREN
- 7 LONE PARENT, INDEP CHILDREN
- 8 2+ FAMILIES
- 9 SAME SEX COHAB
- 10 COHAB COUPLE, DEP CHILDREN
- 11 COHAB COUPLE, INDEP CHILDREN
- 12 COHAB COUPLE, NO CHILDREN
- 13 COUPLE HOUSEHOLD, DEP CHILDREN
- 14 COUPLE HOUSEHOLD, INDEP CHILDREN
- 15 COUPLE HOUSEHOLD, NO CHILDREN
- 5 Missing

### Notes:

This variable is only available from 1984 onwards.

PHHTYF1 is available for all years from 1984, and has been grouped to reflect the information presented in Table 3.5 in *Living in Britain 2000*. The GHS Time Series Dataset version of PHHTYF1 will differ slightly to normal GHS variable, as it contains 15 categories rather than 12.

This variable was derived from the PHHTYPF, for a detailed description of the original values on the GHS data for each year (please see the metadata information for the variable PHHTYPF).

Rather than producing a large table of all 81 original categories relate to the 15 categories a summarized table with relation to the original variable HHTYPF which was used to create PHHTYPF is given below (the full table can be found in the GHS Time Series Excel DataMap)

Value labels/names for the codes on for the original variable	codes from each year of the GHS			New codes <b>HHTYPF1</b>	New Labels
	1984-1992 HHTYPF	1993-1997 HHTYPF	1998-2004 HHTYPF		
1 person only	1,2	1,2	1,2	1	1 PERSON ONLY
2 or more unrelated adults	3,4	3,4	3,4	2	2+ UNRELATED ADULTS
Married/Cohabiting couple					
With dependent children	9thru24	9thru24	-	13	COUPLE HOUSEHOLD, DEP CHILDREN
With non-dependent children	25thru28	25thru28	-	14	COUPLE HOUSEHOLD, INDEP CHILDREN
No children	5thru8	5thru8	-	15	COUPLE HOUSEHOLD, NO CHILDREN
Married couple					
With dependent children	-	-	9thru24	3	MARRIED COUPLE, DEP CHILDREN
With non-dependent children	-	-	25thru28	4	MARRIED COUPLE, INDEP CHILDREN
No children	-	-	5thru8	5	MARRIED COUPLE, NO CHILDREN
Lone parent					
With dependent children	29	29	29	6	LONE PARENT, DEP CHLDREN
With non-dependent children	30	30	30	7	LONE PARENT, INDEP CHILDREN
Two or more families	31	31	31	8	2+ FAMILIES
Same sex Cohabiting	-	33	33	9	SAME SEX COHAB
Cohabiting couple					
With dependent children	-	-	38thru53	10	COHAB COUPLE, DEP CHILDREN

With non-dependent children	-	-	54thru57	11	COHAB COUPLE, INDEP CHILDREN
No children	-	-	34thru37	12	COHAB COUPLE, NO CHILDREN
Missing values	All other missings	All other missings	All other missings	-5	MISSING VALUES

*SPSS syntax to recode the variable PHHTYF1*

\*de-select missing values\*.  
MISSING VALUES phhtyfp (.).  
EXE.

```
RECODE phhtyfp
  (1 , 2 =1)
  (3 , 4 =2)
  (9 thru 24=3)
  (25 thru 28 =4)
  (5 thru 8 =5)
  (29 =6)
  (30 =7)
  (31 =8)
  (33 =9)
  (38 thru 53 =10)
  (54 thru 57=11)
  (34 thru 37 =12)
  (62 thru 77 =13)
  (78 thru 81 =14)
  (58 thru 61 =15)
  (32 =-5) INTO phhtyf1.
EXECUTE.
```

```
VARIABLE LABELS phhtyf1 'Household type F (Grouped)'.
VALUE LABELS phhtyf1
  1 '1 PERSON ONLY'
  2 '2+ UNRELATED ADULTS'
  3 'MARRIED COUPLE, DEP CHILDREN'
  4 'MARRIED COUPLE, INDEP CHILDREN'
  5 'MARRIED COUPLE, NO CHILDREN'
  6 'LONE PARENT, DEP CHILDREN'
  7 'LONE PARENT, INDEP CHILDREN'
  8 '2+ FAMILIES'
  9 'SAME SEX COHAB'
  10 'COHAB COUPLE, DEP CHILDREN'
  11 'COHAB COUPLE, INDEP CHILDREN'
  12 'COHAB COUPLE, NO CHILDREN'
  13 'COUPLE HOUSEHOLD, DEP CHILDREN'
  14 'COUPLE HOUSEHOLD, INDEP CHILDREN'
  15 'COUPLE HOUSEHOLD, NO CHILDREN'
  -5 'Missing'.
```

### *The aggregate household level DVs*

<b>PNPERSON</b>	'No. of persons in the household'
<b>PN0TO4</b>	'No. of children aged <5 in household'
<b>PN0TO15</b>	'No. of children aged <16 in household'
<b>PN5TO15</b>	'No. of children aged 5 to 15 in household'
<b>PN16TO59</b>	'No. of adults aged 16 to 59 in household'
<b>PNAGEa60</b>	'No. of adults aged 60 and over in household'
<b>PNAGE65</b>	'No. of adults aged 65 and over in household'
<b>PTEENAGE</b>	'No. of teenagers in household'
<b>PYOUNGAG</b>	'Youngest age of person in the household'
<b>PNMALES</b>	'No. of males in household'
<b>PNFEMALE</b>	'No. of females in household'
<b>PNADMAL</b>	'No. of adult males in household'
<b>PNADFEMS</b>	'No. of adult females in household'

These aggregate household level derived variables were created using the SPSS syntax below and then mapped back on to the GHS Time Series Dataset.

```
SPSS syntax to recode the aggregate household level DVs

**Syntax file to create aggregate household level derived variables, which are then matched back
onto the GHS Time Series Dataset.
**Please note the actual file names used have been replaced with asterisks**.

MISSING VALUES page psex (-5).

Compute a1 = 1.
if page lt 5 a2 = 1.
if range(page,5,15) a3 = 1.
if range(page,16,59) a4 = 1.
if page gt 59 a5 = 1.
if page gt 64 a6 = 1.
if range(page,13,19) a7 = 1.
if page lt 15 a8 = 1.
if psex = 1 s1 = 1.
if psex = 2 s2 = 1.
if psex = 1 and page gt 15 as1 = 1.
if psex = 2 and page gt 15 as2 = 1.

EXE.

AGGREGATE
/OUTFILE ='*****1.sav'
/BREAK=yhserial
/pnperson = SUM(a1)
/pn0to4 = SUM(a2)
/pn5to15 = SUM(a3)
/pn16to59 = SUM(a4)
/pnage60 = SUM(a5)
/pnage65 = SUM(a6)
/pteenage = SUM(a7)
/pn0to15 = SUM(a8)
/pyoungag = MIN(page)
/pnmales = SUM(s1)
/pnfemales = SUM(s2)
/pnadmale = SUM(as1)
/pnadfems = SUM(as2).
EXE.

**Sort cases for matching.
**Note that every file must be sorted by the same variables
**in the same order to ensure that the data is not corrupted.
SORT CASES BY
```

yhserial ppersno (A).

\*\*Match data from the household level file to the person level file\*\*.

```
MATCH FILES /FILE=*  
/TABLE='*****.sav'  
/BY yhserial.  
EXECUTE.
```

```
RECODE npersons pn0to4 n0to15 pn5to15 pn16to59 ppage60 ppage65 pteenage youngage nmales  
nfemales pnamale pnamfems smkinhh (SYSMIS = 0) (ELSE=COPY).  
EXE.
```

VARIABLE LABELS

nperson	'No. of persons in the household'
pn0to4	'No. of children aged <5 in household'
pn0to15	'No. of children aged <16 in household'
pn5to15	'No. of children aged 5 to 15 in household'
pn16to59	'No. of adults aged 16 to 59 in household'
ppage60	'No. of adults aged 60 and over in household'
ppage65	'No. of adults aged 65 and over in household'
pteenage	'No. of teenagers in household'
pyoungag	'Youngest age of person in the household'
pmales	'No. of males in household'
pnfemales	'No. of females in household'
pnamale	'No. of adult males in household'
pnamfems	'No. of adult females in household'.

```
FORMAT nperson pn0to4 pn0to15 pn5to15 pn16to59 ppage60 ppage65 pteenage pyoungag pmales  
pnfemale pnamale pnamfems (F5.0).
```

## **PAGE5YR 'AGE GROUPED (5 YEARS)'**

1	0-4	11	50-54
2	5-9'	12	55-59
3	10-14	13	60-64
4	15-19	14	65-69
5	20-24	15	70-74
6	25-29	16	75-79
7	30-34	17	80-84
8	35-39	18	85-89
9	40-44	19	90+
10	45-49	-5	missing

### Notes

This variable recodes the variable PAGE. (for a detailed description of the original values on the GHS data for each year please see the metadata information for PAGE).

From the original GHS variable AGE in 2001 the oldest ages are categorised into 85+. In other words for 2001 there will be no people in the 90+ group, anyone aged 85+ will be recoded into the 85-89 age group.

#### SPSS syntax to recode the variable PAGE5YR

```
*de-select missing values*.
**MISSING VALUES page (*).
**EXE*.

IF (page = -5) page5yr = -5 .
EXECUTE .
IF (page>-1 & page<5) page5yr = 1 .
EXECUTE .
IF (page>4 & page<10) page5yr = 2 .
EXECUTE .
IF (page>9 & page<15) page5yr = 3 .
EXECUTE .
IF (page>14 & page<20) page5yr = 4.
EXECUTE .
IF (page>19 & page<25) page5yr = 5 .
EXECUTE .
IF (page>24 & page<30) page5yr = 6 .
EXECUTE .
IF (page>29 & page<35) page5yr = 7.
EXECUTE .
IF (page>34 & page<40) page5yr = 8.
EXECUTE .
IF (page>39 & page<45) page5yr = 9 .
EXECUTE .
IF (page>44 & page<50) page5yr = 10 .
EXECUTE .
IF (page>49 & page<55) page5yr = 11 .
EXECUTE .
IF (page>54 & page<60) page5yr = 12 .
EXECUTE .
IF (page>59 & page<65) page5yr = 13 .
EXECUTE .
IF (page>64 & page<70) page5yr = 14 .
EXECUTE .
IF (page>69 & page<75) page5yr = 15 .
EXECUTE .
IF (page>74 & page<80) page5yr = 16.
EXECUTE .
IF (page>79 & page<85) page5yr = 17 .
EXECUTE .
IF (page>84 & page<90) page5yr = 18 .
EXECUTE .
IF (page>89) page5yr = 19 .
```

EXECUTE .

VALUE LABELS page5yr

- 1 '0-4'
- 2 '5-9'
- 3 '10-14'
- 4 '15-19'
- 5 '20-24'
- 6 '25-29'
- 7 '30-34'
- 8 '35-39'
- 9 '40-44'
- 10 '45-49'
- 11 '50-54'
- 12 '55-59'
- 13 '60-64'
- 14 '65-69'
- 15 '70-74'
- 16 '75-79'
- 17 '80-84'
- 18 '85-89'
- 19 '90+'
- 5 'missing'.

VARIABLE LABELS page5yr 'AGE GROUPED (5 YEARS)'.  
EXE.

## **PAGEAD1 'AGE OF ADULTS GROUPED (6 BANDS)'**

1	16-19
2	20-24
3	25-34
4	35-49
5	50-59
6	60 and over
-5	Child or missing

### Notes

This variable was recoded from the variable PAGE, for a detailed description of the original values on the GHS data for each year please see the metadata information for PAGE.

