ISSN 2042-4116

ESRC Centre for Population Change • Working Paper 47 • April 2014

CPC

centre for population change

User guide to the Centre for Population Change GHS database 1979 - 2009

Éva Beaujouan
Ann Berrington
Mark Lyons-Amos
Máire Ní Bhrolcháin

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

AUTHOR CONTRIBUTIONS

Máire Ní Bhrolcháin originated the proposal to create a time-series database of General Household Survey demographic histories from the 1970s to the present and was Principal Investigator on the project to create the data file. Éva Beaujouan assembled the database, with assistance from Mark Lyons-Amos, under the direction of Máire Ní Bhrolcháin and Ann Berrington. All authors have contributed to the compilation of this User Guide but Éva Beaujouan is its principal author.

Correspondence to cpc@soton.ac.uk

AVAILABILITY OF THE DATABASE

The database is available under Special Licence from the UK Data Archive. Access to the data can be requested online at: http://www.data-archive.ac.uk.

HOW TO CITE THIS USER GUIDE

The User Guide should be cited as follows:

Beaujouan, É., Berrington, A., Lyons-Amos, M. and Ní Bhrolcháin, M. (2014) *User Guide to the Centre for Population Change GHS database 1979-2009*. Centre for Population Change Working Paper 47, ESRC Centre for Population Change, UK. http://www.cpc.ac.uk/publications/cpc_working_papers/pdf/2014_WP47_CPC_GHS_User_Guide_Beaujouan_et_al.pdf

The ESRC Centre for Population Change Working Paper Series is edited by Teresa McGowan

ESRC Centre for Population Change

The ESRC Centre for Population Change (CPC) is a joint initiative between the Universities of Southampton, St Andrews, Edinburgh, Stirling, Strathclyde, in partnership with the Office for National Statistics (ONS) and the National Records of Scotland (NRS). The Centre is funded by the Economic and Social Research Council (ESRC) grant numbers RES-625-28-0001 and ES/K007394/1.

This working paper series publishes independent research, not always funded through the Centre. The views and opinions expressed by authors do not necessarily reflect those of the CPC, ESRC, ONS or NRS.

Website | Email | Twitter | Facebook | Mendeley

USER GUIDE TO THE CENTRE FOR POPULATION CHANGE GHS DATABASE 1979-2009

TABLE OF CONTENTS

1. INTRODUCTION

This is the User Guide for the ESRC Centre for Population Change (CPC) General Household Survey (GHS) time-series database covering the years 1979 to 2009. The key purpose of this database is to provide a consistent time series of demographic data collected in the GHS, particularly childbearing histories and marriage and cohabitation histories.

In 2005, the base for data collection switched from financial year to calendar year. Data for the last quarter of 2004-5 are included in the ONS datafiles for both 2004-05 and 2005 GHS rounds. Since our database includes both survey years, we did not retain this duplication and the last quarter of 2004-5 is allocated to the 2004-05 round only.

From 2005 on, the survey was restructured as a panel rather than as a cross-sectional survey. Three quarters of respondents interviewed in one year are re-interviewed the following year. Therefore, due to the rotating panel only a quarter of the sample in 2006-2009 are new respondents. New respondents in 2006-2009 are identified by code '1' on WAVE. Respondents in waves 2-4 in 2006-2009 are omitted from this database.

Note that from GHS round 1998 onwards, dates of birth of respondents and dates of live births are available only in the special licence version of the GHS datasets deposited by ONS with the Data Archive.

1.1 USING THE CPC GHS TIME-SERIES DATASET

Along with the introductory notes and information on specific variables presented in this User Guide, prospective users of the CPC GHS time series dataset are advised to consult the following documents and papers for essential information on the structure of the dataset and quality of the data on specific topics:

1.1.1 SAMPLE SIZES

Annex H Sample sizes in the CPC time-series dataset with notes of some errors in the original GHS data files

1.1.2 WEIGHTING

Beaujouan, É., J. Brown and M. Ní Bhrolcháin (2011) Reweighting the General Household Survey 1979-2007. *Population Trends* 145: 119-145. http://www.ons.gov.uk/ons/rel/population-trends-rd/population-trends/no--145--autumn-2011/ard-pt145-reweighting-ghs.pdf

1.1.3 FERTILITY HISTORIES

Annex A Quality checks and edits carried out on birth histories

Ní Bhrolcháin, M., E. Beaujouan, and M. Murphy (2011) Sources of error in reported childlessness in a continuous British household survey. *Population Studies* 65:3, 305-318. http://www.tandfonline.com/doi/full/10.1080/00324728.2011.607901

1.1.4 MARRIAGE AND PARTNERSHIP HISTORIES

Annex B Quality checks and edits carried out on marriage histories 1979-1998

Annex C Quality checks and edits carried out on partnership histories from 2000-2007

Annex D Derivation of the quality flags for the marriage and partnership histories

Berrington, A., É. Beaujouan, M. Lyons-Amos and M. Ní Bhrolcháin (2011) Evaluation of the Partnership Histories in the Centre for Population Change GHS Time Series Dataset. CPC Working Paper 12, September 2011, Centre for Population Change, University of Southampton http://www.cpc.ac.uk/publications/2011_WP12_Evaluation_of_the_Partnership_Histories_Berrington_et_al.pdf

1.1.5 VARIABLES CARRIED OVER FROM THE 1972-2004 GHS TIME SERIES DATASET

User Guide to the 1972-2004 GHS Time Series dataset (UK Data Archive Study Number 5664).

1.1.6 MERGING WITH THE ORIGINAL GHS DATA FILES

To merge records between the CPC GHS database and the original GHS data files provided by the UK Data Archive, use, in combination, the CPC variables **year**,

hhserial, and **pno**. See the Data Dictionary for the corresponding GHS variable names in various years.

Note that 2004-05 was the last year in which the survey was fielded on a financial-year basis, and the original dataset deposited with the ESDS Data Archive for this year includes the first quarter of 2005. In 2005, the survey changed back to a calendar year basis, and a full year's sample size was collected in quarters 2, 3, and 4 of 2005. In the ESDS version of the 2005 GHS, the last quarter of 2004-5 is duplicated as the first quarter of 2005. Because the first quarter of 2005 is already present in the 2004-5 dataset, we have excluded it from the 2005 data in our time series dataset. The first quarter of 2005 can be identified in the 2004-5 dataset via **sampq**=4. Therefore to extract the original 2004-5 round data from the CPC dataset, choose year = 2004, and **sampq** = 1 or 2 or 3 or 4

To extract the data deposited as the 2005 round with the Archive, choose

year = 2004 and **sampq** = 4 AND

year = 2005 and **sampq** = 2 or 3 or 4

Note also that the household serial IDs in the original release of the 2005 GHS were changed in a subsequent release of the 2005 data file. The present version of the CPC database includes the revised 2005 household identifiers. If merging this database with the GHS datasets, you should check first that the household identifiers in your 2005 GHS data file are as in the revised data file.

1.2 STRUCTURE OF THE USER GUIDE

Documentation for each variable in the data dictionary starts on a new page. Where variables form part of a set — e.g. **bthyr1-bthyr15**—they appear on one page rather than on several. Some natural sets of variables – e.g. weights, and counts of persons in household – are grouped together (see variable list classified by theme).

The entry for each variable gives the CPC variable name and labels at the top of the page, shows which original GHS variables were used in the construction of the CPC variable and how the codes from the original variables were used to construct the CPC variable. In many cases, the variables used to derive new or consistent variables for the CPC database differ either in name or coding, or both, across GHS years.

1.2.1 1972-2004 GHS TIME-SERIES DATASET

Some of the variables in this database are carried over from the 1972-2004 GHS Time Series dataset for the years 1979-2004, with the variables for 2005-09 being constructed in the same way.

Note that for all the variables carried over from the 1972-2004 time-series, there was only one type of missing value (-5), and this coding was retained as missing for those variables in GHS rounds 2005 and after.

1.2.2 GHS ROUNDS 2008-2009

The database was originally assembled from GHS rounds 1979-2007 and data from GHS rounds 2008 and 2009 were added at a later stage.

- While the data for 2008 and 2009 are consistent with preceding years, the primary documentation of variables refers in most cases to the years 1979-2007. The coding for 2008 and 2009 is consistent with that for 2005-2007.
- Quality checks on the fertility histories, marriage histories and partnership histories were carried out for rounds 1979-2007 only.
- Quality flags have been generated for the 2008 and 2009 GHS rounds, but the results have been scrutinised much less closely than those of preceding years, and 2008-09 data are not included in Annexes A-C.

1.2.3 INCONSISTENCIES IN VARIABLES OVER TIME

In cases where there is a substantive change in the question asked in the GHS which would result in an inconsistency in the interpretation of a variable we have adopted a double digit system of coding.

1.2.4 CHANGING ELIGIBILITY

Some variables are applicable only to certain respondents. Much of the detailed childbearing and partnership history data come from the Family Information Section (FI) of the GHS questionnaire. Children and older persons are not eligible for this section, and men are not asked about their childbearing histories at all, and have been asked about their marriage histories only since 1986. The eligibility criterion for the FI has changed over the years and a summary of these changes is given on the **eligfi** page. See also Annex F and associated Excel file.

1.2.5 PARTNERSHIP

A "partnership" in this documentation refers to one of the following:

a direct marriage

a cohabitation followed by marriage to that partner, or

a free-standing cohabitation.

From 2000 on, the history of such partnerships is near complete. Prior to 2000 the retrospective information collected on cohabitation was more limited; questions were asked about the current partnership which may be a cohabitating union, and also on whether a marriage was preceded by a spell of cohabitation (see Table 1 in Annex B).

1.2.6 MISSING VALUE CODES

For variables generated specifically for this database, a consistent set of missing values is used:

-8: not available/not known

-9: does not apply

However, some variables have been incorporated from the 1972-2004 time series database, and such variables have just one missing value: -5.

While the codes in the original GHS files that are used to arrive at a final coding of the variables in this database are usually separately listed on the page for each variable, in some cases a verbal summary is given in place of the exact codes. This

occurs because relevant values differ substantially across years. This is the case for some variables in relation to missing values, and so we simply summarise verbally by recording “missing values for NA” or “missing values for DNA” or where these are not distinguished “missing codes” to cover the variety of missing codes used (because prior to 1998, NA and DNA codes were not distinguished in the original GHS files).
For all Family Information section variables, in the present database respondents are coded DNA (-9) if they are not eligible for the FI section, or that they are a proxy respondent. When the respondent is eligible but the information is missing the variable is coded NA (-8).

1.2.7 IMPUTATION

Only two kinds of imputation are made in this dataset. One is of the month of an event, where the year was known but the month missing. The imputation was flagged by adding 0.5 to the month. See Annexes A-C. The other imputation is the calculation of the respondent’s date of birth based on age at interview and interview date for GHS rounds 1979 to 1982. In the FI section, for survey years where dates of birth of the individuals were missing (1979-1981) or partly missing, these were calculated: using the month when it was known; attributing a month and deducing a year when necessary. Additional minor corrections have been made, described in Annex H.

1.3 ANNEXES

A. Quality checks and edits carried out on birth histories
B. Quality checks and edits carried out on marriage histories to 1998
C. Quality checks and edits carried out on partnership histories from 2000-2007
D. Derivation of the quality flags for the marriage and partnership histories
E. SPSS coding of **defacto**
F. Eligibility for the specific questions in the Family Information section
G. Quality checks on age at leaving full time education 1979-2007
H. Sample sizes in the CPC time-series dataset with notes of some errors in the original GHS data files
I. Key references

2. THEMATIC LIST OF VARIABLES IN CPC GHS DATABASE

Theme	Variable Name	Location in data dictionary for grouped variables
Survey		
intmt	Month at interview	
intyr	Year at interview	
sampq	Sample quarter	
sched	Full or proxy interview	
wave	Interview wave	
weightn	Original ONS weight	
pcountry	UK country	
year	GHS survey round	
hhserial	Household serial number	
funo	Family unit number	
pno	Person number of the respondent in the household	
selfcomr	Self-completion of the FI section	
wgtcpc	Weights for all individuals in the CPC time-series	
wgtcpcstd	Weights for all individuals in the CPC time-series, normalised	
wgtcpcfi	Weights for individuals answering the Family Information section in the CPC time-series	
wgtcpcfistd	Weights for individuals answering the Family Information section in the CPC time-series, normalised and trimmed at 3	
Household		
acctyp91	'Type of accommodation'	
acctyp01	'Type of accommodation'	
ageldchh	Age of the oldest child in the family unit	Persons in household: Count variables
ageldhh	Age of the eldest person in the household	Persons in household: Count variables
ageygcfu	Age of the youngest child in the family unit	Persons in household: Count variables

ageygchh	Age of the youngest child in the household	Persons in household: Count variables
agygsthh	Age of the youngest person in the household	Persons in household: Count variables
funo	Family unit number	
hhserial	Household serial number	
nadul60	Number of adults aged 60+ in household	Persons in household: Count variables
nadul65	Number of adults aged 65+ in household	Persons in household: Count variables
nadulmen	Number of adult men in the household	Persons in household: Count variables
nadulwom	Number of adult women in the household	Persons in household: Count variables
nchdhh	Number of children in household	Persons in household: Count variables
nfuhh	Number of family units in the household	Persons in household: Count variables
npenshh	Number of pensioners in household	Persons in household: Count variables
npershh	Number of persons in the household	Persons in household: Count variables
npersfu	Number of persons in family unit	Persons in household: Count variables
pseghrp	Socio-economic group of HOH/HRP	
Individual		
age	Age of the respondent	
agelft	Age left full-time education	
brthmt	Month of birth of respondent	
brthyr	Year of birth of respondent	
chdbenef	Whether receives child benefit	
contrapt	Whether currently using contraception	

curtype	Type of current partnership (at survey)	
defacto	Living with a partner/spouse	
degree	Highest qualification = degree or higher	
ethgrp2	Ethnic group	
eligfi	Eligible for Family Information Section	
funo	Family unit number	
lastft	Where last attended course	
marst	Declared marital status	
pno	Person number of the respondent in the household	
morechld	Whether intends to have more children	
moreposs	Could you (and your partner) have more children if you wanted to	
nextage	If intends to have more children, age at birth of next child	
npersfu	Number of persons in family unit	Persons in household: Count variables
pcigsmk	Number of cigarettes smoked per day	
pcigsmk1	Smoking status (ever smoked)	
pcob1	Respondent's country of birth	
pcountry	UK country	
pcutdown	Illness / injury reduce activity	
pdvilo3a	Activity status	
pedfull	Education level (full form)	
pfcob1	Father's country of birth	
pgenhlth	Health on the whole in last 12 months	
phhtyf1	Household type F (grouped)	
phhtype	Household type A (grouped)	
pillness	Any longstanding illness or disability	
plimitac	If longstanding illness limits activity	
pmcob1	Mother's country of birth	
pnumveh	Number of cars	
pregnant	Whether woman pregnant at time of interview	
preltohr	Relationship to HOH/HRP	