#### SPSS syntax to recode the variable PAGEAD1

```
*de-select missing values*.
**MISSING VALUES page (*).
**EXE*.
```

```
IF (page < 16) pagead1 = -5 .
EXECUTE .
IF (page>15 & page<20) pagead1 = 1 .
EXECUTE .
IF (page>19 & page<25) pagead1 = 2 .
EXECUTE .
IF (page>24 & page<35) pagead1 = 3 .
EXECUTE .
IF (page>34 & page<50) pagead1 = 4 .
EXECUTE .
IF (page>49 & page<60) pagead1 = 5 .
EXECUTE .
IF (page>59) pagead1 = 6 .
EXECUTE .
```

```
VALUE LABELS pagead1
  1 '16-19'
  2 '20-24'
  3 '25-34'
  4 '35-49'
  5 '50-59'
  6 '60 and over'
-5 'child or missing'.
```

```
VARIABLE LABELS pagead1 'AGE OF ADULTS GROUPED (6 BANDS)'.
EXE.
```

## **PAGEAD2 'AGE OF ADULTS GROUPED (7 BANDS)'**

1	16-24
2	25-34
3	35-44
4	45-54
5	55-64
6	65-74
7	75 and over
-5	Child or missing

### Notes

This variable was recoded from the variable PAGE, for a detailed description of the original values on the GHS data for each year please see the metadata information for PAGE.

#### SPSS syntax to recode the variable PAGEAD2

```
*de-select missing values*.
**MISSING VALUES page (*).
**EXE*.
```

```
IF (page < 16) pagead1 = -5 .
EXECUTE .
IF (page>15 & page<20) pagead1 = 1 .
EXECUTE .
IF (page>19 & page<25) pagead1 = 2 .
EXECUTE .
IF (page>24 & page<35) pagead1 = 3 .
EXECUTE .
IF (page>34 & page<50) pagead1 = 4 .
EXECUTE .
IF (page>49 & page<60) pagead1 = 5 .
EXECUTE .
IF (page>59) pagead1 = 6 .
EXECUTE .
```

```
VALUE LABELS pagead1
  1 '16-19'
  2 '20-24'
  3 '25-34'
  4 '35-49'
  5 '50-59'
  6 '60 and over'
 -5 'child or missing'.
```

```
VARIABLE LABELS pagead1 'AGE OF ADULTS GROUPED (6 BANDS)'.
EXE.
```

### **PAGEAD3 'AGE OF ADULTS GROUPED (4 BANDS)'**

- 1 16-24
- 2 25-44
- 3 45-64
- 4 45-54
- 5 65 and over
- 5 Child or missing

#### **Notes**

This variable was recoded from the variable PAGE, for a detailed description of the original values on the GHS data for each year please see the metadata information for PAGE.

#### *SPSS syntax to recode the variable PAGEAD3*

```
*de-select missing values*.
**MISSING VALUES page (*).
**EXE*.
```

```
IF (page<16) pagead3 = -5 .
EXECUTE .
IF (page>15 & page<25) pagead3 = 1 .
EXECUTE .
IF (page>24 & page<45) pagead3 = 2 .
EXECUTE .
IF (page>44 & page<65) pagead3 = 3 .
EXECUTE .
IF (page>64) pagead3 = 4 .
EXECUTE .
```

```
VARIABLE LABELS pagead3 'AGE OF ADULTS GROUPED (4 BANDS)'.
```

```
VALUE LABELS pagead3
  1 '16-24'
  2 '25-44'
  3 '45-64'
  4 '65 and over'
 -5 'child or missing'.
```

## **PDVMARD1 'Marital Status (4 Groups)'**

- 1 Single
- 2 Married
- 3 Cohabiting/Same sex
- 4 Widowed/Divorced/Separated
- 5 missing

### **Notes**

This variable was recoded from the variable PDVMARDF, for a detailed description of the original values on the GHS data for each year please see the metadata information for PDVMARDF.

#### *SPSS syntax to recode the variable PDVMARD1*

\*de-select missing values\*.

\*\*MISSING VALUES pdvmardf (\*).

\*\*EXE\*.

RECODE pdvmardf (1=2)(2=3)(3=1)(4,5,6=4)(7=3) into pdvmard1.

VARIABLE LABELS pdvmard1 'Marital Status (4 Groups)'.

VALUE LABELS pdvmard1

1 'Single'

2 'Married'

3 'Cohabiting/Same sex'

4 'Widowed/Divorced/Separated'

-5 'missing'.

## **PDVMARD2 'Marital Status (3 Groups)'**

- 1 Single
- 2 Married/Cohabiting/Same sex
- 3 Widowed/Divorced/Separated
- 5 missing

### Notes

This variable was recoded from the variable PDVMARD1 which was originally derived from PDVMARDF. For a detailed description of the original values on the GHS data for each year please see the metadata information for PDVMARDF.

#### SPSS syntax to recode the variable PDVMARD2

```
*de-select missing values*.
**MISSING VALUES pdvmard1 (*).
**EXE*.

RECODE pdvmard1 (1=1)(2,3=2)(4=3) into pdvmard2.

VARIABLE LABELS pdvmard2 'Marital Status (3 Groups)'.

VALUE LABELS pdvmard2
  1 'Single'
  2 'Married/Cohabiting/Same Sex'
  3 'Widowed/Divorced/Separated'
 -5 'missing'.
```

### ***The DVs relating to HRP/HOH***

#### **PHRPMAR 'marital status of HRP/HOH'**

- 1 HRP/HOH Married
- 2 HRP/HOH Cohabiting
- 3 HRP/HOH Single
- 4 HRP/HOH Widowed
- 5 HRP/HOH Divorced
- 6 HRP/HOH Separated
- 7 HRP/HOH Same sex couple
- 5 Missing

#### **PHRPAGE 'age of HRP/HOH'**

Numeric values

- 99 99 or older
- 5 Missing

#### **Notes**

From the original GHS variable AGE in 2001 the oldest ages are categorised into 85+. In other words for 2001 if the HRP/HOH person is aged 85+ they will be recoded as aged 85.

#### **PHRPSEX 'sex of HRP/HOH'**

- 1 HRP/HOH Male
- 2 HRP/HOH Female
- 5 Missing

*SPSS syntax to recode the 3 variables: PHRPMAR, PHRPAGE & PHRPSEX*

**\*\*Syntax to create the 3 variable involved the principal steps\*\*.**  
**\*\*Please note the actual file names used have been replaced with asterisks\*\*.**

GET

FILE='\*\*\*\*\*1.sav'  
/KEEP = preltohr pDVMarDF page psex YHserial.

SELECT IF(preltohr = 0).

SAVE OUTFILE='\*\*\*\*\*2.sav'.

COMPUTE pHRPMar = pDVMarDF.  
COMPUTE pHRPAge = page.  
COMPUTE pHRPsex = psex.

VARIABLE LABELS

pHRPMar 'marital status of HRP/HOH'  
pHRPAge 'age of HRP/HOH'  
pHRPsex 'sex of HRP/HOH'.

VALUE LABELS pHRPsex 1 'HRP/HOH Male' 2 'HRP/HOH Female' -5 'Missing'.

VALUE LABELS pHRPMar

1 'HRP/HOH Married'  
2 'HRP/HOH Cohabiting'  
3 'HRP/HOH Single'  
4 'HRP/HOH Widowed'  
5 'HRP/HOH Divorced'  
6 'HRP/HOH Separated'  
7 'HRP/HOH Same sex couple'  
-5 'Missing'.

VALUE LABELS pHRPAge 99 '99 or older' -5 'Missing'.

```
FORMAT pHRPMar pHRPAge pHRPSex (F5.0).
SORT CASES BY YHserial (A).

SAVE OUTFILE='*****2.sav'.

GET
  FILE='*****2.sav'
/KEEP = YHserial pHRPSex pHRPMar pHRPAge.

SORT CASES BY YHserial (A).
SAVE OUTFILE='*****2.sav'.

GET
  FILE='*****1.sav'.
SORT CASES BY YHserial (A).
EXE.

**Match data from the household level file to the person level file**.
MATCH FILES /FILE=*
  /TABLE='*****2.sav'
  /BY yhserial.
EXECUTE.

SAVE OUTFILE='*****3.sav'.
```

### **PDVILO3A 'DV for ILO in employment - 3 categories'**

ILO [International Labour Organisation definition of unemployment](#)<sup>xvii</sup>

- 1 In employment
- 2 Unemployed
- 3 Economic inactive
- 5 Missing.

*Names of the original GHS variables used to create PDVILO3A:*

EMPLOYED, ECONSTA, ECSTAA, ECSTILO, ECSTIO96 and DVILO3A

#### Notes:

The variable PDVILO3A is based on DVILO3a from the 2003 dataset.

The main problem in trying to create a comparable variable over time is the nature of definition for inactive and unemployed.

Between 1992 and 2003 these definitions are based in reference to ILO categories. Before 1992 the definitions are not explicitly based on ILO (International Labour Organisation) criteria.

From 1985 full-time students were classified as 'working', 'unemployed' or 'inactive' according to their own reports of what they were doing in the reference week. In previous years they were all classified as 'inactive'.

From 1985 people on Youth Training were classified as working if they were with an employer for work experience in the reference week, or as 'inactive' if at college. Since 1989 all those on schemes Youth Training, Employment Training (ET) and (1989 only) Job Training Scheme (JTS) have been classified as 'working'.

Since 1991 an 'unemployed' person is someone who has been looking for work in the four weeks prior to interview and would be available to start work in the next two weeks, or is waiting to start a job. This has had the effect of reclassifying some informants from 'economically inactive' to economically active and vice versa. In the most part, this definitional change had little effect on the numbers of 'unemployed' men, but numbers of females classified as 'unemployed' increased by one fifth.

For more information on this refer to GHS Report 1991, Chapter 5.

Concerning the original variables DVILO3A, ECSTILO & ECSTILO96:

'5 other unemployed' are those who describe themselves as unemployed but who have either not sought work in the last four weeks or who would not be able to start work in the next two weeks. Under the ILO definition of unemployed these would be included as economically inactive persons.

From 1994 unpaid family workers who were not in a paid job, away from a job, waiting to take up a job or on a Govt scheme are included in code 1 'working'. Men aged 70+ and women aged 65+ no longer code 8 at Wantajob.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS							New codes
	1972	1982	1983	1989	1991	1996	1998	
	1981	1982	1988	1990	1995	1996	2004	
	EMPLOYED	EMPLOYED	ECONSTA	ECSTAA	ECSTILO	ECSTIO96	DVILO3A	<b>PDVILO3A</b>
In Employment	1	1	1	1	1	1	1	1
EMPD: employee								1
EMPD:SELF-EMP								1
GOVT SCH WTH EMP				2	2	2		1
GOVT SCH AT COLL				3	3	3		1
ILO Unemployed					4	4	2	2

UNEMPD:NEVWKD								2
UNEMP:WAS EMPEE								2
UNEMP:WAS SELFEM								2
OTHER UNEMPLOYED					5	5		3
WAITING	3	3	2	4				2
LOOKING	4	4	3	5				2
Economically Inactive					3		3	3
PERM UNABLE WORK	7	7	5	7	6	6		3
RETIRED	8	8	6	8	7	7		3
KEEPING HOUSE	9	9	7	9	8	8		3
STUDENT	6	6	8	10	9	9		3
OTHER ECON INACT					10	10		3
EMPLOYR-BASED ET								3
TEMPORARILY SICK	5	5	4	6				3
ON YOP SCHEME		14	10					3
OTHER	10	10	9	11				3
Missing values	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	-5

SPSS syntax to recode the variable PDVILO3A

```

*de-select missing values*.
MISSING VALUES pdvilo3a (.).
EXE.

DO IF (year<=1982).
+   RECODE
      pdvilo3a
      (1=Copy) (3 thru 4=2) (5 thru HIGHEST=3)(ELSE=-5).
END IF.
EXECUTE .

DO IF (year>=1983 and year<=1988).
+   RECODE
      pdvilo3a
      (1=Copy)(2 thru 3=2)(4 thru HIGHEST=3)(ELSE=-5).
END IF.
EXE.

DO IF (year>=1989 and year<=1990).
+   RECODE
      pdvilo3a
      (1 thru 3=1)(4 thru 5=2)(6 thru HIGHEST=3)(ELSE=-5).
END IF.
EXE.

DO IF (year>=1991 and year<=1996).
+   RECODE
      pdvilo3a
      (1 thru 3=1) (4=2) (5 thru HIGHEST=3) (ELSE=-5).
END IF.
EXE.

DO IF (year>=1998).
+   RECODE
      pdvilo3a
      (1 thru HIGHEST=Copy) (ELSE=-5).
END IF.

VARIABLE LABELS pdvilo3a 'DV for ILO in employment - 3 categories'.
VALUE LABELS pdvilo3a  1 'In employment'
                    2 'Unemployed'
                    3 'Economic inactive'
                    -5 'Missing'.
EXE.

```

## PEDFULL 'Education Level - full form'

0	No Qualifications
1	Higher Degrees
2	1st Degree
3	Teaching Qualification
4	Other Higher Qualification
5	Nursing Qualification
6	A-Level or equivalent (number not specified)
7	A-Level or equivalent (2+)
8	A-Level or equivalent (in 1 subject)
9	GCSE level or equivalent (5 or more)
10	GCSE level or equivalent 1-4, above grade C
11	GCSE level or equivalent 1-4, below grade C
12	Clerical and Commercial Qualification, no O-levels
13	CSE other
14	Apprenticeship
15	STDGRD 6-7NOAWRD
16	Foreign Qualification
17	Other Qualification
18	O-levels (do not know number)
19	Never went to school
-5	Missing

*Names of the original GHS variables used to create PEDFULL:*  
HEDQUAL, EDLEV and EDLEV00

### Notes:

The variable PEDFULL was created to be comparable over time, and retains all the categories, so that an individual researcher can choose how to recode the variable into a format that will suit them. The category labels as close as they were to the original variables used on the original GHS files.

The question is only asked of persons >15 and <70. All other age groups are categorised as missing.

There are some further issues regarding the consistency of the categories overtime for the recode, these are listed in the table below.

Pedfull code	Pedfull category	Inconsistency
6	A-Level or equivalent (number not specified)	From 1972-1982, the number of A-levels or equivalent achieved is not specified. From 1983-2004, this category is divided into those achieving 2 or more A-level's and those achieving just 1 A-level. This causes problems for re-coding into NQF and NVQ level equivalents as 1 A-level equals level 2 whereas 2+ A-levels equals level 3.
7 8	A-Level or equivalent (2+) A-Level or equivalent (in 1 subject)	From 1996-1998, the number of cases with 2+ A-levels increased from 511 to 1,505. During the same period the number with 1 A-level decreased from 1,175 to 497. For both categories the values for 2000 onwards stayed at the same rate as for 1998.
12	Clerical and Commercial Qualification, no O-levels	This category does not exist from 2000 onwards.
13	CSE other	This original GHS category does not exist from 2000 onwards. The respondents with this type of qualification have been included in '10 - CSE below grade 1, GCSE below grade C' from 2000 onwards. To create PEDFULL these were recoded as '11 - GCSE level or equivalent 1-4, below grade C'.
15	STDGRD 6-7NOAWRD	This category only exists for 1989, 1992-1994 and 1996. There are a very small number of cases (a maximum of 2 in any year). DfES advise that this category should be recoded as 'other qualifications'.

16	Foreign Qualification	This category does not exist from 2000 onwards. GHS advise that respondents with this type of qualification have been categorised as 'other qualifications' from 2000 onwards.
17	Other Qualification	GHS advise that this category is not comparable with earlier years from 2000 onwards. It is not possible to separate foreign qualifications and other qualifications given the current set of questions.
18	O-Levels (do not know number)	This category only exists from 1973-1976. There are a very small number of cases (below 50 for all years).
19	Never went to school	This category is only available from 1983-1986 and 1989-2004. From 1989-1993 there are a large number of cases (around 8,000 – 9,000). For the same years there are no missing values (or only 4 in 1992). Published tables in Living in Britain exclude those that 'never went to school' from the baseline.
-5	Missing	There are no cases from 1989-1990 and in 1993. In 1992 there are only 4 cases. In the corresponding years there are a large number of cases who 'never went to school'.

Value labels/names for the codes on the original GHS dataset	codes from each year of the GHS						New codes
	1972 1976	1977 1982	1983 1986	1987 1988/89	1989/90 1998/99	2000/1 2004/5	
	HEDQUAL	HEDQUAL	EDLEV	EDLEV	EDLEV	EDLEV00	PEDUFULL
No Qualifications	0	15	16	17	17	13	0
Higher Degrees	1	1	1	1	1	1	1
1st Degree	2	2	2	2	2	2	2
Teaching Qualification	3	3	3	3	3	3	3
Other Higher Qualification	4	4	4	4	4	4	4
Nursing Qualification	5	5	5	5	5	5	5
A-Level or equivalent (number not specified)	6	6					6
A-Level or equivalent (2+)			6	6	6	6	7
A-Level or equivalent (in 1 subject)			7	7	7	7	8
GCSE level or equivalent (5 or more)	7	7	8	8	8	8	9
GCSE level or equivalent 1-4, above grade C	8	8	9	9	9	9	10
GCSE level or equivalent 1-4, below grade C	9	9	10	10	10	10	11
Clerical and Commercial Qualification, no O-levels	10	10	11	11	11		12
CSE other	11	11	12	12	12		13
Apprenticeship	12	12	13	13	13	11	14
STDGRD 6-7NOAWRD				14	14		15
Foreign Qualification	13	13	14	15	15		16
Other Qualification	14	14	15	16	16	12	17
O-levels (do not know number)	17	17					18
Never went to school			20		-9	-9	19
Missing	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	-5

SPSS syntax to recode the variable PEDFULL

\*\*Create a new variable from pedlev\*\*.  
 COMPUTE pedfull = pedlev00.  
 EXE.

```

*de-select missing values*.
MISSING VALUES pedfull (.).
EXE.

DO IF (year>=1972 and year<=1976).
+   RECODE
      pedfull
(0 thru 6=Copy)(7=9)(8=10)(9=11)(10=12)(11=13)(12=14)(13=16)(14=17)(17=18)(ELSE=-5).

ELSE IF (year>=1977 and year<=1982).
+   RECODE
      pedfull
(15=0)(1 thru 6=Copy)(7=9)(8=10)(9=11)(10=12)(11=13)(12=14)(13=16)(14=17)(17=18)(ELSE=-5).

ELSE IF (year>=1983 and year<=1986).
+   RECODE
      pedfull
(16=0)(1thru5=Copy)(6=7)(7=8)(8=9)(9=10)(10=11)(11=12)(12=13)(13=14)(14=16)(15=17)(20=19)(ELSE=-5).

ELSE IF (year>=1987 and year<=1988).
+   RECODE
      pedfull
(17=0)(1thru5=Copy)(6=7)(7=8)(8=9)(9=10)(10=11)(11=12)(12=13)(13=14)(14=15)(15=16)(16=17)(ELSE=-5).

ELSE IF (year>=1989 and year<=1998).
+   RECODE
      pedfull
(17=0)(1 thru 5=Copy)(6=7)(7=8)(8=9)(9=10)(10=11)(11=12)(12=13)(13=14)(14=15)(15=16)(16=17)(-9=19)(ELSE=-5).

ELSE IF (year>=2000).
+   RECODE
      pedfull
(13=0)(1 thru 5=Copy)(6=7)(7=8)(8=9)(9=10)(10=11)(11=14)(12=17)(-9=19)(ELSE=-5).
END IF.

VARIABLE LABELS pedfull 'Education Level - full form'.
VALUE LABELS pedfull
  0 'No Qualifications'
  1 'Higher Degrees'
  2 '1st Degree'
  3 'Teaching Qualification'
  4 'Other Higher Qualification'
  5 'Nursing Qualification'
  6 'A-Level or equivalent (number not specified)'
  7 'A-Level or equivalent (2+)'
  8 'A-Level or equivalent (in 1 subject)'
  9 'GCSE level or equivalent (5 or more)'
  10 'GCSE level or equivalent 1-4, above grade C'
  11 'GCSE level or equivalent 1-4, below grade C'
  12 'Clerical and Commercial Qualification, no O-levels'
  13 'CSE other'
  14 'Apprenticeship'
  15 'STDGRD 6-7NOAWRD'
  16 'Foreign Qualification'
  17 'Other Qualification'
  18 'O-levels (do not know number)'
  19 'Never went to school'
  -5 'Missing'.
EXECUTE.

```

**DEGREE 'Highest qualification a degree or higher'**

- 1 Degree or higher
- 0 Other or no qualifications
- 5 Missing

**Notes:**

This variable was specifically derived for the GHS Time Series Dataset from the variable PEDFULL, for a detailed description of the original values on the GHS data for each year please see the metadata information for the variable PEDFULL.

Value labels/names for the codes on for the original variable	codes from each year	New codes
	1972	
	2004 PEDUFULL	<b>DEGREE</b>
No Qualifications	0	0
Higher Degrees	1	1
1st Degree	2	1
Teaching Qualification	3	0
Other Higher Qualification	4	0
Nursing Qualification	5	0
A-Level or equivalent (number not specified)	6	0
A-Level or equivalent (2+)	7	0
A-Level or equivalent (in 1 subject)	8	0
GCSE level or equivalent (5 or more)	9	0
GCSE level or equivalent 1-4, above grade C	10	0
GCSE level or equivalent 1-4, below grade C	11	0
Clerical and Commercial Qualification, no O-levels	12	0
CSE other	13	0
Apprenticeship	14	0
STDGRD 6-7NOAWRD	15	0
Foreign Qualification	16	0
Other Qualification	17	0
O-levels (do not know number)	18	0
Never went to school	19	0
Missing	-5	-5

*SPSS syntax to recode the variable DEGREE*

```
*de-select missing values*.
MISSING VALUES pedfull (.).
EXE.
```

```
DO IF pedfull = -5.
+   COMPUTE degree = -5.
ELSE IF (pedfull >=1 and pedfull <= 2).
+   COMPUTE degree=1.
ELSE.
+   COMPUTE degree=0.
END IF.
EXECUTE.
```

```
VARIABLE LABELS degree 'Highest qualification a degree or higher'.
VALUE LABELS degree
    1 'Degree or higher'
    0 'Other or no qualifications'
    -5 'Missing'.
```

```
MISSING VALUES degree (-5).
```

**PNUMVEH 'Number of cars'**

- 0 No car available
- 1 One car
- 2 Two cars
- 3 Three or more cars
- 5 Missing

*Names of the original GHS variables used to create PNUMVEH:*  
 NCARS, CAR (renamed to PCAR for the recode for years 1983-1996/7),  
 TYPEVCL1-7 and NUMCARS

**Notes:**

This variable:

- Counts numbers of cars and vans only (i.e. excluding motorcycles and other vehicles).
- This means creating a new variable for 1998/99 and 2000/01.
- It appears that there is an error in the reporting of GHS data for 1998/99 and 2000/01 – therefore this will mean that our data doesn't correspond exactly with the published data.

For the years 1983-1996 PNUMVEH was derived from 2 variables CAR and NCARS. In these years number of cars was ascertained from two questions. The first asked whether a person had access to a car or not (CAR). If yes, the second question (NCARS) asked how many cars they had access to. In 1983-1996, if someone answered no to the first question (ie they didn't have access to a car (CAR=2)), then they were not categorised to zero for NCARS (for the number of cars they had access to) and therefore the variable CAR was needed to ascertain those households without a car (CAR=2).

In 1998 and 2000 the number of cars has to be derived from a count of 'Vehicle Type' (TYPEVCL). NUMCARS is derived from the variable, vehicle type (TYPEVCL1 – TYPEVCL8). The type of vehicle is recorded for each vehicle available in the household. The recode for 1998 and 2000 counts numbers of cars (1) and light vans (2) only (i.e. excluding motorcycles and other vehicles).