probmore	Further probe on more children for “don't knows” on MORECHLD	
pseghrp	Socio-economic group of HOH/HRP	
ptenure	Tenure	
sched	Full or proxy interview	
selfcomr	Self-completion of the FI section	
sex	Sex of the respondent	
sterild	Whether has been sterilized	
totchld	Total intended family size	
weightn	Original ONS weight	
Quality flags		
nchr	Revised count of number of live births and indicator of quality of revised CPC fertility history	
marqual	Overall quality of the marriage history	
marq1-marq7	Quality of information on marriage number n in marriage history’	
partq1-partq11	Quality of information on partnership number n in partnership history	
partqual	Overall quality of the partnership history	
Marriage and partnership		
curtype	Type of current partnership (at survey)	
curmmt	Month of current marriage	
curmt	Month current partnership started	
curmyr	Year of current marriage	
curyr	Year current partnership started	
defacto	Living with a partner/spouse	
mardmt1-mardmt7	Month of divorce for the n^{th} marriage in marriage history	
mardyr1-mardyr7	Year of divorce for the n^{th} marriage in marriage history	
marend1-marend7	Type of end of n^{th} marriage in marriage history	
maremt1-maremt7	Month of end of n^{th} marriage in marriage history	
mareyr1-mareyr7	Year of end of n^{th} marriage in marriage history	
marmt1-marmt7	Month of n^{th} marriage in marriage history	

marpmt1-marpmt7	Month of premarital cohabitation for the n^{th} marriage in marriage history	
marpyr1-marpyr7	Year of premarital cohabitation for the n^{th} marriage in marriage history	
marpre1-marpre7	Premarital cohabitation for the n^{th} marriage in marriage history	
marst	Declared marital status	
marqual	Overall quality of the marriage history	
marq1-marq7	Quality of information on marriage number n in marriage history'	
maryr1-maryr7	Year of nth marriage in marriage history	
ncoh	Number of free standing cohabitations declared in family information section	
nmaru	Number of valid marriage dates in marriage history in the Family Information section	
nmar	Number of marriages declared in the Family Information section	
npart	Number of partnerships reported in GHS rounds from 2000 onwards	
npartu	Number of valid partnership dates in partnership history in the Family Information section	
pardmt1-pardmt7	Month of divorce in n^{th} partnership of partnership history	
pardyr1-pardyr7	Year of divorce in n^{th} partnership of partnership history	
paremt1-paremt7	Month of end of nth partnership in partnership history	
parend1-parend7	Type of end of n^{th} partnership in partnership history	
pareyr1-pareyr7	Year of end of n^{th} partnership in partnership history	
parmar1-parmar7	Marriage during the n^{th} partnership in partnership history	

parmmt1-parmmt7	Month of marriage during the n^{th} partnership in partnership history	
parmt1-parmt7	Month of n^{th} partnership in partnership history	
parmyr1-parmyr7	Year of marriage during the n^{th} partnership in partnership history	
partq1-partq7	Quality of information on partnership number n in partnership history	
partqual	Overall quality of the partnership history	
paryr1-paryr7	Year of n^{th} partnership in partnership history	
CPC revised fertility variables		
bthmtr1-bthmtr15	Month of birth of 1^{st} to 15^{th} live births in the revised CPC birth history	
bthyrr1-bthyrr15	Year of birth of 1^{st} to 15^{th} live births in the revised CPC birth history	
livwthr1-livwthr15	Does the 1^{st}-15^{th} liveborn child in the birth history live with the respondent	
nchr	Revised count of number of live births and indicator of quality of revised CPC fertility history	
sexbthr1-sexbthr15	Sex of the 1^{st} to 15^{th} live birth in the revised CPC birth history	
Original GHS fertility variables		
bthmt1-bthmt15	Month of birth of 1^{st} to 15^{th} live births in the original GHS birth history	
bthyr1-bthyr15	Year of birth of 1^{st} to 15^{th} live births in the original GHS birth history	
livwth1-livwth15	Does the 1^{st}-15^{th} live-born child in the original GHS birth history live with the respondent	
nbirths	Number of live births declared in the Family Information section	
pregnant	Whether woman pregnant at time of interview	

sexbth1-sexbth15	Sex of the 1st to 15th live birth in the original birth history	
Weights		
weightn	Original ONS weight	
wgtcpc	Weights for all individuals in the CPC time-series	Weights generated by CPC
wgtcpcstd	Weights for all individuals in the CPC time-series, normalised	Weights generated by CPC
wgtcpcfi	Weights for individuals answering the Family Information section in the CPC time-series	Weights generated by CPC
wgtcpcfistd	Weights for individuals answering the Family Information section in the CPC time-series, normalised and trimmed at 3	Weights generated by CPC

3. DATA DICTIONARY

Please note that the database initially contained data from GHS rounds 1979-2007 only, 2008 and 2009 being added at a later stage. The 2008 and 2009 coding is consistent with that of rounds 2005-2007, but the data dictionary has not been updated to reflect this.

acctyp91 ‘Type of accommodation’

1 purpose built flat
2 Other

This variable of stratification is available 1979-1996

acctyp01 ‘Type of accommodation’

1 Detached, semi-detached, terraced
2 Other

The variable is available from 1998 onwards only.

Variables used in the construction of **acctyp91**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979 to 1982	1983 to 1995	1996	**acctyp91**
	TYPEHOUS	TYPACCM	TYPACC96	
Detached house	1	1	1	**1**
Semi-detach. house	2	2	2	**1**
Terraced house	3	3	3	**1**
Purpose built flat	4,5	4,5	4,5	**2**
Partial house	6,7	6,7	6,7	**2**
other	0,8,9	0,8,9	0,8	**2**
Missing values (NA)	No missing values (variables used for stratification)			

Variables used in the construction of **acctyp01**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS	CPC code
	1998 to 2007	**acctyp01**
	ACCOM	
A house or a bungalow	1	**1**
A flat or a maisonette	2	**0**
A room/rooms	3	**0**
something else	4	**0**
Missing values (NA)	No missing values (variables used for stratification)	

age ‘Age of the respondent’

0 to 98	aged 0 to 98 at survey
99	aged 99 or more at survey

The variable age gives the age at last birthday.

Variables used in the construction of **age**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS		CPC code
	1979 to 1982	1983 to 2007	**age**
	AGE	AGE	
0 to 98	0 to 98	0 to 98	**0 to 98**
99+	99	99 to 102	**99**
Missing values (NA)	Any missing codes ever used in GHS	Any missing codes ever used in GHS	**-8**

In 2001, the respondent’s age in completed years has been top-coded at 85; thus for any person aged over 85 years in the 2001 GHS edition, this variable has the value 85. However, month and year of birth are present for these respondents, and so an exact age can be calculated.

Other survey years include a significant number of respondents whose reported age in completed years is more than one year different from the age implied by their reported date of birth. We have not edited the data to make these two reports consistent (see Annex H).

agelft 'Age left full-time education'

Variable giving the age at end of school/studies, reconstructed from the variables AGELFTC and AGELFTS.

-9	DNA
-8	NA
98-99	still at school
1	never went to school
3-66	age at end of continuous education (integer)

Variables used in the construction of **agelft**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979 to 1982	1983 to 1992	2000 to 2007	
			EDAGE	**agelft**
still at school	STUDCUR AGELFTS AGELFTC	STUDCUR AGELFTSC AGELFTFT	96	**98-99**
never went to school	AGELFTS AGELFTC	AGELFTSC AGELFTFT	97	**1**
3 to 66	AGELFTS AGELFTC	AGELFTSC AGELFTFT	3-66	**3 to 66**
Missing (NA)	Missing codes for NA GHS	Missing codes for NA	-8 or 2 for 2001	**-8**
Does not apply (DNA)	Missing codes for DNA GHS	Missing codes for DNA	Missing codes for DNA	**-9**

The filter variables have different names in 1979-82, 1983-84, 1985-92 and 1993-98.

From 2000-2007, **agelft** represents approximately the age at which the respondent completed their continuous full time education; see Ní Bhrolcháin and Beaujouan (2012) for a comparison with national statistics. In rounds before 2000, deficiencies have been found in this variable, as detailed in Annex G, and we do not consider the variable usable before the 2000 GHS round.

Age of persons in household: age variables derived by CPC

ageldchh ‘Age of the eldest child in the household’

ageldhh ‘Age of the eldest person in the household’

ageygcfu ‘Age of the youngest child in the family unit’

ageygchh ‘Age of the youngest child in the household’

agygsthh ‘Age of the youngest person in the household’

brthyr – brthmt 'Year and month of birth of the respondent'

Available from 1982 to 1985 for individuals who completed the Family Information section and from 1986 to 2006 for all respondents. Not originally available for any respondent 1979-81.

Prior to 1982, only age was asked. In all cases where date of birth was either not asked or missing, it has been attributed in the way described below.

Variables used in the construction of **brthyr**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS					CPC code
	1982	1983 to 1985	1986 to 1995	1996	1998-2007	**brthyr**
	FIYRBORN	YEARBORN	DOBIRTHY	DOB	BIRTHY	
1896-1999	age	year in 2 digits	year in 2 digits	date format	year	**1896-2007**
Missing values (NA)	Any missing codes ever used in GHS					**-8**

Because this was originally a 2-digit year of birth, before 1996 we use the age to determine whether the true date is 1800+ 2-digit or 1900+2-digit

Variables used in the construction of **brthmt**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS					CPC code
	1982	1983 to 1985	1986 to 1995	1996	1998-2007	**brthmt**
	FIMONBRN	MONBORN	DOBIRTHM	DOB	BIRTHM	
1-12	month	month	month	Date format	month	**1-12**
If year extracted from age			13			**-8**
Missing values (NA)	Any missing codes ever used in GHS					**-8**

Where date of birth is missing or was not asked for, we calculate date of birth assuming that the respondent is aged "**age**+0.5 years" at the date of interview. Dates used in the file are coded in month and year only.

bthmt1-bthmt15 'Month of birth of 1st to 15th live births in the original birth history'

bthyr1-bthyr15 'Year of birth of 1st to 15th live births in the original birth history'

Year and month of live births to the respondent

Variables used in the construction of **bthyr1-bthyr15**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979 to 1982	1983 to 1996	1998 to 2007	**bthyr1-bthyr15**
	CHDYR1-n	YOB1-n	BIRTHDTE BIRTHDT2-9 BIRTHD10-n	
1941-2007	1945-1982	1941-1997		**1941-2007**
Missing (NA)	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	**-9**

Variables used in the construction of **bthmt1-bthmt15**
n indicates whatever range of the variables is available - the maximum number of birth slots allowed for differs across years.

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979 to 1982	1983 to 1996	1998 to 2007	**bthmt1-bthmt15**
	CHDMTH1-n	MOB1-n	BIRTHDTE BIRTHDT2-9 BIRTHD10-n	
1-12	1-12	1-12		**1-12**
Missing (NA)	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	**-9**

bthmtr1-bthmtr15 ‘Month of birth of 1st to 15th live births in the revised birth history’

bthyrr1-bthyrr15 ‘Year of birth of 1st to 15th live births in the revised birth history’

Year and month of live births to the respondent in the revised birth history.

chdbenef ‘Whether receives child benefit’

-9	DNA
-8	NA
0	no child benefit
1	child benefits

This variable is constructed from the successive variables on child benefit:
1979-1982: CHLDBEN
1983-1991: CHBEN
1992-1996: STATBNM1
1998 2003-2007: BEN1Q1
2000-2002: BEN1QM1

Additionally, SCHED has been used as a filter together with these variables in order to determine the NA (-8) and DNA (-9).

contrapt ‘Whether currently using contraception or not’

1	yes
2	no
-9	DNA
-8	NA

The question is asked only of women and men aged under 50. As for all questions in the contraception section, it is asked for selected years only.

Variables used in the construction of **contrapt**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				CPC code
	1983; 1986	1989; 1993; 1995	1991	1998; 2002	**contrapt**
	PRVNTM1-PRVNTM4	CCMUSUM1-CCMUSUM4	CCCURRM1-CCCURRM4	CCMUSU1-CCMUSU4	
yes	any other valid code	any other valid code	any other valid code	any other valid code	**1**
no	11	16,17,18	15, 16	15,16	**2**
Missing (NA)					**-8**
Does not apply (DNA)	DNA	DNA	14,17,18,19, other DNA	17, other DNA	**-9**

The question is generally not asked of sterilized respondents, but in 1989 and 1991 sterilized persons were also asked the question. In 1998 and 2002, where it was known that both partners were sterilized, some respondents were nevertheless asked this question and reported that their partner had a vasectomy (original code 17): they were recoded -9 on variable **contrapt**.

curmmt ‘Month of current marriage”

curmyr ‘Year of current marriage ”

Year and month of marriage current at survey, if there was a marriage at survey (**curtype**=2 or 3).

Variables used in the construction of **curmyr**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				CPC code
	1979-1988	1989-1996	1998	2000 to 2007	**curmyr**
	YRCUR	YRMAR1-YRMAR7	YRMAR-YRMAR5	YRMAR YRMAR2-7	
4-digit calendar year					**4-digit calendar year**
Missing (NA)	8	-8	-8	Missing codes in GHS	**-8**
Does not apply (DNA)	9	-9	-9	Missing codes in GHS	**-9**

Variables used in the construction of **curmmt**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				CPC code
	1979-1988	1989-96	1998	2000 to 2007	**curmmt**
	MONCUR	MONMAR1-MONMAR7	MONMAR-MONMAR5	MONMAR MONMAR2-7	
1-12 (+0.5 where attributed)					**1-12** *(+0.5 where attributed)*
Missing (NA)	8	-8	-8	Missing codes in GHS	**-8**
Does not apply (DNA)	9	-9	-9	Missing codes in GHS	**-9**

curtype 'Type of current partnership (at survey)'

0 not currently in a partnership
1 cohabitation
2 direct marriage
3 marriage preceded by premarital cohabitation

Information about the current partnership is available since 1979. In 1979 and from 1981 onwards respondents were asked whether they had cohabited premaritally with their current (if currently married) or most recent (if not currently married) spouse. However in 1980 respondents were not asked whether they had cohabited premaritally with their current/most recent spouse. Therefore marriages in 1980 are all coded "2".

Variables used in the construction of **curtype**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979[1],1981-1988	1989-1998	2000 to 2007	**curtype**
no current relationship	MARITAL	MARSTAT	CLEND ENDCOH1-3 HOWENED HOWENDE2-7	**0**
non-marital cohabitation	WHEREWED XSLIVTOG	WHEREWED TGTHR	CLYR STCOY1-3	**1**
direct marriage	MARITAL	MARSTAT	LVTGTHR LVTGTHR2- LVTGTHR7	**2**
marriage with premarital cohabitation	LIVTOG	LVTGTHR1- LVTGTHR7		**3**
Missing (NA)			PARTQUAL=-8	**-8**
Does not apply (DNA)			PARTQUAL=-9	**-9**

[1] Information on premarital cohabitation was not asked in 1980.

curmt ‘Month current partnership started’

curyr ‘Year current partnership started’

Year and month of start of the current partnership = year

- of start living together if cohabiting at survey
- of beginning of premarital cohabitation if married at survey and cohabited premaritally
- of marriage if married at survey and no premarital cohabitation

Variables used in the construction of **curyr**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1981-1982	1983-1998	2000 to 2007	**curyr**
	TOGYRS, CLLIVYR, XSLIVYR	YRSTGTHR, CLYR, STRTYR	STCOY1-3 YRLVTG YRLVTG2-7 YRMAR YRMAR2-7	
4 – digit calendar year				**4 – digit calendar year**
Missing (NA)	8	-8	Missing codes in GHS	**-8**
Does not apply (DNA)	9	-9	Missing codes in GHS	**-9**

Variables used in the construction of **curmt**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1981-1982	1983-1988	2000 to 2007	**curmt**
	TOGMTHS, CLLIVMTH, XSLIVMTH	MONTGTHR, CLMON, STRTMON	STCOM1-3 MONLVTG MONLVTG2-7 MONMAR MONMAR2-7	
1-12 (+0.5 where attributed)				**1-12** *(+0.5 where attributed)*
Missing (NA)	8	-8	Missing codes in GHS	**-8**
Does not apply (DNA)	9	-9	Missing codes in GHS	**-9**

defacto ‘Living with a partner/spouse’

1	Living with spouse-FI
2	Living with partner-FI
3	Not living with anyone- FI
4	Same sex civil partnership-FI
5	Same sex partner-FI
6	Under 16 (non FI)
11	Living with spouse- non FI
12	Living with partner- non FI
13	Not living with anyone- non FI
14	Same sex civil partnership- non FI
15	Same sex partner- non FI
-8	NA
-9	DNA

This variable is a best guess at the *de facto* partnership status current at the time of the survey. Our coding takes the declared marital status in the main section of the questionnaire as the default coding, and amends this according to FI information, where available. The marital status question is primarily about legal marital status, but cannot be relied on as accurate in that respect, and in any case includes the category ‘cohabiting’ for some years (1986-96).

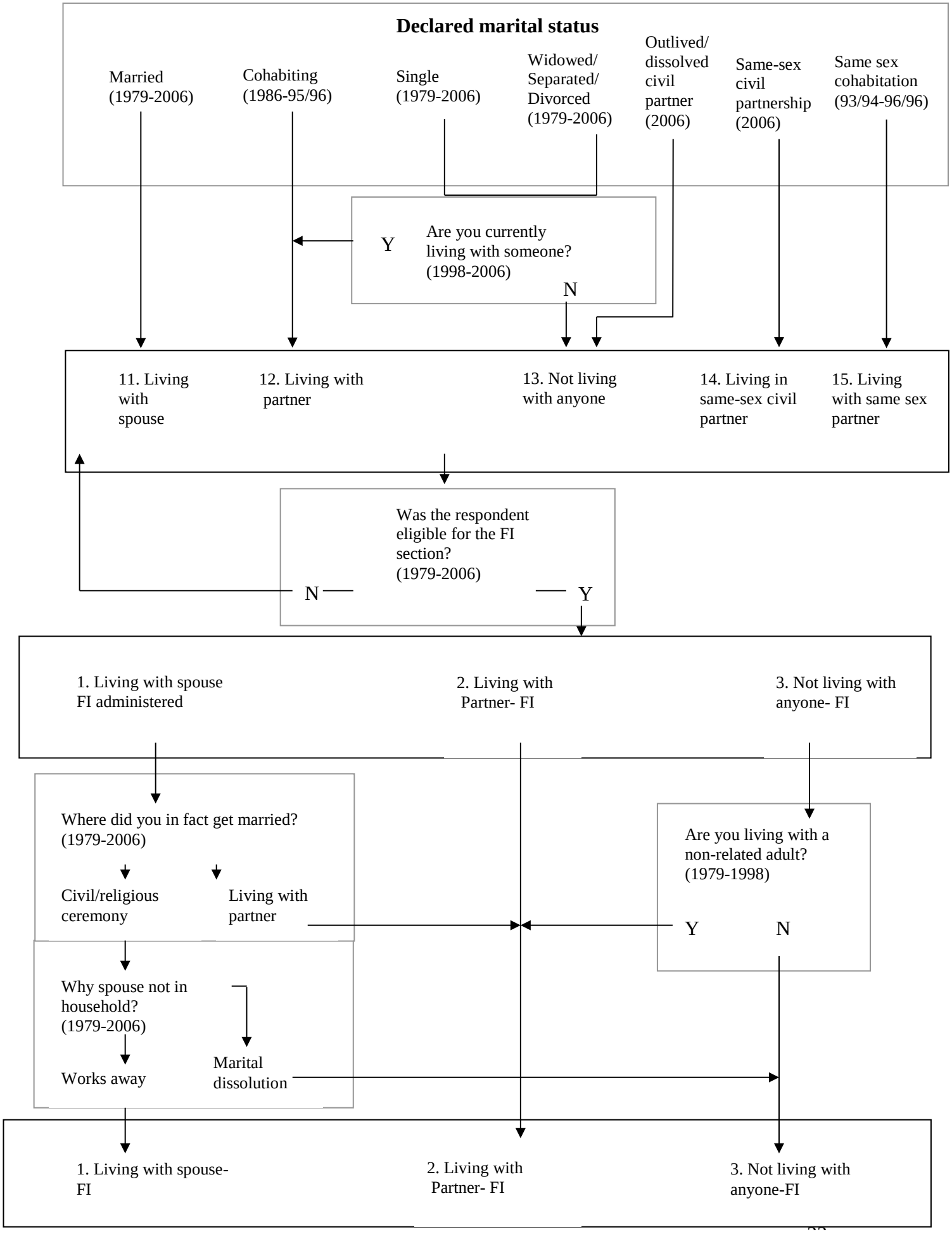
We identify 5 possible partnership states:

1) Living with spouse
2) Living with partner
3) Not living with a partner
4) Living in same sex civil partnership
5) Living with same sex partner.

Two parallel sets of codes are used to distinguish cases where FI information is available (codes 1-5) and is not available (codes 10-15) to modify declared marital status. The accuracy with which de facto status can be determined depends on the availability of the Family Information section of the GHS. Availability of the FI varies by age, sex and survey year, whether the respondent was a proxy and whether the respondent completed the FI section.

The coding of **defacto** is explained by the following table and flow chart. The syntax is given in Annex E.

Flow chart describing defacto



Note: Before 1986, those SWD who did not answer FI section may have been cohabiting, but this will not have been recorded. This is the only limitation on defacto classification. FI denotes that the respondent was eligible for Family Information module and responded completely.

Table: Mapping from categories to component variables to defacto codes

Marital status question[1]	Live with?[2]	FI present[3]	Living with partner[5]	Wherewed[4]	Husbaway[6]	defacto
1979-95						
single, separated, widowed, divorced		no				13
single, separated, widowed, divorced		yes	no			3
single, separated, widowed, divorced		yes	yes			2
single, separated, widowed, divorced		yes	Na/missing			13
Married		no				11
Married		yes		not wed		2
Married		yes		wed	yes	1
Married		yes		wed	no	3
Married		yes		wed	Na/missing	1
Married		yes		Na/missing	yes	11
Married		yes		na/missing	no	13
married		yes		na/missing	na/missing	13
1996-2006						
single, separated, widowed, divorced	no	no				13
single, separated, widowed, divorced	yes	no				12
single, separated, widowed, divorced	na	no				13
single, separated, widowed, divorced	no	yes				3
single, separated, widowed, divorced	yes	yes				2
single, separated, widowed, divorced	na	yes				3
Married		no				11
Married		yes		not wed		2
Married		yes		wed	yes	1
Married		yes		wed	no	3
Married		yes		wed	na/missing	1
Married		yes		na/missing	yes	11
Married		yes		na/missing	no	13
Married		yes		na/missing	na/missing	13
1993-2005						
same sex cohabitation		no				15
same sex cohabitation		yes				15
2006-						
same sex cohabitation		no				15
same sex cohabitation		yes				5
same sex civil partnership		no				14
same sex civil partnership		yes				14
dissolved same sex civil partnership		no				13
dissolved same sex civil partnership		yes				3

Original GHS variables used in the construction of defacto

Year	Variable 1 (Legal marital status)	Variable 2 (Currently cohabiting)	Variable 3 (Type of self-completion)	Variable 4 (Type of marriage ceremony)	Variable 5 (Cohabiting with unrelated adult in household)	Variable 6 (Reason for spouse not in household)
1979	MARITAL	-	FISELF	WHEREWED	XSLIVTOG	HUSAWAY
1980	MARITAL	-	FISELF	WHEREWED	XSLIVTOG	HUSAWAY
1981	MARITAL	-	FISELF	WHEREWED	XSLIVTOG	HUSAWAY
1982	MARITAL	-	FISELF	WHEREWED	XSLIVTOG	HUSAWAY
1983	MARSTAT	-	CHECKB	WHEREWED	STGTHR, WDSTGTHR	HUSBAWAY
1984	MARSTAT	-	CHECKB	WHEREWED	STGTHR, WDSTGTHR	HUSBAWAY
1985	MARSTAT	-	CHECKB	WHEREWED	STGTHR, WDSTGTHR	HUSBAWAY
1986	MARSTAT	-	CHECKB	WHEREWED	STGTHR, WDSTGTHR	HUSBAWAY
1987	MARSTAT	-	CHECKB	WHEREWED	STGTHR, WDSTGTHR	HUSBAWAY
1988	MARSTAT	-	CHECKB	WHEREWED	TGTHR	HUSBAWAY
1989	MARSTAT	-	FAMINFSG	WHEREWED	TGTHR	HUSBAWAY
1990	MARSTAT	-	FAMINFSG	WHEREWED	TGTHR	HUSBAWAY
1991	MARSTAT	-	FAMINFSG	WHEREWED	TGTHR	HUSBAWAY
1992	MARSTAT	-	FAMINFSG	WHEREWED	TGTHR	HUSBAWAY
1993	MARSTAT	-	FAMINFSG	WHEREWED	TGTHR	HUSBAWAY
1994	MARSTAT	-	SELFCOM3	WHEREWED	TGTHR2	HUSBAWAY
1995	MARSTAT	-	SELFCOM3	WHEREWED	TGTHR2	HUSBAWAY
1996	MSTAT	COHABIT	SELFCOM3	WHEREWED	TGTHR2	HUSBAWAY
1997	----------	----------	----------	----------	----------	----------
1998	MARSTAT	LIVEWITH	SELFCOM3	WHEREWED	TGTHR2	HUSBAWAY
1999	----------	----------	----------	----------	----------	----------
2000	MARSTAT	LIVEWITH	SELFCOM3	WHEREWED	-	HUSBAWAY
2001	MARSTAT	LIVEWITH	SELFCOM3	WHEREWED	-	HUSBAWAY
2002	MARSTAT	LIVEWITH	SELFCOM3	WHEREWED	-	HUSBAWAY
2003	MARSTAT	LIVEWITH	SELFCOM3	WHEREWED	-	HUSBAWAY
2004	MARSTAT	LIVEWITH	SELFCOM3	AREWED	-	HUSBAWAY
2005	MARSTAT	LIVEWITH	SELFCOM3	AREWED	-	HUSBAWAY
2006	MARSTA	LIVWTH	SELFCOM3	AREWED	-	HUSBAWAY

degree ‘Highest qualification = degree or higher’

-5	missing
0	other or no qualification
1	degree or higher

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.
For the additional years 2005-2007, the more detailed variable **pedfull** (derived itself from EDLEV00 which is also described in this userguide) has been used to derive the variable the way it was constructed in the 1972-2004 time-series.

eligfi ‘Eligible for the Family Information section’

1	eligible for FI section and not proxy
2	eligible but proxy
3	not eligible

This variable summarizes which individuals are interviewed in the Family Information section. Only individuals coded **eligfi**=1 are asked the FI section. Proxy respondents have been coded since 1980 and are never asked the Family Information (FI) section. Thus they are always coded “does not apply” on all the variables of the FI section.

In the FI section, an additional filter is active when asking all the details on births, fertility history and intentions: only women are asked these questions. Further details can be found in Annex F.

Who is eligible for the Family Information section across GHS rounds?
Between 1979 and 1985 inclusive, non-proxy women aged 18-49 or married and aged 16-17, are asked the Family Information questions. From 1986 onwards, both men and women aged 16-59, who are not proxies, are asked the Family Information section, but men are not asked the birth history.

Variables used in the construction of **eligfi:**

age, **sex**, **marst** and **sched**

(SAS) Coding:

```
if year<=1985 then do;
        if sex=2 and ((16<=age<50 and marst=2) or 18<=age<50) then eligfi=1; *eligible;
                else eligfi=3; *not eligible;
        if sex=2 and ((16<=age<50 and marst=2) or 18<=age<50) and sched=2 then eligfi=2;
        *eligible but proxy;
end;
if year>1985 then do;
        if 16<=age<60 then eligfi=1;
                else eligfi=3;
        if 16<=age<60 and sched=2 then eligfi=2;
end;
```

ethgrp2 ‘Ethnic group’

1 White
2 Non white
-8 NA
-9 DNA

The only variable that can be produced reasonably consistently across GHS rounds is a binary classification of ethnic group. For the years 1979-82, it is based on interviewer assessment of the respondents’ ethnic group rather than a direct question.