From 2001 onwards the variable NUMCARS could be used.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS					New codes
	1972 1978	1979 1982	1983 1996	1998 2000	2001 2004	
	<b>NCARS</b>	<b>NCARS</b>	<b>NCARS (&amp; CAR)</b>	<b>TYPEVCL1-8</b>	<b>NUMCARS</b>	<b>PNUMVEH</b>
No car available	0	0	0	See below	0	0
Three or more cars			3		3	3
Missing	All other missings	All other missings	All other missings		All other Missings	-5

In the years 1983-1996/97 PNUMVEH was derived from 2 variables CAR and NCARS. The number of households without access to a car/van has been derived from CAR (Access to a car – Yes/No, where no=2). The number of cars available to the household is derived from NCARS.

In 1998/99 and 2000/1 the number of cars has to be derived from a count of 'Vehicle Type' (TYPEVCL1 – TYPEVCL8). The type of vehicle is recorded for each vehicle available in the household. To create PNUMCARS the SPSS recode counts only the cars (1) and light vans (2).

	1998/9 and 2000/1							
	TYPEVCL1	TYPEVCL2	TYPEVCL3	TYPEVCL4	TYPEVCL5	TYPEVCL6	TYPEVCL7	TYPEVCL8
Car	1	1	1	1	1	1	1	1

Light van	2	2	2	2	2	2	2	2
Motor cycle	3	3	3	3	3	3	3	3
Some other motor vehicle	4	4	4	4	4	4	4	4

SPSS syntax to recode the variable PNUMVEH

```

**1998 & 2000 variable is derived from typevc1.
**1983-1996, no access to a car is derived from pcar.
*****

*de-select missing values*.
MISSING VALUES pnumveh (.).
EXE.

DO IF (year<1983).
+   RECODE pnumveh (0=0)(1=1)(2=2)(3 thru HIGHEST=3)(ELSE=-5) .

ELSE IF (year>=1983 and year<=1996).
+   DO IF (pcar=2).
+     COMPUTE pnumveh=0.
+   ELSE.
+     RECODE pnumveh (1=1)(2=2)(3 thru HIGHEST=3)(ELSE=-5) .
+   END IF.

ELSE IF (year>=1998 and year<=2000).
+   COUNT pnumveh= ptypevc1 ptypevc2 ptypevc3 ptypevc4 ptypevc5 ptypevc6 ptypevc7
+   ptypevc8 (1 thru 2).
+   RECODE pnumveh (0=0)(1=1)(2=2)(3 thru HIGHEST=3)(ELSE=-5) .

ELSE IF (year>=2001).
+   RECODE pnumveh (0=0)(1=1)(2=2)(3 thru HIGHEST=3)(ELSE=-5) .
END IF.

VARIABLE LABELS pnumveh 'Number of cars'.
VALUE LABELS pnumveh
    0   'No car available'
    1   'One car'
    2   'Two cars'
    3   'Three or more cars'
   -5   'Missing'.
EXECUTE .

```

## PGENHLTH 'Health on the whole in last 12 months'

- 1 Good
- 2 Fairly Good
- 3 Not Good
- 5 Missing

*Names of the original GHS variables used to create PGENHLTH:*  
HEALTH and GENHLTH

### Notes:

Introduced in 1977. Asked of adults only up to 1998 and asked of all persons from 2000 onwards.

Not asked of proxies (assuming this to be the case before 1984).

In 1977 the question is introduced by the following sentence: 'Quite a lot is known about people's visits to doctors and hospitals, but little is known about people who don't go to a doctor and the sorts of difficulties they may have'. - this is assumed not to have an effect on answers given.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS		New codes
	1977	1983	
	1982	2004	
	<b>HEALTH</b>	<b>GENHLTH</b>	<b>PGENHLTH</b>
GOOD	1	1	1
FAIRLY GOOD	2	2	2
NOT GOOD	3	3	3
Missing	All other missings	All other missings	-5

### *SPSS syntax to recode the variable PGENHLTH*

```
*de-select missing values*.
MISSING VALUES pgenhlth ().
EXE.

RECODE
  pgenhlth
  (1 thru 3 =Copy) (ELSE= -5).
VARIABLE LABELS pgenhlth 'Health on the whole in last 12 months'.
VALUE LABELS pgenhlth 1 'Good' 2 'Fairly Good' 3 'Not Good' -5 'Missing'.
EXECUTE .
```

## **PILLNESS 'Any longstanding illness or disability'**

Do you have any long-standing illness, disability or infirmity? By long-standing, I mean anything that has troubled you over a period of time or that is likely to affect you over a period of time?

Note: the phrase 'By longstanding, I mean anything with has troubled you over a period of time or which is likely to affect you over a period of time' was not included in 1972-1974 but was included from then onwards.

- 1 Yes
- 2 No
- 5 Missing

*Names of the original GHS variables used to create PILLNESS:*  
 CHRON and ILLNESS

### Notes:

In 1972, the question asked was 'Do you (or any of your children aged under 15) suffer from any LLTI, disability or infirmity? In 1972 there are two variables for illness: CHRON (for persons aged 15+) and CHRONKID (aged under 15). CHRONKID has not been included on the GHS Time Series Dataset. Since 1972, the question asks about the informant's LLTI.

In 1973 there labels 'IN HOSPITAL NOW' for code 3, investigation showed that there was no data for value 3 even though there were labels. The labels and data do not match. The label 'IN HOSPITAL NOW' was excluded for the GHS Time Series Dataset.

In 1977 and 1978 the respondent was given a check list of chronic illnesses and then asked which problems they themselves had. A higher percentage reported illnesses in these years. Therefore 1977 and 1978 are not comparable to other years and are not included on the dataset.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS				New codes
	1972	1973	1974 1982	1983 2004	
	CHRON	CHRON	CHRON	ILLNESS	
YES	1	1	1	1	1
NO	2	2	2	2	2
Missing	All other missings	All other missings	All other missings	All other missings	-5

(NOT ASKED 1977 AND 1978)

### SPSS syntax to recode the variable PILLNESS

```
*de-select missing values*.
MISSING VALUES pillness (.).
EXE.
RECODE
  pillness
    (1 thru 2=Copy) (ELSE=-5).
VARIABLE LABELS pillness 'Any longstanding illness or disability'.
VALUE LABELS pillness 1 'Yes' 2 'No' -5 'Missing'.
EXECUTE .
```

## PLIMITAC 'If longstanding illness limits activity'

### *Limiting long-standing illness*

For each illness mentioned above

Does this illness or disability (Do any of these illnesses or disabilities) limit your activities in any way? Yes, No

- 1 Yes
- 2 No
- 5 Missing

*Names of the original GHS variables used to create PLIMITAC:*

ANYLIMIT, for years 1974-1976 ANYLIM1-ANYLIM4 and LIMITACT

### Notes:

Details for each year:

1973 - ANYLIMIT

For each illness mentioned above does (quote complaint) limit you activity in any way? This is asked for the 3 complaints listed. In 1973 ANYLIMIT is available on the dataset and it is assumed that this combines the answers to all the complaints - i.e. yes if any of the complaints limit activities.

1974-1976 - ANYLIM1-ANYLIM4

As above, but asks up to 4 complaints, and there are 4 variables on the dataset - one for each variable. However, in 1974 it would appear that the variable is yes if any of the complaints limit activity - rather than presented as 4 variables in 1974. For the years 1974-1976 the variables ANYLIM1-ANYLIM4 were recoded to into 1 variable at an earlier stage, the SPSS syntax for this recode is given below:

```
MISSING VALUES anylim1 anylim2 anylim3 anylim4 (.).
DO IF (anylim1 = 1 OR anylim2 = 1 OR anylim3 = 1 OR anylim4 = 1).
  COMPUTE anylimit = 1.
ELSE IF (anylim1 = 2 OR anylim2 = 2 OR anylim3 = 2 OR anylim4 = 2).
  COMPUTE anylimit = 2.
ELSE.
  COMPUTE anylimit = -5.
END IF.
EXE.
```

1977-1978 - Not asked.

1979-1982 - ANYLIMIT

Does this illness or disability limit your activities in any way?

1983-2004 - LIMITACT

As for 1979-82, but new variable name.

When checking the GHS Time Series Dataset it was found that in 1976 1 person who has no longstanding illness but has been recoded as 'Yes' if longstanding illness limits activity.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS				New codes
	1973	1974	1979	1983	
		1976	1982	2004	
	<b>ANYLIMIT</b>	<b>ANYLIM1/ANYLIM2/ ANYLIM3/ANYLIM4 (recoded to ANYLIMIT at an earlier stage)</b>	<b>ANYLIMIT</b>	<b>LIMITACT</b>	<b>PLIMITAC</b>
YES	1	1	1	1	1
NO	2	2	2	2	2
Missing	All other Missings	All other missings	All other missings	All other Missings	-5

(NOT ASKED 1977 AND 1978)

*SPSS syntax to recode the variable PLIMITAC*

```
*de-select missing values*.
MISSING VALUES plimitac (.).
EXE.
```

```
RECODE
  plimitac
    (1 thru 2=Copy) (ELSE=-5).
```

```
VARIABLE LABELS plimitac 'If longstanding illness limits activity'.
VALUE LABELS plimitac 1 'Yes' 2 'No' -5 'Missing'.
EXECUTE .
```

**PCUTDOWN 'Illness/injury reduce activity'**

- 1 Yes
- 2 No
- 5 Missing

*Names of the original GHS variables used to create PCUTDOWN:*  
 RESTRACT and CUTDOWN

Notes:

Asked of all persons. Question not asked 1977–78.

In 1972 there are two variables for 'illness restricting activity'; RESTRACT (respondents aged 15+) and RESTRKID (those aged <15). Only RESTRACT has been included on the GHS Time Series Dataset.

There have been slight changes to the wording over time:

1972 - RESTRACT

Now think about the things you usually do, did you have to cut down on these at all because of (this illness/disability, or some other) illness or injury during the two weeks ending last Sunday?

1973 - RESTRACT

Now think about the things you usually do. During the 2 weeks ending last Sunday, did you have to cut down on the things you usually do because of (this illness/disability, or some other) illness or injury?

1974-1976 - RESTRACT

Now I'd like you to think about all the things you usually do. For example, the things you do every day (at work/school) about the house, during your free time, etc. During the 2 weeks ending last Sunday, did you have to cut down on the things you usually do because of (this illness/disability, or some other) illness or injury?

1977-1978 - Not asked.

1979-1981 - RESTRACT

Now I'd like you to think about the 2 weeks ending yesterday and the things you usually do - for example, the things you do every day (at work/school) about the house, during your free time, etc. During the 2 weeks ending yesterday, did you have to cut down on the things you usually do because of (this illness/disability, or some other) illness or injury?

1982 - RESTACT

Now I'd like you to think about the 2 weeks ending yesterday. During those 2 weeks, did you have to cut down on any of the things you usually do (about the house/at work or in your free time) because of (answers given in the previous question) or some other illness or injury?

1983 - 2004 - CUTDOWN

As above but new variable name.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS		New codes
	1972	1983	
	1982	2004	
	<b>RESTRACT</b>	<b>CUTDOWN</b>	<b>PCUTDOWN</b>
YES	1	1	1
NO	2	2	2
Missing	All other missings	All other missings	-5

(NOT ASKED IN 1977 AND 1978)

SPSS syntax to recode the variable PCUTDOWN

```
*de-select missing values*.
MISSING VALUES pcutdown (.).
EXE.

RECODE
  pcutdown
    (1 thru 2=Copy) (ELSE=-5).
VARIABLE LABELS pcutdown 'Illness/Injury reduce activity'.
VALUE LABELS pcutdown 1 'Yes' 2 'No' -5 'Missing'.
EXECUTE .
```

## **PLONGILL 'LLSI OR NON-LLSI'**

LLSI Limiting Longstanding Illness

- 1 Limiting longstanding illness
- 2 Non-limiting longstanding illness
- 3 No longstanding illness
- 5 Missing

### Notes:

This variable was specifically derived for the GHS Time Series Dataset from the variables PAGE, PSCHEDTY, PLIMITAC and PILLNESS, please see the individual variables for further details of how they were originally derived.