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979-1982	1983-1996	1998-2006	**ethgrp2**
	COLOUR	ORIGIN	ETHNIC	
White	2,5	1, 98	1, 2	1
Non-white	1,4	2-77	3-15	2
Missing values (-8)	3,8			

funo ‘Family unit number’

Within each household, a number is given to each family unit. **funo** is a one-digit variable in 1979-1996, and is the household serial number plus one digit in 1998-2007.

Variables used in the construction of **funo**

Value labels/names for the codes on the original GHS dataset	CPC code		
	1979 to 1996	1998 to 2007	**funo**
	FAMUNIT	FSERIAL	
	one digit	HHSERIAL + one digit	

Not present for 2008 and 2009**.**

hhserial ‘Household serial number’

A household serial number is given each year to each household in the survey. Household serial numbers are re-used each year and hence the same serial number in two different survey years does not refer to the same household.

Variables used in the construction of **hhserial**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979 to 1982	1983 to 1998	2000 to 2007	**hhserial**
	HOUSENO	HSERNO	HSERIAL	
1-30102201				

See also section in Introduction on merging with original GHS datasets.

intmt ‘Month of interview’

intyr ‘Year of interview’

Year and month of interview. Note that **intyr** can differ from the survey year (**year**)

Variables used in the construction of **intyr**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS					CPC code
	1979 to 1986	1987 to 1995	1996	1998	2000 to 2007	**intyr**
	YEAR	YEAR	INTDATE	STARTDAT	STARTDAT	
1979-2007	year in 2 digits	year in 2 digits	Special date format	Standard date format	Standard date format	**1979-2007**
Missing values (NA)	Any missing codes ever used in GHS					**-8**

The interview year is not always equal to the survey edition’s year, and especially in 1996. This is developed in appendix.

Variables used in the construction of **intmt**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS					CPC code
	1979 to 1986	1987 to 1995	1996	1998	2000 to 2007	**intmt**
	year	year	intdate	startdat	startdat	
1-12	1-12	1-12	Special date format	Standard date format	Standard date format	**1-12**
Missing values (NA)	Any missing codes ever used in GHS					**-8**

[1]In the original ESDS GHS file for 1996-7 dates were represented as “days since 14 October 1582” with 15 October 1582 = 1.

lastft ‘Where last attended course’

-9	DNA
-8	NA
-6	child/out age/no int
1	elementary or secondary school
2	university
3	polytechnic
4	nursing school or teaching hospital
5	some other type of college

lastft variable is available *only before 1998*.

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				CPC code
	1979	1980 to 1982	1983 to 1984	1985 to 1998	**lastft**
	EDTYPE	EDTYPE	LASTSCH1	LASTSCH2	
elementary or secondary school	1	1	1	1	**1**
university	3	2	2	2	**2**
polytechnic	-	-	5	3	**3**
nursing school or teaching hospital	2	3	3	4	**4**
some other type of college	4,6	4	4	5	**5**
Missing values (NA)	Missing codes for NA in GHS				**-8**
Does not apply (DNA)	Missing codes for DNA in GHS				**-8**

FURTHRED and AGELFTS/AGELFTSC are used as filter variables, and people who have not prolonged their education but have been to school are considered as having been to an elementary/secondary school even if originally they are not directly coded as such.

livwth1-livwth15 'Does the 1st-15th live-born child in the original birth history live with the respondent'

1 yes
2 no, lives elsewhere
3 no, deceased

Whether the child lives with the respondent

Variables used in the construction of **livwth1-livwth15**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979 to 1982	1983 to 1996	1998 to 2007	
	CHDLIV1-*n*	LIVE1-*n*	CHLDLIVE CHLDLIV2-9 CHLDLI10-*n*	**livwth1-livwth15**
yes	1	1	1	**1**
no, lives elsewhere	2	2	2	**2**
No, deceased	3	3	3	**3**
Missing (NA)	Missing codes in GHS	Missing codes in GHS	-8	**-8**
Does not apply (DNA)	Missing codes in GHS	Missing codes in GHS	-6, -9	**-9**

livwthr1-livwthr15 ‘Does the 1st-15th live-born child in the revised birth history live with the respondent’

1	yes
2	no, lives elsewhere
3	no, deceased

Whether the child lives with the respondent. Note that all the live births recovered in the CPC revised birth history live with the respondent.

mardmt1-mardmt7 ‘Month of divorce for n-th marriage in marriage history’

mardyr1-mardyr7 ‘Year of divorce for n-th marriage in marriage history’

Year and month of divorce of each marriage of the respondent in the sequence of marriages.

Variables used in the construction of **mardyr1-mardyr7**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				CPC code
	1979-83	1983-88	1989-98	2000 to 2007	**mardyr1-mardyr7**
	DECR1YR-DECR3YR	YRDIV1-YRDIV3	YRDIV1-YRDIV7	YRDIV YRDIV2-7	
4 - digit calendar year					**4 - digit calendar year**
Missing (NA)	8	-8	-8	Missing codes in GHS	**-8**
Does not apply (DNA)	9	-9	-9	Missing codes in GHS	**-9**

Variables used in the construction of **mardmt1-mardmt7**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				CPC code
	1979-83	1983-88	1989-98	1998 to 2007	**mardmt1-mardmt7**
	DECR1MTH-DECR3MTH	MONDIV1-MONDIV3	MONDIV1-MONDIV7	MONDIV MONDIV2-7	
1-12 (+ 0.5 where attributed)					**1-12** *(+ 0.5 where attributed)*
Missing (NA)	8	-8	-8	Missing codes in GHS	**-8**
Does not apply (DNA)	9	-9	-9	Missing codes in GHS	**-9**

maremt1-maremt7 ‘Month of end of n-th marriage in marriage history’

mareyr1-mareyr7 ‘Year of end of n-th marriage in marriage history’

Year and month of end (by separation or death of the partner) of each marriage of the respondent in the sequence of marriages.

Variables used in the construction of **mareyr1-mareyr7**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				CPC code
	1979-82	1983-88	1989-98	2000 to 2007	**mareyr1-mareyr7**
	DEP1YR-DEP3YR	YREND1-YREND3	YRDIE1-YRDIE7 YRSEP1-YRSEP7	YRDIE YRDIE2-7 YRSEP YRSEP2-7	
4 - digit calendar year					**4 - digit calendar year**
Missing (NA)	8	-8	-8	Missing codes in GHS	**-8**
Does not apply (DNA)	9	-9	-9	Missing codes in GHS	**-9**

Variables used in the construction of **maremt1-maremt7**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				CPC code
	1979-82	1983-88	1989-98	2000 to 2007	**maremt1-maremt7**
	DEP1MTH-DEP3MTH	MONEND1-MONEND3	MONDIE1-MONDIE7 MONSEP1-MONSEP7	MONDIE MONDIE2-7 MONSEP MONSEP2-7	
1 to 12 (+ 0.5 where attributed)					**1 to 12** *(+ 0.5 where attributed)*
Missing (NA)	8	-8	-8	Missing codes in GHS	**-8**
Does not apply (DNA)	9	-9	-9	Missing codes in GHS	**-9**

marend1-marend7 ‘Type of end of n-th marriage in marriage history’

0 partnership current at survey
1 death of the partner
2 divorce
3 separation

Variables used in the construction of **marend1-marend7**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				CPC code
	1979-82	1983-1996	1998	2000 to 2007	**marend1-marend7**
	MAR1END-MAR3END	WHYEND1-WHYEND7	HOWENDED, HOWENDE2-HOWENDE5, CURRENT, CURRENT2-7	HOWENDED HOWENDE1-7 CURRENT CURRENT2-7	
current marriage			CURRENT=1	CURRENT=1	**0**
death	1	1	HOWENDED=1	HOWENDE=1	**1**
divorce	2	2	HOWENDED=2	HOWENDE=2	**2**
separation	3	3	HOWENDED=3	HOWENDE=3	**3**
Missing (NA)	8	-8	-8	-8	**-8**
Does not apply (DNA)	9	-9	-9	**eligfi** ne 1	**-9**

marmt1-marmt7 ‘Month of n-th marriage in marriage history’

maryr1-maryr7 “Year of n-th marriage in marriage history”

Year and month of marriage number n of the respondent in the sequence of marriages

Variables used in the construction of **maryr1-maryr7**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				CPC code
	1979-82	1983-88	1989-98	2000 to 2007	**maryr1-maryr7**
	MARCYR, MAR1YR-MAR3YR	YRCUR, YRMAR1-YRMAR3	YRMAR1-YRMAR7	YRMAR YRMAR2-7	
4 - digit calendar year					**4 - digit calendar year**
Missing (NA)	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	**-9**

Variables used in the construction of **marmt1-marmt7**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				CPC code
	1979-82	1983-88	1989-98	1998 to 2007	**marmt1-marmt7**
	MARCMTH, MAR1MTH-MAR3MTH	MONCUR, MONMAR1-MONMAR3	MONMAR1-MONMAR7	MONMAR MONMAR2-7	
1-12 (+0.5 where attributed)					**1-12** *(+0.5 where attributed)*
Missing (NA)	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	**-9**

marpmt1-marpmt7 ‘Month of premarital cohabitation for n-th marriage in marriage history’

marpyr1-marpyr7 ‘Year of premarital cohabitation for n-th marriage in marriage history’ maryr1-maryr7 ‘Year of n-th marriage in marriage history’

Year and month of premarital cohabitation before each marriage of the respondent in the sequence of marriages.

Note that in 1979 respondents were only asked whether they had cohabited prior to their current/most recent marriage and not the start date of this premarital cohabitation. In 1980 no information about premarital cohabitation was collected. In the period 1981-1988 inclusive, respondents were only asked about the duration of premarital cohabitation prior to their current/most recent marriage. From 1989 respondents are asked about the duration of premarital cohabitation prior to their current and previous marriages.

Variables used in the construction of **marpyr1-marpyr7**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1981-1988	1989-98	2000 to 2007	**marpyr1-marpyr7**
	TOGMTHS	YRLVTG1 - YRLVTG7	YRLVTG YRLVTG2-7	
4 - digit calendar year				**4 - digit calendar year**
Missing (NA)	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	**-9**

Variables used in the construction of **marpmt1-marpmt7**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1981-1988	1989-98	1998 to 2007	**marpmt1-marpmt7**
	TOGYRS	MONLVTG1 - MONLVTG7	MONLVTG MONLVTG2-7	
1-12 (+0.5 where attributed)				**1-12** *(+0.5 where attributed)*
Missing (NA)	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	Missing codes in GHS	Missing codes in GHS	**-9**

marpre1-marpre7 ‘Premarital cohabitation for n-th marriage in marriage history’

1 yes
2 no

Variables used in the construction of **marpre1-marpre7**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979 and 1981-88	1989-98	2000 to 2007	**marpre1-marpre7**
	LIVEDTOG*	LVTGTHR1-LVTGTHR7	LVTGTHR LVTGTHR2-7	
yes	1	1	1	**1**
no	2	2	2	**2**
Missing (NA)	8	-8	-8	**-8**
Does not apply (DNA)	9	-9	MAR=2	**-9**

* for GHS rounds 1979, and 1981 to 1988 LIVEDTOG was asked only about the current/most recent marriage, and so is attributed to the relevant marriage in the marriage history. From 1989- the question was asked about all marriages. When the question was not asked, these variables are coded -9.

No information about premarital cohabitation was collected in 1980 and so the data are set to -9 for 1980.

marq1-marq7 ‘Quality of information on marriage number n in marriage history’

Quality flags generated by CPC for the marriage history, see Annex D.

Before 1989

0	no n-th marriage declared (**nmar**<n)
1	OK (marr date OK + end date OK + ordered)
2	OK after editing
3	partially OK
4	invalid date or date missing where **nmar**>=n
-9	DNA: not eligible FI, proxy

From 1989 on

0	no n-th marriage declared (**nmar**<n)
10	OK (no precohab + marr date OK + end date OK)
11	OK (precohab date OK + marr date OK + end date OK)
12	OK (precohab date edited + marr date OK + end date OK)
13	OK (precohab date unresolved + marr date OK + end date OK)
20	OK after editing (no precohab + marr date + end date)
21	OK after editing (precohab date OK + edited marr or end date)
22	OK after editing (precohab date edited + edited marr or end date)
23	OK after editing (precohab date unresolved + edited marr or end date)
31	partially OK
41	invalid date or date missing where **nmar**>=n
-9	DNA: not eligible FI, proxy

For the coding of “partially OK”, see also Annex D

marqual ‘Overall quality of the marriage history’

This variable is entirely CPC generated and is a summary of the individual **marq1-n**s.

0	no marriage
1	OK
2	OK after editing
3	partially OK
4	unusable
-9	DNA: not eligible FI, proxy

marqual is coded partially OK either if all marriages are partially OK, or some are OK and some partially OK on the marq1-n indicators. Partially OK does not include any histories with an unusable or missing spell.

marqual is coded OK after editing either if all marriages are OK after editing, or some are OK and some are OK after editing.

marst ‘Declared marital status’

1	single
2	married
3	separated
4	divorced
5	widow
6	cohabitation (1985-96 only)
7	civil partnership (2006-)
-9	not concerned

MARSTAT is the declared marital status recorded for all individuals in the household in the main part of the questionnaire. In the years 1979-1984, valid responses for this variable are Married, Single, Widowed, Divorced and Separated. In the years 1985-92 there is an additional category for cohabiting with a partner. From 1998-present, cohabitation is not included as a pre-coded category of the marital status question, and hence the valid responses are identical to the period 1972-1984 (Married, Single, Widowed, Divorce and Separated). In 2006-7, civil partnerships are added as response categories. Persons who dissolved a civil partnership are considered as single. The **defacto** variable provides further information on same-sex partnerships.

Variables used in the construction of **marst**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS						CPC code
	1979 to 1982	1983 to 1985	1986 to 1992	1993 to 1996	1998 to 2005	2006-2007	**marst**
	MARITAL	MARSTAT	MARSTAT	MARSTAT	MARSTAT	MARSTA	
Married	1	1	1	1	2	2	**2**
Single	2	2	3	3	1	1	**1**
Divorced	4	4	5	5	4	4	**4**
Separated	5	5	6	6	3	3	**3**
Widowed	3	3	4	4	5	5	**5**
Cohabiting			2	2			**6**
Same sex cohabiting				7			**6**
Civil partnership						6	**7**
Separated from civil partnership						**7-8**	**1**
Civil partner died						9	**1**
Less than 16 years	0 or missing			0 or missing	-6	-6	**-9**

morechld ‘Whether intends to have more children’

1	yes
2	no
3	don’t know
11	yes
12	probably yes
13	probably not
14	no
15	don’t know
-9	DNA
-8	NA

Since the label values of the variable changed in 1991, we use the double digit coding to identify the potential inconsistency before and from that year.