#### SPSS syntax to recode the variable PLONGILL

```
*de-select missing values*.
**MISSING VALUES page pschedty plimitac (*).
**EXE*.
```

```
DO IF (page GT 16 AND pschedty=3).
+   COMPUTE plongill = -5.
ELSE IF (pillness = 2).
+   COMPUTE plongill = 3.
ELSE IF (pillness = -5).
+   COMPUTE plongill = -5.
ELSE.
+   COMPUTE plongill = plimitac.
END IF.
```

```
VALUE LABELS plongill
  1 'Limiting longstanding illness'
  2 'Non-limiting longstanding illness'
  3 'No longstanding illness'
 -5 'Missing'.
```

```
VARIABLE LABELS plongill 'LLSI OR NON-LLSI'.
```

## PLSIRA 'IF LLSI or restricted activity'

- 1 LLSI and Rest Activity
- 2 Rest Activity and non-LLSI
- 3 LLSI ONLY
- 4 Non-LLSI only
- 5 Restricted Activity only
- 6 No reported illness
- 7 Anything else
- 5 Missing

### Notes:

This variable is derived from the variables PAGE, PSCHEDTY, PCUTDOWN PLIMITAC, PLONGILL and PILLNESS, please see the individual variables for further details of how they were originally derived.

### SPSS syntax to recode the variable PLSIRA

```
*de-select missing values*.
**MISSING VALUES page pschedty plongill pcutdown plimitac (*).
**EXE*.
```

```
DO IF (page GT 15 AND pschedty = 3).
+   COMPUTE plsira = -5.
ELSE IF (plongill = -5 OR pcutdown = -5).
+   COMPUTE plsira = -5.
ELSE IF (plimitac = 1).
+   DO IF (pcutdown = 1).
+       COMPUTE plsira = 1.
+   ELSE IF (pcutdown = 2).
+       COMPUTE plsira = 3.
+   END IF.
ELSE IF (pillness = 1).
+   DO IF (pcutdown = 1).
+       COMPUTE plsira = 2.
+   ELSE IF (pcutdown = 2).
+       COMPUTE plsira = 4.
+   END IF.
ELSE IF (pillness = 2).
+   DO IF (pcutdown = 1).
+       COMPUTE plsira = 5.
+   ELSE IF (pcutdown = 2).
+       COMPUTE plsira = 6.
+   END IF.
ELSE.
+   COMPUTE plsira=-5.
END IF.
```

```
VALUE LABELS plsira
-5 'Missing'
1 'LLSI and Rest Activity'
2 'Rest Activity and non-LLSI'
3 'LLSI ONLY'
4 'Non-LLSI only'
5 'Restricted Activity only'
6 'No reported illness'
7 'Anything else'.
```

```
VARIABLE LABELS plsira 'IF LLSI or restricted activity'.
EXE..
```

## PDOCTALK 'Consulted doctor last 2 wks (exc.hosp)'

- 1 Yes
- 2 No
- 5 Missing

*Names of the original GHS variables used to create PDOCTALK:*  
DOCCHAT and DOCTALK

### Notes:

For 1972 the number of 'No' responses is unusually low. From checking the 1972 questionnaire, this question was asked about ALL respondents, not just those that had specified a limiting illness or disability. Also in 1972 there were two variables for adults aged 15+, and children.

In 1977, the number of 'No' responses is low. The frequencies are correct but there are a large number of Missing values.

In 1982, this variable only has data for 'Yes' (DOCCHAT=1) and Missing. There are no values for No (DOCCHAT=2).

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS		New codes
	1972	1983	
	1982	2004	
	<b>DOCCHAT</b>	<b>DOCTALK</b>	<b>PDOCTALK</b>
YES	1	1	1
NO	2	2	2
Missing	All other missings	All other missings	-5

### SPSS syntax to recode the variable PDOCTALK

```
*de-select missing values*.
MISSING VALUES pdoctalk (.).
EXE.
```

```
RECODE
  pdoctalk
```

```
(1 thru 2=Copy) (ELSE=-5).
```

```
VARIABLE LABELS pdoctalk 'Consulted doctor last 2 wks (exc hospital)'.
VALUE LABELS pdoctalk 1 'Yes' 2 'No' -5 'Missing'.
EXECUTE .
```

## PNCHATS 'Number of GP consultations in the last 2 weeks'

Numeric values

0 No consultations with GP

-5 Missing

*Names of the original GHS variables used to create PCUTDOWN:*  
NDOCCHAT and NCHATS

### Notes:

In the years 1972, 1973 and 1974 in some cases people who did not consult a doctor (PDOCTALK=0), have a value for the number of consultations as greater than one.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS						New codes
	1972	1978	1983	1987	1998		
	1977	1982	1986	1996	2003	2004	<b>PNCHATS</b>
	<b>NDOCCHAT</b>	<b>NDOCCHAT</b>	<b>NCHATS</b>	<b>NCHATS</b>	<b>NCHATS</b>	<b>NCHATS</b>	
NO CONS WITH GP				0	0		0
NUMERIC	√	Once, Twice etc	√	√	√	√	(numeric)
Missing	All other missings	All other missings	All other missings	All other missings	All other missings	All other missings	-5

### SPSS syntax to recode the variable PNCHATS

\*de-select missing values\*.

MISSING VALUES pnchats ().

EXE.

RECODE

pnchats

(0 thru HIGHEST=Copy) (ELSE=-5).

VARIABLE LABELS pnchats 'Number of GP consultations in the last 2 weeks'.

VALUE LABELS pnchats 0 'No consultations with GP' -5 'Missing'.

EXECUTE .

## POUTPATNT 'Hospital outpatient attend - last 3mths'

- 1 Yes  
 2 No  
 -5 Missing

*Names of the original GHS variables used to create POUTPATNT:*  
 OUTPAT and OUTPATNT

### Notes:

In 1972, children aged under 15 are excluded. In this year there were two variables for outpatient; one for adults aged 15+ (OUTPAT) and one for children aged <15 (OUTPAKID). OUTPAKID has not been included in the GHS Time Series Dataset.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS						New codes
	1972	1983	1984	1989	1994	1998	
	1982		1988	1993	1996	2004	
	<b>OUTPAT</b>	<b>OUTPATNT</b>	<b>OUTPATNT</b>	<b>OUTPATNT</b>	<b>OUTPATNT</b>	<b>OUTPATNT</b>	<b>POUTPATN</b>
YES	1	1			1	1	1
ATTD AS OUTPATNT			1	1			1
NO	2	2			2	2	2
NO OUTPATNT ATT.				2			2
NO OUTPAT ATTEND			2				2
Missing	All other missings	All other missings	All other Missings	All other missings	All other missings	All other missings	-5

### SPSS syntax to recode the variable POUTPATNT

```
*de-select missing values*.
MISSING VALUES poutpatn (.).
EXE.

RECODE
  poutpatn
    (1 thru 2=Copy) (ELSE=-5).
VARIABLE LABELS poutpatn 'Hospital outpatient attend - last 3 mths'.
VALUE LABELS poutpatn 1 'Yes' 2 'No' -5 'Missing'.
EXECUTE .
```

**PDAYPATN 'In hospital as day patient in last year'**

- 1 Yes
- 2 No
- 5 Missing

*Names of the original GHS variables used to create PDAYPATN:*  
DAYPATNT

**Notes:**

Only available from 1992 onwards.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS	New codes
	1992 2004	
	<b>DAYPATNT</b>	<b>PDAYPATN</b>
YES	1	1
NO	2	2
Missing	All other missings	-5

SPSS syntax to recode the variable PDAYPATN

\*de-select missing values\*.  
MISSING VALUES pdaypatn (.).  
EXE.

RECODE  
pdaypatn

(1 thru 2=Copy) (ELSE=-5).

VARIABLE LABELS pdaypatn 'In hospital as day patient in last year'.

VALUE LABELS pdaypatn 1 'Yes' 2 'No' -5 'Missing'.

EXECUTE .

**PINPATNT 'Hospital in patient last year'**

- 1 Yes
- 2 No
- 5 Missing

*Names of the original GHS variables used to create PINPATNT:*  
INPATIEN and INPATNT

Notes:

PINPATNT is only available from 1982 onwards.

Although a variation of this variable is available in some years of the GHS before 1982 respondents were only asked whether they had been a hospital inpatient in the last 3 months (rather than over 12 months from 1983 onwards).

From 1983, respondents were asked whether they had been to hospital as an inpatient in the past 12 months

From 2003 the question is: "During the last year, that is, since (DATE 1 YEAR AGO), have you been in hospital as an inpatient, overnight or longer?"

Value labels/names for the codes on the original GHS datasets	Codes from each year of the GHS				New codes
	1982	1983 1996	1998 2003	2004	
	<b>INPATIEN</b>	<b>INPATNT</b>	<b>INPATNT</b>	<b>INPATNT</b>	
YES	1	1	1	1	<b>PINPATNT</b> 1
NO	2	2	2	2	2
Missing	All other missings	All other missings	All other missings	All other missings	-5

```

SPSS syntax to recode the variable PINPATNT

*de-select missing values*.
MISSING VALUES pinpatnt (.).
EXE.

RECODE
  pinpatnt
    (1 thru 2=Copy) (ELSE=-5).
VARIABLE LABELS pinpatnt 'Hospital in patient last year'.
VALUE LABELS pinpatnt 1 'Yes' 2 'No' -5 'Missing' .
EXECUTE .
    
```

**PCIGSMK1 'Smoking status (ever smoked)'**

- 1 Current cigarette smoker
- 2 Ex-cigarette smoker
- 3 Never smoked
- 5 Missing

For years 1972-1988 this variable is constructed using 3 variables which formed an in between step (SPSS syntax is not provided for this step, however the details are provided below including the variable names and how the codes were recoded). See 'notes' section for further details. For years 1990 onwards this variable uses the derived variable on Smoking Status (PCIGSMK on the GHS Time Series Dataset).

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS		New codes
	1972	1984	
	1982	1988	
	SMEVER	SMOKEVER	<b>PSMOKEVE</b>
Yes	1	1	1
No	2	2	2
Missing	All other missings	All other missings	-5

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS		New codes
	1972	1988	
		CIGNOW	<b>PCIGNOW</b>
Yes	1	1	
No	2	2	
Missing	All other missings	-5	

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS		New codes
	1972	1984	
	1982	1988	
	CIGREG2	CIGEVER	<b>PCIGEVER</b>
Yes	1	1	1
No	2	2	2
Missing	All other missings	All other missings	-5

For these 3 variables the questions are not asked in 1977, 1979, 1981, 1983, 1985, 1987, 1989, 1991, 1993, 1995. The 3 variables have not been tracked from 1990 onwards as a DV on smoking status already exist.

*Names of the original GHS variables used to create PCIGSMK1:*

SMEVER, SMOKEVER, CIGNOW, CIGREG2, CIGEVER and from 1990 onwards the GHS time series variable PCIGSMK (based on the variables CIGSMKNG and CIGSMK)

Value labels/names for the codes on the original GHS datasets	codes from each year				New codes
	1972-1988			1990-2004	
	PSMOKEVE	PCIGNOW	PCIGEVER	PCIGSMK	<b>PCIGSMK1</b>
N = code as "never smoked cigarettes"	2				3
Y = code as "current cigarette smoker"		1			1
Y = code as "ex-cigarette smoker"			1		2
N = code as "never smoked cigarettes"			2		3
20+ cigs a day				1	1
10-19 cigs a day				2	1

0-9 cigs a day				3	1
10-19 cigs a day				4	1
20-29 cigs a day				5	2
30-39 cigs a day				6	3
All other values including missing values					-5

Notes:

Questions asked of those aged 16+ (except for 1972 where there are some cases for 15 year olds).

For the years 1977, 1979, 1981, 1983, 1985, 1987, 1989 there were no questions asked on smoking. These values were therefore all SYSMIS and have been converted to -5.

There is no data in 1973 and 1974 for those who completed a self-completion questionnaire. Therefore this group (self-completion) are included as missing for all years.

For 1972 to 1988 the 3 variables used to derive this variable are detailed below:

This process required the use of 3 variables as each question routed to the next question i.e.

EVER SMOKED CIGARETTE, CIGAR OR PIPE

Y = routed to next question

N = code as "never smoked cigarettes"

Missing = missing"

SMOKES CIGARETTES AT ALL NOW

Y = code as "current cigarette smoker"

N = routed to next question

Missing = missing"

EVER SMOKED CIGS.REGULARLY

Y = code as "ex-cigarette smoker"

N = code as "never smoked cigarettes"

Missing = missing"

In 1972 the variable CIGREG2 'Have you ever smoked cigarettes regularly' gives a smaller number of yes answers than CIGNOW 'Do you currently smoked cigarettes'. Therefore researchers may want to remove the year 1972 from investigations.

For more information on the variable PCIGSMK used from 1990 onwards, see metadata information for the variable PCIGSMK.

*SPSS syntax to recode the variable PCIGSMK1*

```
*de-select missing values*.
MISSING VALUES pcigever psmokeve pcignow ().
EXE.
```

```
**Recode for years 1972 to 1988/89**.
DO IF (year>=1972 and year<=1988).
+ DO IF (psmokeve=2).
+ COMPUTE pcigsmk1=3.
+ ELSE IF (pcignow=1).
+ COMPUTE pcigsmk1=1.
+ ELSE IF (pcigever=1).
+ COMPUTE pcigsmk1=2.
+ ELSE IF (pcigever=2).
+ COMPUTE pcigsmk1=3.
+ ELSE.
+ COMPUTE pcigsmk1=-5.
+ END IF.
END IF.
```

```
EXE.
```

```
**Recode for years 1990 onwards**.
```

```
DO IF (year>=1990).
```

```
+   RECODE pcigsmk (1 THRU 4=1) (5=2) (6=3) (ELSE=-5) INTO pcigsmk1.
```

```
END IF.
```

```
EXE.
```

```
RECODE pcigsmk (SYSMIS = -5) (ELSE = COPY).
```

```
EXE.
```

```
.
```

```
EXECUTE .
```

```
VARIABLE LABELS pcigsmk1 'smoking status (ever smoked)'.  
.
```

```
VALUE LABELS pcigsmk1
```

```
1 'current cigarette smoker'
```

```
2 'ex-cigarette smoker'
```

```
3 'never smoked'
```

```
-5 'missing'.
```

## PCIGSMK 'Number of cigarettes smoked per day'

- 1 20+ cigs a day
- 2 10-19 cigs a day
- 3 0-9 cigs a day
- 4 na to cigs a day
- 5 ex-cig smoker
- 6 Never smoked
- 5 Missing

For the years prior to 1990 several variables were used to create this variable. All steps are detailed below and the final table shows how they relate to the variable PCIGSMK.

Firstly a variable PCIGSADA (equivalent to CIGSADAY available on the original GHS for years 1984 onwards) relating to the average number of cigarettes smoked per day was created using 2 variables which calculate the number of cigarettes smoked at weekends, and the number smoked during week days:

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS			New codes
	1972 1976	1978	1980 1982	
	NCIGWEND	NCIGWEND	NCIGWEND	<b>PNCIGWEN</b>
NUMERIC	√	√	√	√
100 or more a day	99			
99 or more a day		99		
Less than 1 a day	0	0	0	-5
All other values including missing values				-5

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS				New codes
	1972 1974	1975 1976	1978	1980 1982	
	CIGWEEK	NCIGWEEK	NCIGWEEK	NCIGWEEK	<b>PNCIGWEE</b>
NUMERIC	√	√	√	√	√
100 or more	99	99	-	-	
99 or more	-	-	99	-	
Less than 1	0	0	0	0	-5
All other values including missing values					-5

These 2 variables have not been tracked from 1984 onwards, as a variable on the average number of cigarettes smoked per day already exists from 1984 onwards. These smoking questions were not asked in 1977, 1979, 1981, 1983.

Next the variable PCIGSADAY was created using these 2 variables for years prior to 1984, the SPSS syntax is given below:

```
DO IF (year<=1982).
+ COMPUTE PCIGSADA = ((PQTYWKDY * 5)+(PQTYWKEN * 2))/7 .
END IF.
EXECUTE .
```

The original GHS variable CIGSADAY (renamed to PCIGSADA for the GHS Time Series Dataset) available from 1984 onwards has not been tracked for all years, as a more complete DV on smoking status exists from 1990

(see the table below for more details on the DV). Note: the variable PCIGSADA is not available on the GHS Time Series Dataset.

*Names of the original GHS variables used to create PCIGSMK:*

CIGSADAY (renamed to PCIGSADA to create the variable), CIGSMKNG and CIGSMK.

Pre-1984 the variables NCIGWEND, CIGWEEK and NCIGWEEK were used to create an in between step variable PCIGSADA (see syntax above).

Pre-1990 the GHS time series variable PCIGSMK1 was also used (based on the original GHS variables SMEVER, SMOKEVER, CIGNOW, CIGREG2 and CIGEVER) along with variable PCIGNOW which created while creating the variable PCIGSMK1 (based on the original GHS variable CIGNOW).

Notes:

This variable is available on the GHS Time Series Dataset for 72, 73, 74, 75, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 00, 01, 02, 03, 04.

Questions asked of those aged 16+ (except for 1972 where there are some cases for 15 year olds).

PCIGSMK exists on the original GHS as DV (CIGSMKNG and CIGSMK) but only from 1990 onwards.

For 1972 to 1988 this variable has to be derived, details given below:

For 1972 to 1982 the variable PCIGSADA was derived from PNCIGWEE and PNCIGWEN before the final variable PCIGSMK could be derived.

$$PCIGSADA = (PNCIGWEN*2 + PNCIGWEE*5)/7$$

For 1972 to 1988 the derived variable PCIGSMK1 is used, for information on how it was derived see the metadata for PCIGSMK1.

Please note that in 1972 the variable CIGREG2 'Have you ever smoked cigarettes regularly' gives a smaller number of yes answers than CIGNOW 'Do you currently smoked cigarettes'. Therefore researchers may want to remove the year 1972 from investigations.

There is no data in 1973 and 1974 for those who completed a self-completion questionnaire. Therefore this group (self-completion) are included as missing for all years.

Note: In PCIGSMK there is an explicit category called "na to cigs a day". This is a category for those who smoke, but did not answer the number of cigarettes smoked.

Value labels/names for the codes	codes from each year							New codes			
	1972		1988	1990	1992	1994	1996		1998	2000	2004
	PCIGSADAY	PCIGNOW	PCIGSMK1	CIGSMKNG	CIGSMKNG	CIGSMKNG	CIGSMKNG		CIGSMK	CIGSMK	
20+ cigs a day	>19			1	1	1	1	1	1	1	
10-19 cigs a day	>=10 and <=19			2	2	2	2	2	2	2	
0-9 cigs a day	>=0 and <=9			3	3	3	3	3	3	3	
NA to cigs a day		1		4	4	4	4	4	4	4	
Ex-Cig smoker			2	5	5	5	5	5	5	5	
Never smoked			3	6	6	6	6	6	6	6	
All other values including missing values										-5	

SPSS syntax to recode the variable PCIGSMK

```
*de-select missing values*.
MISSING VALUES pcigsmk pcigsmk1 pcigsada pcignow().
EXE.
```

```

**Recode for years 1972 to 1988**.
DO IF (year>=1972 and year<=1988).
+   DO IF (pcigsada>19).
+       COMPUTE pcigsmk=1.
+   ELSE IF (pcigsada>=10 and pcigsada<=19).
+       COMPUTE pcigsmk=2.
+   ELSE IF (pcigsada>0 and pcigsada<=9).
+       COMPUTE pcigsmk=3.
+   ELSE IF (pcignow=1).
+       COMPUTE pcigsmk=4.
+   ELSE IF (pcigsmk1=2).
+       COMPUTE pcigsmk=5.
+   ELSE IF (pcigsmk1=3).
+       COMPUTE pcigsmk=6.
+   ELSE.
+       COMPUTE pcigsmk=-5.
+   END IF.
END IF.
EXE.

DO IF (year>=1990).
+   RECODE pcigsmk (1 thru 6=Copy)(ELSE=-5) .
END IF.

RECODE pcigsmk (SYSMIS = -5) (ELSE = COPY).
EXE.

VARIABLE LABELS pcigsmk 'Number of cigarettes smoked per day'.
VALUE LABELS pcigsmk
    1      '20+ cigs a day'
    2      '10-19 cigs a day'
    3      '0-9 cigs a day'
    4      'na to cigs a day'
    5      'ex-cig smoker'
    6      'Never smoked'
   -5     'Missing'.
EXECUTE .

```

## **PNOSMOKE 'Go without smoking for whole day'**

- 1 Very easy
- 2 Fairly easy
- 3 Fairly difficult
- 4 Very difficult
- 5 Missing

*Name of the original GHS variables used to create PNOSMOKE: NOSMOKE*

### **Notes:**

Available in 1992, 1994 and from 1996 to 2004.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS	New codes
	1992	<b>PNOSMOKE</b>
	2004	
	NOSMOKE	
Very easy	1	<b>1</b>
Fairly easy	2	<b>2</b>
Fairly difficult	3	<b>3</b>
Very difficult	4	<b>4</b>
Missing	All other missings	<b>-5</b>

### *SPSS syntax to recode the variable PNOSMOKE*

```
*de-select missing values*.
MISSING VALUES pnosmoke (.).
EXE.
```

```
RECODE
```

```
    pnosmoke
```

```
    (1 thru 4=Copy) (ELSE=-5).
```

```
VARIABLE LABELS pnosmoke 'Go without smoking for whole day'.
```

```
VALUE LABELS pnosmoke 1 'Very easy' 2 'Fairly easy' 3 'Fairly difficult' 4 'Very difficult' -5 'Missing'.
```

```
EXECUTE .
```

## PGIVEUP 'Like to give up smoking altogether'

- 1 Yes
- 2 No
- 3 Don't know
- 5 Missing

Name of the original GHS variables used to create PGIVEUP: GIVEUP

### Notes:

Available in 1992, 1994 and from 1996 to 2004.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS			New codes
	1992	1994 1998	2000 2004	
	GIVEUP	GIVEUP	GIVEUP	<b>PGIVEUP</b>
Yes	1	1	1	<b>1</b>
No	2	2	2	<b>2</b>
Don't Know	3		3	<b>3</b>
Missing	All other missings	All other missings	All other missings	<b>-5</b>

### SPSS syntax to recode the variable PGIVEUP

```
*de-select missing values*.
MISSING VALUES pgiveup (.).
EXE.
```

```
RECODE
  pgiveup
    (1 thru 3=Copy) (ELSE=-5).
VARIABLE LABELS pgiveup 'Like to give up smoking altogether'.
VALUE LABELS pgiveup 1 'Yes' 2 'No' 3 'Dont know' -5 'Missing'.
EXECUTE .
```

### **PFIRSTCI** 'When first cigarette of day'

- 1 Less than 5 minutes
- 2 5-14 minutes
- 3 15-29 minutes
- 4 30 minutes but less than 1 hour
- 5 1 hour but less than 2 hours
- 6 2 hours or more
- 5 Missing

Name of the original GHS variables used to create **PFIRSTCI**: FIRSTCIG

#### **Notes:**

Available in 1992, 1994 and from 1996 to 2004.