Variables used in the construction of **morechld**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979 to 1982	1983 to 1990	1991 to 2007	**morechld**
	CHLDMORE	MORECHLD	MORECHLD	
yes	1	1		**1**
no	2	2		**2**
don’t know	3	3		**3**
yes			1	**11**
probably yes			2	**12**
probably no			3	**13**
no			4	**14**
don’t know			5	**15**
Missing (NA)				**-8**
Does not apply (DNA)				**-9**

Note that from 1991, this variable should be used in conjunction with **probmore**

moreposs 'Could you (and your partner) have more children if you wanted to'

1	could have more children
2	would be difficult, impossible
-9	DNA
-8	NA

The form of the question varies across survey years. We refer the user to the questionnaires for each round for precise details.

The sub-sample who were asked this question differs from the one to which **contrapt** relates, as follows:

before 1989
women aged less than 45 and eligible for the FI section and not pregnant
1989
women aged less than 45 and eligible for the FI section
from 1991 onwards
women aged less than 50 and eligible for the FI section and not pregnant and not sterile

Variables used in the construction of **moreposs**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1983; 1984; 1986; 1987	1989	1991; 1993; 1995; 1998; 2002	**moreposs**
	MOREPOSS	MOREPOSS	MOREPOSS	
could have more children	1	1	1	**1**
would be difficult, impossible	2	2	2	**2**
Missing (NA)				**-8**
Does not apply (DNA)			if not pregnant and not sterile	**-9**

nbirths ‘Number of live births declared in the Family Information section’

Number of live births reported by the respondents answering the fertility history questions (non-proxy women eligible for the FI section).

nbirths is built as a combination of the variable that indicates whether the respondent ever had a live birth and, if so, the number of her live births.

Variables used in the construction of **nbirths**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979 to 1982	1983 to 1996	1998 to 2007	
	CHILD + NCHILDRN	BABY + NUMBABY	BABY + NUMBABY	**nbirths**
0	CHILD=2	BABY=2	BABY=2	**0**
1 to 15	NCHILDRN=1 to 15	NUMBABY=1 to 15	NUMBABY=1 to 15	**1 to 15**
Missing (NA)	Missing codes for NA GHS	Missing codes for NA	-8	**-8**
Does not apply (DNA)	Missing codes for DNA GHS	Missing codes for DNA	-6, -9	**-9**

nbirths is based on the original GHS variables. In a sizeable minority of cases this variable is erroneous, due to an over-reporting of childlessness, as documented in Murphy (2009). The errors are largely concentrated in the rounds from 1998 onwards--see Ní Bhrolcháin, Beaujouan and Murphy (2011) for an account of the source of the errors in the original fertility histories, and for the procedure used to correct these. An alternative variable, **nchr**, has been constructed by CPC giving the number of live births in the birth histories as revised by CPC.

nchr ‘Revised count of number of live births’

This variable gives the number of live births in the birth histories as revised by CPC. The birth histories were revised by using information on own children in the household, from the household grid. See Ní Bhrolcháin et al. (2011) for details. Revised fertility histories were produced from 1994 onwards only. For GHS rounds before 1994, the original and revised fertility histories are identical.

Quality indicator for revised fertility histories
Where nchr>=0, this means that an ordered sequence of births is available for analysis. Any negative code implies that the birth history is either not available or completely incorrect, and so unusable. Fertility histories that are partially in error– e.g. second or later birth date missing—are considered partially usable (e.g. for the date of first birth) and have nchr >=0.

ncoh 'Number of free-standing cohabitations declared in the Family Information section'

Number of free-standing cohabitations originally reported by the respondents in the Family Information section, including the current cohabitation. They cover relationships "in which [respondent] lived together with someone as a couple but did not get married". The questions on dates of all such cohabitation were asked only from 2000 onwards.[1] In GHS rounds before 2000, the variable cannot be calculated and is coded -9.

Free standing cohabitations refer to spells of cohabitation that either had ended in separation or were current at survey.

Variables used in the construction of **ncoh**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS	CPC code
	2000 to 2007	**ncoh**
	NUMCOHAB COHAB **defacto**	
0	COHAB=2	**0**
1 to 8	NUMCOHAB + 1 if **defacto**=2	**1 to 8**
Missing (NA)	-8, **selfcomr**=-8	**-8**
Does not apply (DNA)	**eligfi** ne 1	**-9**

[1] In 1998 a question was asked about whether the respondent had ever cohabited without marrying that partner, and if so, how many such partnerships they had. This variable for 1998 has not been included in the dataset.

nextage 'If intends to have more children, age at birth of next child'

17 to 62	age at which next child expected
-9	DNA
-8	NA

Only asked of those who either intend to have at least one more child, or don't know, or, from 1991 onwards, answered "probably yes" to MORECHLD or PROBMORE.

Variables used in the construction of **nextage**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979 to 1982	1983 to 1990	1991 to 2007	**nextage**
	CHLDMORE AGENXTCH	MORECHLD NEXTAGE	MORECHLD NEXTAGE	
Range 17 to 62	AGENXTCH	NEXTAGE	NEXTAGE	**17 to 62**
Missing (NA)				**-8**
Does not apply (DNA)	when CHLDMORE not eq.1	when MORECHLD not eq.1 and not eq.3	when MORECHLD not eq.11,12 and PROBMORE not eq.1	**-9**

Same conditions as for **totchld**.

nmar 'Number of marriages declared in the Family Information section'

Number of marriages reported by respondents asked the Family Information section.

Variables used in the construction of **nmar**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				CPC code
	1979-82	1983-1988	1989-1998	2000 to 2007	**nmar**
	NLEGMAR, NCLMD	NUMMAR CLNUMMAR	NUMMAR CLNUMMAR	NUMMAR CLNUMMAR	
0	MARITAL	MARSTAT	MARSTAT	**marst**[2] not eq. (2,3,4)	**0**
1 to 7	NLEGMAR, NCLMD	NUMMAR CLNUMMAR	NUMMAR CLNUMMAR	NUMMAR, CLNUMMAR	**1 to 7**
Missing (NA)	FISELF=8	CHECKB=-8	FAMINFSG=-8	-8, **selfcomr**=-8	**-8**
Does not apply (DNA)	FISELF=8	CHECKB=-9	FAMINFSG=98/99	**eligfi** ne 1	**-9**

Detail of the construction of the variable can be found in the accompanying Excel file on "eligibility criteria".

[2] **marst** here is the marital status variable that we have constructed from GHS marital status variable—i.e. the declared marital status variable that appears in the CPC database.

nmaru ‘Number of valid marriage dates in marriage history in the Family Information section’

Variable generated by CPC.

Number of marriage dates with a quality indicator coded either 1 to 3, or 10 to 31. That is final number of marriages for which we have at least partially valid information.

This variable is obtained by counting the number of marq1-n that are coded 1 to 3, or 10 to 31.

npart ‘Number of partnerships reported in GHS rounds from 2000 onwards’

Total number of partnerships reported in the Family Information section, constructed from the number of marriages and cohabitations reported (editing described in Annex C).

Construction of **npart:**
This is the total number of partnerships where a partnership is: direct marriage, cohabitation followed by marriage, cohabitation only (including those current at survey). It is constructed from **ncoh** and **nmar**.

Distinguish from **npartu** (next page), the number of valid partnerships in the history.

As for other variables relating to the (near-) complete partnership history, the number of partnerships is available only from GHS round 2000 onwards.

Note that the GHS questionnaire and documentation does not have a concept of partnership – only of marriages and cohabitations. The concept of a partnership in the CPC database combines these two.

npartu ‘Number of valid partnership dates in partnership history in the Family Information section’

This variable relates to the combined sequence of marriages and cohabitations, and is constructed from the edited sequence of partnerships (editing described in Annex C). Note that where a person who first cohabits and then marries their cohabiting partner, this is counted as one partnership.

From 2000 onwards:
Number of partnership dates with a quality indicator “ok” to “partially ok” (codes 1 to 3). This is the number of partnerships for which we have at least partially valid information.

This variable is obtained by counting the number of partq1-7 that are coded 1 to 3.

This variable applies to respondents in GHS rounds 2000 onwards only.

Number of persons in household: count variables

Count variables derived by CPC using the characteristics of people in the household or in the family unit(s) in the household. The derivation is based on age and sex.

A child in these variables is a person aged <=15, an adult a person aged 16+.

None of these variables has missing values because the age of the individuals in the counted units (household or family unit) is always known.

nadul60 ‘Number of adults aged 60+ in the household’

nadul65 ‘Number of adults aged 65+ in the household’

nadulmen ‘Number of adult men in the household’

nadulwom ‘Number of adult women in the household’

nchdhh ‘Number of children in the household’

nfuhh ‘Number of family units in the household’

npenshh ‘Number of pensioners in the household’

npersfu ‘Number of persons in the family unit’

npershh ‘Number of persons in the household’

pardmt1-pardmt7 ‘Month of divorce in partnership history’

pardyr1-pardyr7 ‘Year of divorce in partnership history’

Year and month of divorce, when there has been a marriage and a divorce during the partnership.

Variables used in the construction of **pardyr1-pardyr7**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS	CPC code
	2000 to 2007	**pardyr1-pardyr7**
	YRDIV YRDIV2-7	
4 - digit calendar year		**4 - digit calendar year**
Missing (NA)	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	**-9**

Variables used in the construction of **pardmt1-pardmt7**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS	CPC code
	1998 to 2007	**pardmt1-pardmt7**
	MONDIV MONDIV2-7	
1-12 (+0.5 where attributed)		**1-12** *(+0.5 where attributed)*
Missing (NA)	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	**-9**

This variable applies to respondents in GHS rounds 2000 onwards only.

paremt1-paremt7 ‘Month of end of n-th partnership in partnership history’

pareyr1-pareyr7 ‘Year of end of n-th partnership in partnership history’

Where the partnership n ended by separation or death of the partner, year and month of end.

Variables used in the construction of **pareyr1-pareyr7**
n indicates any ranges on which the variables are available the different years

	Codes from each year of the GHS	CPC code
Value labels/names for the codes on the original GHS dataset	2000 to 2007 YRDIE YRDIE2-7 YRSEP YRSEP2-7 ENDCOHY1-3 ENDLIVY1-3	**pareyr1-pareyr7**
4- digit calendar year		**4 - digit calendar year**
Missing (NA)	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	**-9**

Variables used in the construction of **paremt1-paremt7**
n indicates any ranges on which the variables are available the different years

	Codes from each year of the GHS	CPC code
Value labels/names for the codes on the original GHS dataset	1998 to 2007 MONDIE MONDIE2-7 MONSEP MONSEP2-7 ENDCOHM1-3 ENDLIVM1-3	**paremt1-paremt7**
1-12 (+0.5 where attributed)		**1-12** *(+0.5 where attributed)*
Missing (NA)	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	**-9**

This variable applies to respondents in GHS rounds 2000 onwards only.

parend1-parend7 'Type of end of n-th partnership in partnership history'

0 partnership current at survey
1 stopped living together
2 death of the partner

Variables used in the construction of **parend1-parend7**

Cohabitations:

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS 2000 to 2007 ENDCOH1-3 **defacto**	CPC code **parend1-parend7**
current partnership	last cohab slot+**defacto**=2	**0**
stopped living together	ENDCOHN=1,2,3,5	**1**
death of the partner	ENDCOHN=4	**2**
Missing (NA)	-8	**-8**
Does not apply (DNA)	**eligfi** ne 1	**-9**

Marriages:

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS 2000 to 2007 HOWENDED HOWENDE1-7 CURRENT CURRENT2-7	CPC code **parend1-parend7**
current partnership	CURRENT=1	**0**
stopped living together	HOWENDE=2,3	**1**
death of the partner	HOWENDE=1	**2**
Missing (NA)	-8	**-8**
Does not apply (DNA)	**eligfi** ne 1	**-9**

This variable applies to respondents in GHS rounds 2000 onwards only.

parmar1-parmar7 ‘Marriage during the n-th partnership in partnership history’

1 marriage
2 no marriage

This variable relates to the combined sequence of marriages and cohabitations. Note that where a person first cohabits and then marries their cohabiting partner, this is counted as one partnership.

Variables used in the construction of **parmar1-parmar7**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS	CPC code
	2000 to 2007	**parmar1-parmar7**
	YRMAR YRMAR2-7	
yes	1	**1**
no	2	**2**
Missing (NA)	-8	**-8**
Does not apply (DNA)	PAR=2	**-9**

Note: This variable records whether the respondent got married during the spell number n, i.e. whether a marriage date is available for this union. It is possible to deduce from this variable and the date of marriage in the partnership (below) a variable indicating whether the union: was a direct marriage, included a marriage, or was a cohabitation only. For instance, if in one spell **parmar***n*=1 and date marriage=date partnership, then it was a direct marriage; if **parmar***n*=1 and date marriage does not equal date partnership, then marriage occurred after a period of cohabitation. If in one spell **parmar***n*=2 then it was/is a free standing cohabitation. This variable applies to respondents in GHS rounds 2000 onwards only.

parmmt1-parmmt7 ‘Month of marriage during the n-th partnership in partnership history’

parmyr1-parmyr7 ‘Year of marriage during the n-th partnership in partnership history’

Where respondent married during the partnership spell n, year and month of the marriage.

Variables used in the construction of **parmyr1-parmyr7**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS	CPC code
	2000 to 2007	**parmyr1-parmyr7**
	YRMAR YRMAR2-7	
4- digit calendar year		**4 - digit calendar year**
Missing (NA)	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	**-9**

Variables used in the construction of **parmmt1-parmmt7**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS	CPC code
	2000 to 2007	**parmmt1-parmmt7**
	MONMAR MONMAR2-7	
1-12 (+0.5 where attributed)		**1-12** *(+0.5 where attributed)*
Missing (NA)	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	**-9**

Note that there could be e.g. a marriage during the first partnership, no marriage during the second (which therefore was a free-standing cohabitation) and another (second) marriage during the third partnership.

This variable applies to respondents in GHS rounds 2000 onwards only.

parmt1-parmt7 'Month of n-th partnership in partnership history'

paryr1-paryr7 'Year of n-th partnership in partnership history'

This variable relates to the combined (and ordered) sequence of marriages and cohabitations. Note that where a person who first cohabits and then marries their cohabiting partner, this is counted as one partnership.