Value labels/names for the codes on the original GHS datasets	codes from each year of the GHS	New codes
	1992	
	2004	
	FIRSTCIG	<b>PFIRSTCI</b>
Less than 5 minutes	1	<b>1</b>
5-14 minutes	2	<b>2</b>
15-29 minutes	3	<b>3</b>
30minutes but less than 1 hour	4	<b>4</b>
1 hour but less than 2 hours	5	<b>5</b>
2 hours or more	6	<b>6</b>
Missing	All other missings	<b>-5</b>

#### SPSS syntax to recode the variable **PFIRSTCI**

```
*de-select missing values*.
MISSING VALUES pfirstci (.).
EXE.
```

```
RECODE
  pfirstci
```

```
(1 thru 6=Copy) (ELSE=-5).
```

```
VARIABLE LABELS pfirstci 'When first cigarette of day'.
```

```
VALUE LABELS pfirstci 1 'Less than 5 minutes' 2 '5-14 minutes' 3 '15-29 minutes' 4 '30 minutes but less than 1 hour' 5 '1 hour but less than 2 hours' 6 '2 hours or more' -5 'Missing'.
```

```
EXECUTE .
```

## Using the GHS Time Series Dataset for pseudo-cohort analysis

Note: this part of the User Guide is reproduced from the ONS publication [The GHSPCD: Introduction and Methodology](#)<sup>xviii</sup> (by Zoë Uren, 2006). Originally the GHS Time Series Dataset was referred to as the GHS Pseudo-Cohort Dataset (GHSPCD).

The annual feature of the GHS makes it possible to generate pseudo-cohorts. Rather than track individuals (as in longitudinal studies), the pseudo-cohort tracks the average experiences of persons belonging to a particular birth cohort. Although the GHS Time Series Dataset does not track individual experiences, it does however have the advantage of no sample attrition due to non-participation and mortality.

The main objectives of creating the GHS Time Series Dataset were to analyse trends in social inequalities over the last 30 years, and to analyse these inequalities using pseudo-cohort analysis. While this approach is not new, and other researchers have combined GHS data series, previous analyses have been limited to specific areas of research or time periods<sup>xix</sup>.

To give an idea of the sizes of the cohorts involved these are given below:

5-year estimated birth cohort			
1871-1875	8	1941-1945	50,834
1876-1880	69	1946-1950	60,478
1881-1885	417	1951-1955	53,332
1886-1890	1,614	1956-1960	58,207
1891-1895	4,250	1961-1965	66,439
1896-1900	8,906	1966-1970	65,460
1901-1905	16,330	1971-1975	54,010
1906-1910	24,720	1976-1980	37,621
1911-1915	32,090	1981-1985	30,394
1916-1920	34,797	1986-1990	24,104
1921-1925	43,877	1991-1995	15,343
1926-1930	43,899	1996-2000	7,981
1931-1935	43,559	2001-2005	2,575
1936-1940	46,187		

### What is pseudo-cohort analysis?

Pseudo-cohort analyses track the average experiences of people who were born in the same time period. Unlike conventional longitudinal analyses, individuals surveyed are not the same from year to year. As the GHS is a cross-sectional survey, the experiences of respondents' in a given survey year are taken to represent the average experiences of their particular birth cohort.

For example, those aged 20 to 24 in the 1980 GHS survey, are taken as being representative of the birth cohort born in 1956 to 1960. The same birth cohort is then represented by those aged 21 to 25 in 1981, and 22 to 26 in 1982, and finally aged 44 to 48 in 2004. This constitutes a pseudo-cohort, and because the pseudo-cohort does not contain the same individuals from one year to the next, it measures the average experience for a given cohort.

The GHS Time Series Dataset can be used to explore differences between birth cohorts for outcomes such as self-reported health, unemployment, qualifications and heavy smoking, to name but a few. Pseudo-cohorts can be further defined in terms of sex, social class, and other social characteristics and the differences in experiences between these groups can be examined.

People in different birth cohorts would have grown-up in different economic, technological and social climates. Those in the same birth cohort share similar experiences in terms of:

- Availability of material items e.g. consumer durables, cars, new technology
- Access to services e.g. health, education
- Attitudes and social acceptability
- Historical and political experiences

These factors can be hugely influential and can impact on outcomes via peoples' experiences and the life choices they make. For example teenagers growing up in the 1960s would have had very different experiences to those in the 1940s, or teenagers of today.

The GHS does however only cover private households in Great Britain. Therefore it is not representative of the whole population as it excludes those individuals that live in communal establishments such as care homes, hospitals and prisons. The experiences and social outcomes can differ considerably between these two populations. This is particularly pertinent when looking at the experiences of older people. The older people captured by the GHS may not be representative of this age group as the characteristics of those older people who remain within a household may be different to those of older people that reside in communal establishments. This can lead to problems when analysing the health of older people, as less healthy people in residential care would be excluded from the sample.

**Box 1** Adapted from *The GHSPCD: Introduction and Methodology* ([Uren 2006](#))<sup>xx</sup>

#### **Benefits of the GHS Time Series Dataset to the researcher**

##### *1. The GHS is a highly valued survey*

The General Household Survey (GHS) is a highly regarded government survey, and is considered a good quality source of data. The survey is conducted to provide a representative picture the population of Great Britain. The GHS Time Series Dataset adopts certain benefits from the original survey design. For example the survey is nationally representative, has a large sample size and has a good response rate over time.

##### *2. Facilitates the analysis of time trends*

The GHS provides a good data source for analysing changes in society since the 1970s both because of the long time that the survey has been running, as well as its large sample size. However, using the GHS to analyse trends over time can be both time consuming and complex. This is because GHS data is presently only available in individual annual datasets, and separate extracts need to be taken for each year of interest. The GHS Time Series Dataset presents a large number of variables in a consistent format over time, facilitating time series analysis.

##### *3. Provides a research tool to measure the interactions between social outcomes*

There are various types of inequalities that can exist between different groups in society such as life expectancy, educational attainment, access to services, and the likelihood of certain behaviours such as poor diet and smoking. These inequalities rarely exist alone and there is considerable interaction between them.

As the GHS covers a vast range of topics, including demographics, living standards, employment, education and health, the relationships between these issues can be explored.

*4. Enables research of cohort effects on outcomes using pseudo-cohort analysis*

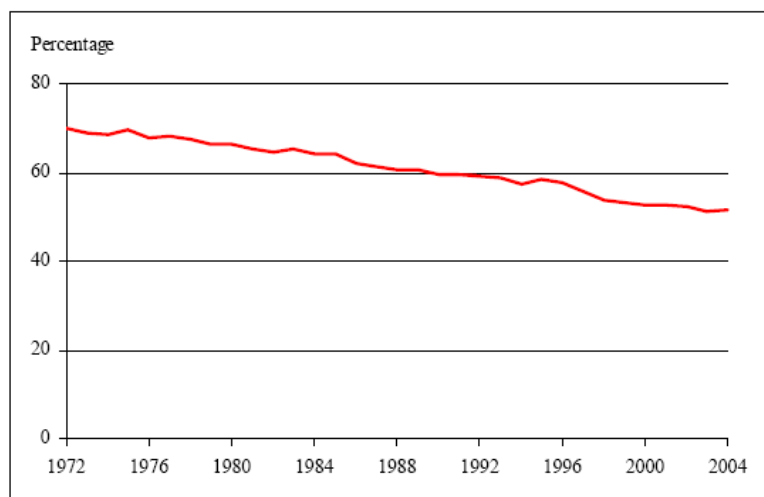
Longitudinal data enables the researcher to track individuals' experiences and changing circumstances. These data sources are the preferred approach to understanding causal pathway factors influencing individual life courses. However at present there is no single source of longitudinal data in Britain that provides demographic and socio-economic information both across an extended time period and for a range of different birth cohorts. The British Household Panel Survey (BHPS) at the University of Essex is restricted in that data collection only started in 1991. The ONS Longitudinal Study (LS) covers the period since 1971, but the range of socio-economic data is limited. It uses census data which is only collected decennially. The British Birth Cohort Studies (1970 and 1946) only collect data for these specific cohorts.

### Examples of pseudo-cohort analysis using the GHS Time Series Dataset<sup>xxi</sup>

Pseudo-cohort analysis can be used to distinguish between the effects of period, age and birth cohort on an outcome. Using the example of trends in marriage since the early 1970s, it is possible to demonstrate this. Marriage trends have been used here to provide a simple and well known example in which to demonstrate how pseudo-cohort analysis works rather than getting into more complex research questions.

Since the early 1970s the number of people getting and remaining married has decreased<sup>xxii</sup>. Results from the GHS Time Series Dataset confirm this. Figure 1 shows that the percentage of people aged over 16 that are married has fallen from 70% to 52% between 1972 to 2004. To understand the reasons for this decline over time it is necessary to explore these changes by age and birth cohort.

Figure 1 Percentage of married people, 1972-2004

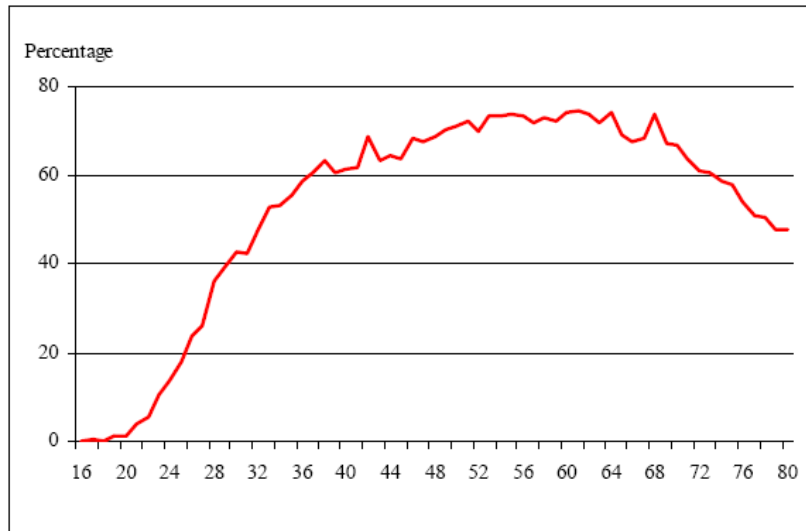


Source: General Household Survey 1972 to 2004/5.

Figure 2 examines the percentage of people that were married by single year of age in 2003. For those people aged under 25 in 2003, very few were married. The proportion increased among those in their 30s, and for people aged 40 in 2003, approximately 60% were married. The

proportions declined with older ages, representing an increase in the numbers that became widowed.

**Figure 2 Percentage of married people by single year of age, 2003 (3-year moving averages)**



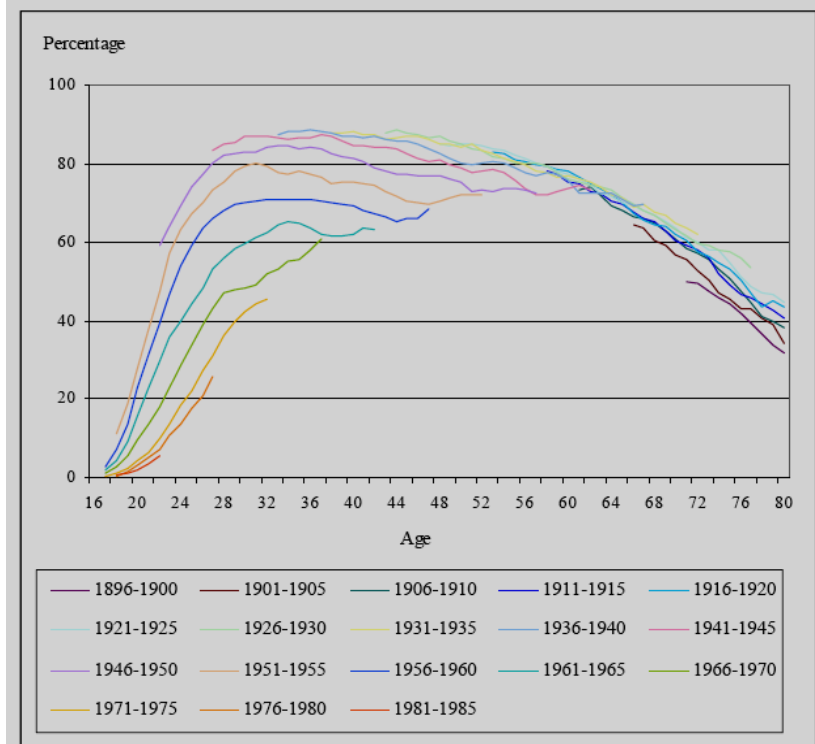
Source: General Household Survey 2002/3 to 2004/5.