Year and month of 1st-7th partnership

Variables used in the construction of **paryr1-paryr7**
n indicates any ranges on which the variables are available the different years

	Codes from each year of the GHS	CPC code
Value labels/names for the codes on the original GHS dataset	2000 to 2007 STCOY1-3 YRLVTG YRLVTG2-7 YRMAR YRMAR2-7	**paryr1-paryr7**
4- digit calendar year		**4 - digit calendar year**
Missing (NA)	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	**-9**

Variables used in the construction of **parmt1-parmt7**
n indicates any ranges on which the variables are available the different years

	Codes from each year of the GHS	CPC code
Value labels/names for the codes on the original GHS dataset	2000 to 2007 STCOM1-3 MONLVTG MONLVTG2-7 MONMAR MONMAR2-7	**parmt1-parmt7**
1-12 (+0.5 where attributed)		**1-12** *(+0.5 where attributed)*
Missing (NA)	Missing codes in GHS	**-8**
Does not apply (DNA)	Missing codes in GHS	**-9**

This variable applies to respondents in GHS rounds 2000 onwards only.

partq1-partq7 ‘Quality of information on partnership number n in partnership history’

The partnership history is an ordered sequence of partnerships of the respondent, where one partnership relates to cohabitation and/or marriage with one partner. In the survey there is no indication of minimum time spent in the cohabitation/marriage. Note that where a person who first cohabits and then marries their cohabiting partner, this is counted as one partnership.

Before 2000:
No retrospective information on the dates of free-standing cohabitations except if current at survey, and partq1-n not derived prior to 2000.

From 2000 onwards:

0	no date in the nth spell[3]
1	OK (beg date OK + end date OK + ordered)
2	OK after editing
3	partially OK
4	unusable
-9	DNA: not eligible FI, proxy

This variable applies to respondents in GHS rounds 2000 onwards only.

[3] Note that code 0 for PARTQ1-n differs slightly from code 0 for MARQ1-n: here 0 means no date in spell, irrespective of how many partnerships declared; in MARQ1-n 0 means no n-th marriage declared.

partqual ‘Overall quality of the partnership history’

This variable is entirely CPC generated and is a summary of the individual **partq1-n**s.

0	No partnership
1	OK
2	OK after editing
3	Partially OK
4	Unusable
-9	DNA: not eligible FI, proxy

partqual is coded partially OK either if all partnerships are partially OK, or some are OK and some partially OK on the **partq1-n** indicators. Partially OK does not include any histories with an unusable or missing spell.

partqual is coded OK after editing either if all partnerships are OK after editing, or some are OK and some are OK after editing.

This variable applies to respondents in GHS rounds 2000 onwards only.

pcigsmk ‘Number of cigarettes smoked per day’

-5	missing
1	20+ cigarettes a day
2	10-19 cigarettes a day
3	0-9 cigarettes a day
4	NA to cigarettes a day
5	ex-cigarette smoker
6	never smoked

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable CIGSMK has been used to derive the variable in the same way.

pcigsmk1 ‘Smoking status (ever smoked)’

-5	missing
1	current cigarette smoker
2	ex-cigarette smoker
3	never smoked

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable CIGSMK has been used to derive the variable in the same way.

pcob1 ‘Country of birth’

-5	missing
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

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable COB1 has been used to derive the variable in the same way.

pfcob1 ‘Father’s country of birth’

-5	missing
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

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable FCOB1 has been used to derive the variable in the same way.

pcountry ‘UK country’

-5	missing
1	England
2	Wales
3	Scotland
4	Scottish Supplement

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable COUNTRY has been used to derive the variable in the same way.

pcutdown ‘Illness/Injury reduce activity’

-5	missing
1	yes
2	no

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable CUTDOWN has been used to derive the variable in the same way.

pdvilo3a ‘Activity status’

-5	missing
1	in employment
2	unemployed
3	economically inactive

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable DVILO3A has been used to derive the variable in the same way.

pedfull ‘Education level, full form’

-5	missing
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 6-7noawrd (should be grouped with “other qualifications”
16	Foreign Qualification
17	Other Qualification
18	O-levels (do not know number)
19	Never went to school

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable EDLEV00 has been used to derive the variable in the same way.

pgenhlth ‘Health on the whole in last 12 months’

-5	missing
1	good
2	fairly good
3	not good

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable GENHLTH has been used to derive the variable in the same way.

phhtyf1 ‘Household type F (Grouped)’

-5	missing
1	1 person only
2	2+ unrelated adults
3	married couple, dependent children
4	married couple, independent children
5	married couple, no children
6	lone parent, dependent children
7	lone parent, independent children
8	2+ families
9	same sex cohab
10	cohab couple, dependent children
11	cohab couple, independent children
12	cohab couple, no children
13	couple household, dependent children
14	couple household, independent children
15	couple household, no children

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable HHTYPF has been used to derive the variable in the same way. This variable is not available in 1979-83.

phhtype ‘Household type A (Grouped)’

-5	missing
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, one or both aged 60+

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series dataset.

For the additional years 2005-2007, the variable HHTYPA has been used to derive the variable in the same way.

pillness ‘Any longstanding illness or disability’

-5	missing
1	yes
2	no

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable ILLNESS has been used to derive the variable in the same way.

plimitac ‘If longstanding illness limits activity’

-5	missing
1	yes
2	no

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable LIMITACT has been used to derive the variable in the same way.

pmcob1 ‘Mothers country of birth’

-5	missing
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

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable MCOB1 has been used to derive the variable in the same way.

pno ‘Person number of the respondent in the household’

A unique number for each person in the household.

Variables used in the construction of **pno**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS		CPC code
	1979 to 1982	1983 to 2007	**pno**
	PERNO	PERSNO	
1-13			

pnumveh ‘Number of cars’

-5	missing
0	no car available
1	one car
2	two cars
3	three or more cars

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable NUMCARS has been used to derive the variable in the same way.

pregnant ‘Woman pregnant at time of interview’

1	yes
2	no
-8	not known
-9	DNA

Whether the respondent is pregnant, among eligible women.

This variable was filtered on several criteria. A table with eligibility is available in Annex F with the eligibility for the whole contraception/fertility intentions section.
B*efore 1986* women aged less than 45 and eligible for the FI section
From 1986 onwards women aged less than 50 and eligible for the FI section

Variables used in the construction of **pregnant:**

PREGNANT (always the same across years), but also **age**, **eligfi** and **sex**.
A subset of those eligible for the FI section is asked this question.

```
if year<=1985 then do;
        if sex ne 2 or eligfi ne 1 or age>=45 then pregnant=-9;
        else if pregnant notin (1,2) then pregnant=-8;
end;
if year>1985 then do;
        if sex ne 2 or eligfi ne 1 or age>=50 then pregnant=-9;
        else if pregnant notin (1,2) then pregnant=-8;
end;
```

preltohr ‘Relationship to head of household /household reference person’

-5	missing
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

This variable is carried over from the variable of the same name of the 1972-2004 GHS time-series dataset. For the additional years 2005-2007, the variable RELTOHRP has been used to derive the variable in the same way.

probmore ‘Further probe on more children for ‘don’t knows’ on morechld’

1	you will probably have any more children (probably yes)
2	you will probably not have any more children (probably not)
-9	DNA
-8	NA

This question is asked of respondents who said “don’t know” to the previous question, **morechld**

This variable equals PROBMORE of GHS, with values of 3 and 9 recoded to -8.

pseghrp ‘Socio-economic group of HOH/HRP’

-5	missing
1	employers; large
2	managers; large
3	employers; small
4	managers; small
5	prof-self employed
6	prof-employee
7	interm non-man anc
8	interm 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

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable HRPNWSEG has been used to derive the variable in the same way.

ptenure ‘Tenure’

-5	missing
1	owns outright
2	buying with a mortgage
3	private Rented
4	social Rented

For the years 1979-2004, when the GHS 1972-2004 time series and the CPC GHS time series datasets overlap, this variable is carried over from the variable of the same name in the 1972-2004 time series.

For the additional years 2005-2007, the variable TENURE has been used to derive the variable in the same way.

sampq ‘Sample quarter’

1	First quarter
2	Second quarter
3	Third quarter
4	Fourth quarter
-9	not available that year

The calendar month corresponding to these codes differs according to whether the survey is carried out on a calendar year or a financial year basis. For calendar years, code 1 = January to March quarter. For financial years, code 1 = April to June quarter.

The 1998 GHS data file from ESDS was found to have an error in codes 3 and 4 and an attempt has been made to correct these using the month of interview. All interviews either coded 3 originally or coded 4 and taking place in November and December 1998 are now coded 3. The rest of those originally coded 4 retain the code 4. This is not exactly accurate as interviews scheduled for a particular quarter sometimes take place in a later quarter. So a few of those coded 4 in the 1998 survey may in fact belong to quarter 3 but the interview took place in January to March 1999.

Variables used in the construction of **sampq**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979 to 1996	1998 to 2004	2004 to 2007	**sampq**
	none	SAMPQTR	SAMPQTR	
First quarter		1	1	**1**
Second quarter		2	2	**2**
Third quarter		3	3	**3**
Fourth quarter		4	4	**4**
Not available that year	no value			**-9**
Missing (NA)		Any missing codes ever used in GHS	Any missing codes ever used in GHS	**-8**

Note that 2004-05 was the last year in which the survey was fielded on a financial-year basis, and the original dataset deposited with the ESDS Data Archive for this year includes the first quarter of 2005. In 2005, the survey changed back to a calendar year basis, and a full year’s sample size was collected in quarters 2, 3, and 4 of 2005. In the ESDS version of the 2005 GHS, the last quarter of 2004-5 is duplicated as the first quarter of 2005. Because the first quarter of 2005 is already present in the 2004-5 dataset, we have excluded it from the 2005 data in our time series dataset. The first quarter of 2005 can be identified in the 2004-5 dataset via **sampq**=4.

Therefore to extract the original 2004-5 round data from the CPC dataset, choose year = 2004, and **sampq** = 1 or 2 or 3 or 4

To extract the data deposited as the 2005 round with the Archive, choose **year** = 2004 and **sampq** = 4 AND **year** = 2005 and **sampq** = 2 or 3 or 4

sched ‘Full or proxy interview’

1 Full interview
2 Proxy
3 Eligible but non-contact or refusal
4 Child under 16
5 No information available for this year (1979 only)

Variables used in the construction of **sched**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1980 to 1982	1983 to 2001	2002 to 2006	**sched**
	COMPLETE	SCHEDTYP	SCHEDTYP	
Fully Co-op	1			**1**
Eligible, proxy	2			**2**
Eligible, non-contact	3			**3**
Eligible, refusal	4			**3**
Child strictly under 16	Not coded	Not coded	0	**4**
Full interview		1	1	**1**
Proxy		2	2	**2**
Eligible but NC or Refusal		3	3	**3**
Missing (NA)	*missing code and not a child*	*missing code and not a child*	*missing code and not a child*	***-8***

Note that in 1979, no variable is available on which to base **sched**. Thus for that year, **sched** is coded 5.

selfcomr ‘Self completion of the FI section’

1 self-completion
2 not self-completion
-8 not known
-9 DNA

This variable applies only to respondents who were eligible for the FI section. Given the complexity of the initial variables, the table has been simplified and the codes only are given, not the labels.

Variables used in the construction of **selfcomr**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS						CPC code
	1979 to 1982	1983 to 1985	1989 and 1990 to 1993	1989 to 1993	1994 to 2002	2003 to 2007	**selfcomr**
	FISELF	CHECKB	FAMINFSG	FAMINFSG	SELFCOM3	SELFCOM3	
Self completion	1	1	1, 3, 5, 7, 9, 11, 13, 15	1, 3, 5, 7	2, 3		**1**
Self completion laptop						1	**11**
Self completion paper						2, 3, 5	**12**
Not self completion	2, 3	2, 3, 4, 5	2, 4, 6, 8, 10, 12, 14, 16	2, 4, 6, 8	1	4	**2**
Refused whole or partial section			17, 18	9, 10	5	6	**3**
Does not apply (DNA)	-6, -9	-6, -9	-6, -9	-6, -9	4, -6, -9	-6, -9	**-9**
Missing (NA)	missing code	missing code	missing code	missing code	6, missing code	missing code	**-8**

sex ‘Sex of the respondent’

1 male
2 female

The **sex** variable is consistent across time, and indicates the sex of the respondent.

Variables used in the construction of **sex**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS	CPC code
	1979 to 2007	**sex**
	SEX	
Male	1	**1**
female	2	**2**
Missing (NA)	Any missing codes ever used in GHS	**-8**

sexbth1-sexbth15 'Sex of the 1st-15th live birth in the original birth history'

1 male
2 female

Sex of live births to respondent

Variables used in the construction of **sexbth1-sexbth15**
n indicates any ranges on which the variables are available the different years

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1979 to 1982	1983 to 1996	1998 to 2007	
	CHDSEX1-*n*	SEX1-*n*	BIRTHSEX BIRTHSE2-9 BIRTHS10-*n*	**sexbth1-sexbth15**
Male	1	1	1	**1**
Female	2	2	2	**2**
Missing (NA)	Missing codes in GHS	Missing codes in GHS	-8	**-8**
Does not apply (DNA)	Missing codes in GHS	Missing codes in GHS	-6, -9	**-9**

sexbthr1-sexbthr15 ‘Sex of the 1st-15th live birth in the revised birth history’

1	male
2	female

Sex of live births to respondent in the revised birth history

sterild ‘Whether has been sterilized’

1	yes, woman
2	yes, man
3	yes, both
4	no
-9	DNA
-8	NA

The question is asked only of women aged less than 50. As for all questions in the contraception section, it is asked in selected years only, as given in the table below.

A table on eligibility is given in Annex F with the eligibility for the whole contraception/fertility intentions section.

Variables used in the construction of **sterild**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS			CPC code
	1983; 1984; 1986; 1987	1989	1991; 1993; 1995; 1998; 2002	**sterild**
	STERILSD + WHOSTLSD	STERCHKA	STERILA + STERILB + WHOSTLSD	
yes, woman	1+1		1+DNA+1	**1**
yes, man	1+2		1+DNA+2	**2**
yes, both	1+3		1+DNA+3	**3**
no	2+any		(2 or 2) + DNA	**4**
Missing (NA)	.	.	. or -8	**-8**
Does not apply (DNA)	.	.	. or -9	**-9**

The variable combines the variables on sterilisation and the information on which partner has been sterilised.

From 1991 onwards, answers were coded separately for women with a partner (STERILA) and without partner (STERILB). We thus group the two variables.
The variable available for 1989, STERCHKA, has been constructed in the original GHS using the questions on contraception. Users have to keep in mind that the way it is constructed differs from the other years, since no specific question on sterilization has been asked.

In all years where a contraception section exists, except 1989, a direct question was asked about whether the respondent and her partner were sterilized. In 1989 the information appears as a code on a question about current contraception.

totchld ‘Total intended family size’

1 to 25	range of the total number of children wanted
-9	DNA
-8	NA

Only asked of those who either intend to have at least one more child, or don’t know, or, from 1991 onwards, answered “probably yes” to **morechld** or **probmore**.

Variables used in the construction of **totchld**

Value labels/names for the codes on the original GHS dataset	Codes from each year of the GHS				CPC code
	1979 to 1982	1983 to 1985	1986 to 1990	1991 to 2007	**totchld**
	morechld CHLDTOT	**morechld** TOTCHILD	**morechld** TOTCHLD	**morechld** TOTCHLD	
range 1 to 25	CHLDTOT	TOTCHILD	TOTCHLD	TOTCHLD	**1 to 25**
Missing (NA)					**-8**
Does not apply (DNA)	when **morechld** not eq.1	when **morechld** not eq.1	when **morechld** not eq.1 and not eq.3	when **morechld** not eq.11 and not eq.12 and **probmore** not eq.1	**-9**

wave ‘Wave’

1	First wave
2	Second wave
3	Third wave
4	Fourth wave

This is the wave variable exactly as in the original datasets. This variable is used to represent the stage at which respondents are in the rolling sample from 2005. Those to be dropped in the following year are coded 4, those due to remain over three more years are coded 1. In 2005, all respondents are new. In 2006 all respondents, except those coded wave=4 in 2005, continue in the sample. For respondents new in a particular round, **wave** takes value 1, for the others the wave variable is incremented by 1. This applies between 2006 and 2009 rounds.

weightn ‘Original ONS weights’

Available for years 1996 to 2009 only.

Weights generated by CPC

Because ONS provided weights from 1996 only, CPC generated two principal sets of weights applicable to the entire 1979-2007 dataset. Variable names starting with **wgtcpc** refer to weights generated by CPC. See Beaujouan, Brown and Ní Bhrolcháin (2011) for details. New weights were produced in exactly the same way for 2008 and 2009.

All normalised weights are normalised by survey year; hence the sum of the normalised weights for a given GHS survey round equals the sample size in that round.

wgtcpc 'Weights for all individuals in the CPC time-series'

wgtcpcstd 'Weights for all individuals in the CPC time-series, normalised'

wgtcpcfi 'Weights for individuals answering the Family Information section in the CPC time-series'

wgtcpcfistd 'Weights for individuals answering the Family Information section in the CPC time-series, normalised and trimmed at 3'

The weights appropriate for different types of analysis are:

Samples analysed	**CPC weights**	**Referred to in Beaujouan, Brown and Ní Bhrolcháin (2011) as:**
General samples	**wgtcpc**, **wgtcpcstd**	CPC-ALL
Respondents to the Family Information Section	**wgtcpcfi**, **wgtcpcfistd**	CPC-FI

We recommend to use the normalised weights (normalisation is done by year), as the size of the sample varies from one year to the other.

year ‘Year of the survey’

1979	1979
1980	1980
1981	1981
1982	1982
1983	1983
1984	1984
1985	1985
1986	1986
1987	1987
1988	1988-89
1989	1989-90
1990	1990-91
1991	1991-92
1992	1992-93
1993	1993-94
1994	1994-95
1995	1995-96
1996	1996-97
1998	1998-99
2000	2000-01
2001	2001-02
2002	2002-03
2003	2003-04
2004	2004-05
2005	2005
2006	2006
2007	2007

The variable **year** gives the calendar year of the survey round (1979-87, and 2005-2007) or the first year of the two-year financial year span (e.g. 1988 for 1988-9 from 1988-9 to 2004-5). This is done in order to simplify the reading of the tables and to have a numeric variable, but of course the whole dataset corresponding to the original survey round, whether on a calendar or financial year basis, is present for that year.

ANNEX A QUALITY CHECKS AND EDITS CARRIED OUT ON BIRTH HISTORIES 1979-2007[4]

Organisation of birth history files for cleaning

We first built a data file, extracted from the original GHS files for the years 1979-2007 with variables relevant to birth histories and cleaning. The variables were made consistent across years with standardized value labels. This original set of fertility histories is available in the deposited CPC dataset.

However, it was clear that a number of births had not been reported within the original fertility histories and so these original histories have been augmented using information on children living in the household. (See the end of this Annex and Ni Bhrolchain, Beaujouan and Murphy (2011) for further details.) It is this augmented file which was used as the basis for the data checks reported below. Variables from the CPC revised birth history appear in the dataset postscripted with an r.

Checks and edits for birth history

General principles

- In general, we count anomalies and list the frequency of anomalies by survey year.
- Valid date defined as: year present and <= year of interview and >= date of birth
- For convenience, dates of birth are referred to in this document as DOB; it will be clear from the context whether the date of birth in question relates to the respondent herself or to live births in the fertility history.

To summarize, the checks and edits are as follows:

(a) Checks on the year and month of live births
   - ⇨ order
   - ⇨ number of live births declared by the respondent and count of the number of dates of birth in the birth history
   - ⇨ frequency of missing month or year
   - ⇨ consistency with DOB and age of parents and with date of interview

The same checks have been carried out for the age of mother at live births as for the dates of live births, but tables in this document are confined to checks on the live birth dates.

Remark

It seems that the mother's age at birth has been edited by ONS since 1988 (i.e. they have corrected some ages at birth when the childbearing history was not consistent) but the associated dates of live births were not edited correspondingly. In the original

[4] Quality flags were generated for the 2008 and 2009 rounds when these were added but they have not been examined closely; the tabulations in this Annex do not include the 2008 and 2009 data.

1996 data file, for some individuals the sequence of age of the mother at births were in reverse order, but the corresponding dates of birth were in the correct order (i.e. earliest birth first etc.). Sometimes, the dates of birth are out of order in the original datasets.

The age of mother at birth(s) (and all the variables that go with them) have been reordered when in complete reverse order. The years of birth (and all the variables that go with them if not already reversed in the process of reordering age of mother at birth(s)) have been reordered. For the births of children, the sequence of year of birth is updated using these corrected sequences of ages of mother at birth and the month of birth of the child when it exists. Before 1988 where we do not have the age of mother at birth and in 1996 where age sequences are problematic, the original years of live births are retained.
One special check is done to see how many individuals' records have been edited. In all cases, checks are applied again after the edits described above. As a result, the tables below are in pairs, giving the results of the pre-edit and post-edit checks.

Following the editing process, two kinds of anomalies remain:

- years of live birth that are out of order (but not in exact reverse order), in very few cases that already had other amendments.
- difference between the declared number of births and number of dates of births actually present on the file.

Detail of the checks

Check 1
Declared number of births in FI section = number of valid dates of birth.
(0) same number of births declared as number of valid birth dates
(1) no valid birth dates despite at least one birth declared
(2) number of valid dates of birth > number of births declared > 0
(3) 0 < number of valid dates of birth < number of births declared

Table BH1 gives the number of women in each GHS year for whom the number of dates of live birth differs from the declared number of births. We see that the number of people who do not give enough dates of births of children compared with their number of children is especially high in 1982 and 2006.

Table before edit BH1

year	checknb					
		same nb births declared and nb dob	nb births declared>0 but no dob	nb births declared<dob	nb births declared>dob	Total
1979	2753	6003	0	0	1	8757
1980	2565	6202	0	0	1	8768
1981	2549	6528	0	0	1	9078
1982	2106	5153	10	0	339	7608
1983	2038	5294	0	0	0	7332
1984	1778	5251	0	0	0	7029
1985	1849	5364	0	0	0	7213
1986	147	7205	0	0	0	7352
1987	163	7346	0	0	0	7509
1988	150	7073	0	0	0	7223
1989	151	7087	0	0	0	7238
1990	89	6677	0	0	0	6766
1991	97	6928	0	0	0	7025
1992	88	6975	0	0	0	7063
1993	102	6618	0	0	0	6720
1994	139	6462	2	0	2	6605
1995	84	6455	0	0	0	6539
1996	91	6105	0	0	0	6196
1998	36	5555	5	0	9	5605
2000	79	5299	8	0	10	5396
2001	135	5906	9	1	12	6063
2002	203	5596	8	0	15	5822
2003	226	6516	6	1	24	6773
2004	217	5655	3	0	15	5890
2005	258	6640	14	0	20	6932
2006	738	5492	87	1	10	6328
2007	429	5371	0	0	4	5804
Total	19260	166756	152	3	463	186634

Corrections

The 339 histories with fewer years of live birth than declared number of live births in 1982 are due to missing values for year of live birth in the middle of the birth history series.

Between 1988 and 2007:

- As noted above, it seems that the birth histories were cleaned by ONS, but they only edited the ages of the mother at birth, and left the original, unedited, and presumably mistaken, year of birth on the file.
- we have therefore recalculated the year of live birth using the recorded month of birth together with the date of birth of the mother, thus replacing the existing year when it is missing or in error.

Table after edit BH1c

year	checknbl					
		same nb declared and nb dob	nb births but no dob	nb births declared<dob	nb births declared>dob	Total
1979	2753	6003	0	0	1	8757
1980	2565	6202	0	0	1	8768
1981	2549	6528	0	0	1	9078
1982	2106	5422	5	0	75	7608
1983	2038	5294	0	0	0	7332
1984	1778	5251	0	0	0	7029
1985	1849	5364	0	0	0	7213
1986	147	7205	0	0	0	7352
1987	163	7346	0	0	0	7509
1988	150	7069	0	0	4	7223
1989	151	7085	0	0	2	7238
1990	89	6677	0	0	0	6766
1991	97	6926	0	0	2	7025
1992	88	6972	0	0	3	7063
1993	102	6617	0	0	1	6720
1994	139	6460	4	0	2	6605
1995	84	6455	0	0	0	6539
1996	91	6105	0	0	0	6196
1998	36	5551	8	0	10	5605
2000	78	5297	11	0	10	5396
2001	135	5904	12	0	12	6063
2002	203	5595	12	0	12	5822
2003	226	6522	11	0	14	6773
2004	217	5658	8	0	7	5890
2005	258	6642	21	0	11	6932
2006	585	5645	88	1	9	6328
2007	350	5451	3	0	0	5804
Total	19027	167246	183	1	177	186634

Comment

The number of missing years of live births has increased between Table BH1 and BH1c. This is because when there was a mistake in the reported age of the mother at birth, ONS suppressed the age of mother at birth but left the incorrect year of the child's birth. The edit carried out restores a few problematic dates previously suppressed.

The declared number of births is not, as we see, always equal to the number of birth slots appearing on the file. We have no basis for choosing between these two counts of live births, and so we retain on the database the original declared number of births, although it may be inconsistent with the number of birth slots for some respondents.

Concerning the second last column (declared number of births> number of dates of birth), this occurs either (1) when there are gaps in the sequence of births; or (2) when the maximum number of birth slots provided for in the file is not high enough compared to the real maximum declared number of births. However the limitation due to the number of slots is very low and begins in 2001-02, as seen in the table BH1d below.

Table BH1d

	resp. whose nb children too high compared with nb dob allowed	max allowed
1979	0	12
1980	0	12
1981	0	11
1982	0	9
1983	0	10
1984	0	8
1985	0	12
1986	0	13
1987	0	13
1988-89	0	14
1989-90	0	12
1990-91	0	10
1991-92	0	12
1992-93	0	11
1993-94	0	11
1994-95	0	10
1995-96	0	10
1996-97	0	9
1998-99	0	10
2000-01	0	10
2001-02	4	**7**
2002-03	7	**7**
2003-04	11	**7**
2004-05	8	**7**
2005	9	**7**
2006	4	**7**
2007	4	**7**

Check 2

DOB respondent<dates of live birth<=date of interview

We created a quality indicator for each birth: '0'='no problem' '1'='born before parent' '2'='born before parent was 10' '3'='born after interview'.

Table BH2 shows that there are somewhat more inconsistencies from 2001 onwards in the dates of live birth relative to the mother's date of birth.

Table before edit BH2

year	chpar				Total
	no problem	born before parent	born before parent was 10	born after interview	
1979	30705	0	0	0	30705
1980	31443	0	0	0	31443
1981	32410	0	0	0	32410
1982	27160	0	0	0	27160
1983	26587	0	0	0	26587
1984	25354	0	0	0	25354
1985	25555	0	0	0	25555
1986	26073	0	0	0	26073
1987	26419	0	0	0	26419
1988	25350	0	0	0	25350
1989	25269	0	0	0	25269
1990	23663	0	0	0	23663
1991	24657	0	0	0	24657
1992	24534	0	0	1	24535
1993	24079	0	0	0	24079
1994	23620	0	2	0	23622
1995	23385	0	0	0	23385
1996	22274	0	0	0	22274
1998	20394	1	0	1	20396
2000	19266	0	0	0	19266
2001	21110	49	13	8	21180
2002	20130	4	6	9	20149
2003	24489	0	0	0	24489
2004	20419	2	0	0	20421
2005	25070	14	0	9	25093
2006	22913	2	2	7	22924
2007	21470	2	0	0	21472
Total	663798	74	23	35	663930

Problem in 2001

Many dates of birth of children are quite outside the valid limits and some of them are obviously mistyped.

Edits

- same general edits as in Check 1 using the ages of mother at birth already edited by ONS
- in addition and as far as possible, we attribute a corrected year to the birth. Two methods of attribution are used:
    - obvious data entry problems such as a year 2987 are replaced with e.g. 1987
    - the years of birth of the siblings are also used to derive a likely year

Table after correction BH2c

year	chparb				Total
	no problem	born before parent	born before parent was 10	born after interview	
1979	30705	0	0	0	30705
1980	31443	0	0	0	31443
1981	32410	0	0	0	32410
1982	27160	0	0	0	27160
1983	26587	0	0	0	26587
1984	25354	0	0	0	25354
1985	25555	0	0	0	25555
1986	26073	0	0	0	26073
1987	26419	0	0	0	26419
1988	25350	0	0	0	25350
1989	25269	0	0	0	25269
1990	23663	0	0	0	23663
1991	24657	0	0	0	24657
1992	24535	0	0	0	24535
1993	24079	0	0	0	24079
1994	23622	0	0	0	23622
1995	23385	0	0	0	23385
1996	22274	0	0	0	22274
1998	20396	0	0	0	20396
2000	19266	0	0	0	19266
2001	21180	0	0	0	21180
2002	20149	0	0	0	20149
2003	24489	0	0	0	24489
2004	20421	0	0	0	20421
2005	25093	0	0	0	25093
2006	22921	0	2	1	22924
2007	21471	1	0	0	21472
Total	663926	1	2	1	663930

Check 3-4

3- Where date of live birth (i+1) exists, does date of live birth (i) exist?