The two preceding graphs show how overall marriage patterns have changed over time and vary by age. In summary, marriage rates have declined over time and older people are more likely to be married than younger people in 2003.

However it is not possible to understand the changes in marriage patterns from exploring relationships over time and by age in isolation. It is indeterminable to what extent the decline in marriage rates is due to changes in the age structure of the population over time or the behaviour of each subsequent generation? Cohort analysis allows the researcher to examine the interaction between time and age.

While the effects of period and age have been illustrated, marriage patterns have also varied between different birth cohorts over the last 30 years. Figure 3 shows the percentage of married people by age for each five-year birth cohort. The percentages are calculated using three-year moving averages (calculations are given in appendix 2). To compensate for the effect of small numbers, the analyses are only presented up to age 80 and exclude any percentages with a denominator of less than 50.

**Figure 3 Percentage of married people by birth cohort**



Source: General Household Survey 1972 to 2004/5, using 3 year moving averages.

#### *How to interpret the graph*

Each line on the graph represents a birth cohort, and maps the relationship between age and the outcome of interest. In this example the percentage of married people is shown on the y-axis and single year of age on the x-axis. Some of the cohorts are incomplete, as information is not available for every age in each cohort. This is because the GHS is cross-sectional and only includes data from 1972 to 2004. For example, people born between 1901 and 1905 were aged 67 to 71 in 1972 when the survey started.

Using the example of the 1956-60 birth cohort, the proportion of this cohort that was married increased substantially when aged in their 20s, and peaked in their early 30s. By the end of their 30s, the percentage of this cohort that were married began to decline, probably as respondents became separated and divorced. Each cohort can be analysed individually to understand how, on average, marriage patterns have changed by age.

Comparing all cohorts shown in the graph, it is possible to infer that:

- i. Differences between cohorts exist at younger generations (cohort effect).*
- ii. No differences between cohorts exist at older generations (no cohort effect).*

#### *i. Cohort effect at younger generations*

Cohort effects for marriage patterns are more pronounced for younger generations than older generations as shown by divergent plots in figure 3. Among younger generations (i.e. those born after the Second World War), we can see their patterns of marriage at age 25 in the table below. For each successive generation, smaller proportions were married at this age.

**Table 1 Birth Cohort by Percentage married at age 25**

Birth Cohort	Percentage married at age 25
1946-50	73
1951-55	68*
1956-60	61*
1961-65	42*
1966-70	35*
1971-75	24*
1976-80	18*

\* Significantly different to the 1946-50 birth cohort

*ii. No cohort effect at older generations*

For older generations (i.e. those born before the start of the Second World War) there are no cohort effects as shown by the converged lines in figure 3. It is not possible to compare their marriage patterns at age 25 because the GHS was not running when they were at this age. However, we can compare marriage patterns at age 55.

**Table 2 Birth Cohort by Percentage married at age 55**

Birth Cohort	Percentage married at age 55
1916-1920	82
1921-1925	83
1926-1930	81
1931-1935	82
1936-1940	79
1941-1945	76**
1946-1950	74**

\*\* Significantly different to the 1916-20 birth cohort

We can see that in each of the older cohorts, approximately the same proportion was married at this age. However, there is a significant difference between the proportion of people who were married at this age in the post-war cohort and the earlier born cohorts.

*Discussion*

Pseudo-cohort analysis is not straightforward, either to present or to interpret. For any one graph, up to twenty-one cohorts can be illustrated at a time. From this the any user needs to understand how the outcome differs by age for each cohort, as well as, infer what differences occur between cohorts. This is further complicated if comparing between different groups, such as by sex or social class.

Further, despite the large sample size of the dataset, when disaggregating the data by birth cohort and single year of age the numbers are considerably reduced. As a consequence it is not possible to run detailed analysis for very specific groups.

*General Limitations of pseudo-cohort analysis*

Recall bias occurs when different groups of respondents remember (or tell us) things differently. For example, people may not always remember what qualifications they have, or they may not know how their qualifications equate to modern qualifications. Therefore it could be invalid

to compare qualifications between older people and the rest of the respondents.

In addition, since pseudo-cohort analysis represents the average experiences of people belonging to a particular cohort, and does not track the experiences of actual people, it is not possible to determine the duration of a particular event. For example, it is not possible to know how long a person has been married or for how long an individual had smoked before giving up.

*Conclusions on using the GHS Time Series Dataset*

Limitations aside, the GHS Time Series Dataset will provide a powerful analysis tool. It will enable researchers to analyse key socio-economic trends over the past thirty years on an (almost) annual basis. The dataset can be used to measure interactions between social outcomes and subsequently analyse the cohort effects on these outcomes. Additionally, the General Household Survey is a highly valued survey and the GHS Time Series Dataset adopts many of the benefits from the original survey.

## Appendix 1

### Further information on the Scottish supplement

From 1971 the sample of addresses selected in Scotland was doubled so as to improve estimates in separate analyses for Scotland. The interview on the supplementary sample was just the household questionnaire with one or two minor changes. The supplementary sample was discontinued in 1994. Although the 1994-1996 datasets contain the variable 'GB', frequencies show that all values for this are either 'missing' or systematic.

Following the consultation exercise, and the subsequent advice received, the Scottish supplement was removed from the final dataset. Details of how the Scottish supplement was removed from the GHS Time Series Dataset for different years are given below. In general, where both gb and region are available, the variable gb is used.

#### 1972, 1983 -1993

GB variable was used to exclude the Scottish Supplement by selecting those where the variable 'gb' = 1 (main sample) or -9 (missing).

#### 1973

There were about 7000 cases for Scotland – about 20% of the sample. Comparing this to 1974 (where there is an indicator of whether the Scottish supplement exists) there were 3000 cases for Scotland and about 3000 extra Scottish supplements. Therefore 1973 will contain a Scottish supplement, but it cannot be distinguished and have to therefore weight the Scottish cases in 1973 by 0.5.

#### 1974, 1977, 1980-81

'Region' (region) variable was used to exclude the Scottish supplement, as it had a value for the Scottish Supplement

#### 1975-76

The variable 'region' does not have a label attached which has the title "Scottish supplement" although all other categories are labelled. It is presumed that the category "12" of 'region' is the Scottish supplement (as this was the case in 1974, where the variable was labelled) Therefore these cases were excluded from the dataset.

#### 1978 & 1979 & 1982

There was no Scottish supplement on these years. They may already have been excluded from the dataset that was at the Essex Archive. There was one case in 1978 where the variable 'region' (which has been renamed in the GHSPCD to 'pcountry') has the value of 13 i.e. "Scottish Supplement". It was decided to retain this case so the data matched the 1978 data and it is assumed this is an error in the original 1978 data.

1994 onwards - Scottish Supplement was not asked.

## Appendix 2

### Three-year moving averages

Despite the large sample size of the GHS, when disaggregating the data into different groups or outcomes, the sample size becomes too small and there is too much noise in the analyses. To resolve this three-year moving averages, using the combined method, have been used.

$$\bar{y}_c = \frac{\hat{Y}_1 + \hat{Y}_2 + \hat{Y}_3}{n_1 + n_2 + n_3}$$

where the  $\hat{Y}$  are the sample counts of those with a characteristic and the  $n$  are the sample sizes.

All pseudo-cohort analyses presented in the examples are based on three-year moving averages.

For a single cohort the outcome at a particular age is based on the average experience of those people either side of that age i.e. the outcome at age 17, is the average experience of those aged 16, 17 and 18. For those aged 18 it is the average experience of those aged 17, 18 and 19, and so on.

- 
- i <http://www.esds.ac.uk/government/>
- ii <http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=5756&Pos=2&ColRank=1&Rank=272>
- iii <http://www.statistics.gov.uk/cci/article.asp?ID=1637&Pos=1&ColRank=1&Rank=1>
- iv <http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=5756&Pos=2&ColRank=1&Rank=272>
- v <http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=5756&Pos=2&ColRank=1&Rank=272>
- vi <http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=5756&Pos=2&ColRank=1&Rank=272>
- vii <http://www.statistics.gov.uk/StatBase/Source.asp?vlnk=263&More=Y#general>
- viii <http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=5756&Pos=2&ColRank=1&Rank=272>
- ix <http://www.statistics.gov.uk/cci/article.asp?ID=1637&Pos=1&ColRank=1&Rank=1>
- x <http://www.statistics.gov.uk/LIB2002/default.asp>
- xi <http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=5756&Pos=2&ColRank=1&Rank=272>
- xii <http://www.statistics.gov.uk/LIB2002/default.asp>
- xiii <http://www.esds.ac.uk/government/dv/cob/>
- xiv [www.statistics.gov.uk/lib2000/Section243.html](http://www.statistics.gov.uk/lib2000/Section243.html)
- xv [http://www.statistics.gov.uk/methods\\_quality/ns\\_sec/default.asp](http://www.statistics.gov.uk/methods_quality/ns_sec/default.asp)
- xvi [http://www.statistics.gov.uk/methods\\_quality/ns\\_sec/default.asp](http://www.statistics.gov.uk/methods_quality/ns_sec/default.asp)
- xvii <http://www.statistics.gov.uk/statbase/Product.asp?vlnk=2054>

---

<sup>xviii</sup> <http://www.statistics.gov.uk/cci/article.asp?ID=1637&Pos=1&ColRank=1&Rank=1>

<sup>xix</sup>

References:

1. Evandrou, M. and Falkingham, J (2000) 'Looking back to look forward: lessons from four birth cohorts for ageing in the 21st century'. *Population Trends*, 99, Spring 2000 pp27-36.
  2. Evandrou, M and Falkingham, J (2001) 'Smoking behaviour and socio-economic status: a cohort analysis, 1974-1998'. *Health Statistics Quarterly*, 14, Summer 2002 pp30-39.
  3. Dawson, A. and G. Evans, Pensioners' incomes and expenditure 1970-1985. *Employment Gazette*, 1987. 95(5): p. 243-252.
  4. Hancock, R. and P. Weir, Pensioners' income in Great Britain, 1979-1989: detailed analyses of the General Household Survey and Family Expenditure Surveys. Working Paper 7, ACIOG: London, 1994.
  5. Evandrou, M. and K. Glaser, Changing economic and social roles: the experience of four cohorts of mid-life individuals in Britain, 1985-2000. *Population Trends*, 2002(110): p. 19-30.
  6. Kemm, J.R., A birth cohort analysis of smoking by adults in Great Britain 1974-1998. *Journal of Public Health Medicine*, 2001. 23(4): p. 306-311.
  7. Kemm, J., An analysis by birth cohort of alcohol consumption by adults in Great Britain 1978-1998. *Alcohol and Alcoholism*, 2003. 38(2): p. 142-147.
  8. McIntosh, S (2005) 'Using Pseudo Cohorts to Track Changes in the Qualifications of National Populations', Research report No 621, Department for Education and Skills.
- <sup>xx</sup> <http://www.statistics.gov.uk/cci/article.asp?ID=1637&Pos=1&ColRank=1&Rank=1>
- <sup>xxi</sup> The examples are taken from the methodology paper in [The GHSPCD: Introduction and Methodology](#) (Uren 2006).
- <sup>xxii</sup> Babb P *et al.* (eds) (2006) Chapter 2: Households and Families, in *Social Trends*, No 36