4- Chronological sequence of years of birth dates correct? : i.e. date of live birth (1) <= date of live birth (2) <= date of live birth (3) etc.

Quality indicator

'0'='ordered no problem' '1'='missings in the middle' '2'='not ordered'

Table BH3 shows that 1982 is a problematic year, as explained before, with missing dates of births in the middle of many birth histories. The dates of births are more often out of sequence from 1998 onwards than earlier.

Table before edit BH3

year	flag			Total
	ordered no problem	missings in the middle	not ordered if no missing	
1979	4246	0	1	4247
1980	4378	0	0	4378
1981	4558	0	0	4558
1982	3486	268	0	3754
1983	3726	0	0	3726
1984	3641	0	0	3641
1985	3623	0	0	3623
1986	4804	0	0	4804
1987	4984	0	0	4984
1988	4782	0	0	4782
1989	4826	0	0	4826
1990	4599	0	0	4599
1991	4775	0	0	4775
1992	4780	0	0	4780
1993	4624	0	0	4624
1994	4467	0	1	4468
1995	4519	0	0	4519
1996	4288	0	1	4289
1998	3808	6	10	3824
2000	3450	5	3	3458
2001	3856	5	15	3876
2002	3722	3	9	3734
2003	4039	6	0	4045
2004	3486	5	2	3493
2005	4250	5	38	4293
2006	3886	4	17	3907
2007	3623	0	0	3623
Total	113226	307	97	113630

Problems

For the missing values in the middle of the series of dates of live birth in 1982, the month, the sex, the age of the mother at birth and other variables concerning the children are often present, nevertheless.

Remark

Strangely, in the original file the year of birth of first, second… birth is a derived variable, while month etc. are original variables.

Edits

- The main edit described at the beginning of this document (replace date of birth with the ONS-edited age of mother at birth), resolved most of those classified (1) in Table BH3c and so an additional edit was not required for this particular issue
- For those classified (2) we reorder the births in ascending order, together with all associated variables.

Table after edit BH3c

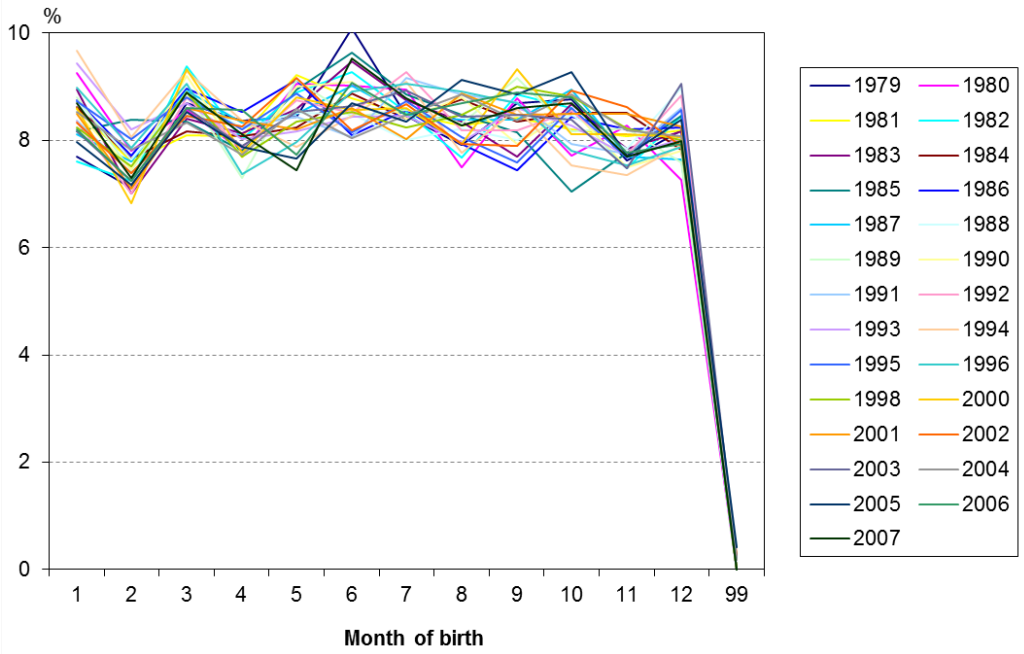
year	flagb			Total
	ordered no problem	missings in the middle	not ordered if no missing	
1979	8757	0	0	8757
1980	8768	0	0	8768
1981	9078	0	0	9078
1982	7542	64	2	7608
1983	7332	0	0	7332
1984	7029	0	0	7029
1985	7213	0	0	7213
1986	7352	0	0	7352
1987	7509	0	0	7509
1988	7223	0	0	7223
1989	7238	0	0	7238
1990	6766	0	0	6766
1991	7025	0	0	7025
1992	7062	0	1	7063
1993	6720	0	0	6720
1994	6605	0	0	6605
1995	6539	0	0	6539
1996	6196	0	0	6196
1998	5598	7	0	5605
2000	5391	5	0	5396
2001	6054	9	0	6063
2002	5816	6	0	5822
2003	6764	8	1	6773
2004	5885	5	0	5890
2005	6927	4	1	6932
2006	6324	3	1	6328
2007	5804	0	0	5804
Total	186517	111	6	186634

Check 5
Frequency of missing months in dates of live birth.

No cases were found of a missing month when a valid year was present, and so no edits were required.

The distribution of month of live births was also checked and month 6 is not especially more frequent than other months, and so we infer that there is not a great deal of imputation to June. Thus the missing months, if they have been imputed by ONS, were distributed across the calendar year. In the original GHS data files the imputation of any part of the dates of live births is not flagged.

Graph BH1 Distribution of month of birth for first births



Following the five checks and edits detailed above, a temporary woman-level quality flag for the birth history, called bthhist[5] in the tables below, was computed, as follows:

BTHHIST
(1) no anomalies in birth history, no edits required
(2) no anomalies in birth history after edits
(2) no date of live birth, but declared number of births >=1
(3) number of dates of live birth >=1 and > declared number of births or < declared number of births
(4) some invalid/missing dates of live birth; inconsistencies in history, cannot be resolved, unusable
(-9) DNA: not eligible FI, proxy

Remark
From table BH4c, some errors remain in 1982 that we couldn't correct.

In 2006, the especially high number of people not reporting births while they had some cannot be corrected. This is due to a small issue with the transition from a cross-sectional to a longitudinal design, and either these persons, whose most were present in 2005, have not given information to update their 2005 records, or their birth history has not been retrieved correctly from the previous year (where the dates of living births were present). The errors are not due to the erroneous recoding of some males to female in the second+ wave of the 2005-6 round, nor to problems with those who have just turned 16 and so were newly eligible for the FI section in 2006, but not in 2005.

[5] Note bthhist is a working variable and does not appear in the final dataset.

Table BH4c

year	brthhist				Total
	no anomalie	no dob while births declared	too few or too many dob	missing in the middle	
1979	8756	0	1	0	8757
1980	8767	0	1	0	8768
1981	9077	0	1	0	9078
1982	7528	5	11	64	7608
1983	7332	0	0	0	7332
1984	7029	0	0	0	7029
1985	7213	0	0	0	7213
1986	7352	0	0	0	7352
1987	7509	0	0	0	7509
1988	7219	0	4	0	7223
1989	7236	0	2	0	7238
1990	6766	0	0	0	6766
1991	7023	0	2	0	7025
1992	7060	0	3	0	7063
1993	6719	0	1	0	6720
1994	6599	4	2	0	6605
1995	6539	0	0	0	6539
1996	6196	0	0	0	6196
1998	5587	8	3	7	5605
2000	5375	11	5	5	5396
2001	6039	12	3	9	6063
2002	5798	12	6	6	5822
2003	6748	11	6	8	6773
2004	5875	8	2	5	5890
2005	6900	21	7	4	6932
2006	6231	88	6	3	6328
2007	5801	3	0	0	5804
Total	186274	183	66	111	186634

Recovered birth history

We describe here the procedure by which own children have been recovered in the GHS from 1994 onwards. First, we give an outline, then give details of the recovery, present some information on the numbers of recovered children and finally provide details of a consistency check against existing GHS derived variables.

Outline

In addition to the dated birth history collected in the Family Information (FI) section of the questionnaire ("original fertility history"), we have constructed a birth history which includes both the live births declared by women in the Family Information section, and those "own children" recorded in the household who were not declared in the FI section. We refer to these as "recovered" births, and we refer to the amended birth history that includes them as the "recovered birth history". The recovery has been carried out from1994 onwards only. This is because it is only from 1994 on that a full matrix of relationships between household members is available from which the relationship between a mother and child can be identified unambiguously. Recovered birth histories therefore differ from the original birth histories from 1994 on only; for 1979-1993, the two sets of histories are identical. Prior to 1994, an own child reconstruction is in principle possible using each person's relationship to head of household, but that has not been attempted thus far (the birth histories did not appear to have major deficiencies up to 1994).

Our initial finding was that some women who had declared no births in the FI section had in fact previously reported own children in the household, at the household interview stage. We also found that the number of a woman's own children in the household could sometimes be greater than the number of births declared in the FI section. In exploiting this information, three situations arise:

- For women declared childless in the FI section, attribute to them the live births deduced from the presence of own children in the household. That is, when there is/are own child(ren) in the household, assign their dates of birth as the dates of live birth of the woman. Note that in these cases, any children who are not present in the household are not identified, and so are missing from the recovered birth history.
- For women who have declared >=1 live birth in the FI section and have the same number of own children, or fewer, in the household, compare the dates of birth of the children in the household and the dates in the FI section birth history to see whether the children are the same, or appear to be different children. Recover the children when their birth dates are different from those given in the FI section, following the criteria set out below.
- For women who have declared >=1 live birth in the FI and have a greater number of own children in the household, do the same and recover the additional child/children, subject to the criteria set out below.

Detail of the recovery

<u>Identifying own children in household</u>

Each year since 1994, either the household or individual file gives a household grid that codes the relationship of each individual in the household to every other member of the household (before 1998: REL02T01 etc.; after 1998: RELTO01-14). These

include a code of "3" for "child", including natural or adopted children; step and foster children are separately identified and coded. We define as a woman's "own child": any child coded "3" in the relationship to the woman. In each year from 1994 on, we recovered these own children in the household, together with their characteristics from the household section of the questionnaire. The relationships are not checked to be reciprocal – i.e. that a child is coded "child" of the woman and that the women is coded "mother" of the child. The matrix is however constructed symmetrically at the GHS data processing stage. In 1998, just one half of the matrix of relationships was coded (R01-09), and we have accordingly constructed the other half from it.

The characteristics recovered are:
- month and year of birth
- age at survey
- sex of the child
- child present in the household

Adding the children in the household to the recovered fertility history when absent from the original fertility history in the FI section

We merge the vector of own children in the household with the vector of children declared in the FI section. The three situations identified above raise different issues.

- For women childless according to the original fertility history (according to BABY/NUMBABY in the FI section), all own children in the household are considered as live births.
- Among women declaring 1+ live births in the original fertility history (FI), several cases arise:

- one or more dates of birth of own children in the household and of children in the FI section are very similar: either the year of birth is the same but the month of birth is 9 months or less different; or the month of birth is exactly the same but the year is one year different: in these cases, we consider these are the same children in the two sources. The dates of birth of the children declared in the FI section are preferred to those given at the household interview stage, because we consider them more likely to be accurate
- a date of birth is very different in the two series (i.e. not meeting the criteria of similarity identified above): we consider that this child was indeed not declared in the FI section and in this case, the child and associated characteristics are added to the vector of the FI section children, forming the "recovered vector".

The table below gives the proportion of women who had at least one own child added by the recovery procedure to the original fertility history given in the Family Information section.

Table BHR2: Proportion of women with at least one recovered own child by original parity

NBIRTHS	1994	1995	1996	1998	2000	2001	2002	2003	2004	2005	2006	2007
-8	35.3	35.7	20.9	36.1	37.2	37.0	38.9	32.7	36.9	38.0	34.7	52.0
0	2.9	1.5	1.9	4.3	9.9	6.3	4.9	16.7	12.9	11.2	13.6	14.2
1	0.5	1.1	0.4	1.5	0.4	1.0	1.1	1.7	1.2	2.0	4.4	2.2
2	0.5	0.4	0.2	0.7	0.3	1.3	0.2	0.8	1.2	1.0	1.4	1.4
3	0.6	0.3	1.0	0.7	0.3	0.9	0.4	1.0	1.1	1.6	1.4	1.4
4	0.8	0.0	0.3	1.1	0.9	0.0	0.8	1.2	1.7	0.4	2.8	1.0
5	0.9	1.9	3.1	2.7	0.0	0.0	1.1	0.0	2.9	1.3	0.0	1.7
6	2.2	2.3	0.0	0.0	0.0	3.6	5.6	0.0	4.3	0.0	0.0	0.0
7	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.1	0.0
Total	2.0	1.3	1.2	2.2	4.2	3.7	3.4	7.9	7.0	6.2	8.9	9.1

Consistency check

To verify our recovery procedure, we have compared a subset of the children recovered from the household, those under 16, with derived variables on the GHS files. Recovered children under 16 are used here for comparability with the GHS derived variables. We have compared these with

CHNFUOWN: number of children aged under 16 in the family unit (GHS documentation says that this includes stepchildren)

DPCHOWNA: number of dependent children in the household (children aged under 16 or 16-18 and in full-time education)

We would expect that our count of own children < 16 should be less than or equal to these variables, as CHNFUOWN may include step children and DPCHOWNA will include some children 16+ who are in full time education. Our recovery is almost perfectly in line with these variables (we show only the tabulation against CHNFUOWN below). Hence, we believe our procedure to be accurate.

Table BHR1 Number of own children aged less than 16 recovered from the household (biological or adopted), compared with the number of dependent children less than 16 in the Family unit given by the CHNFUOWN variable (own + step children)

nchhh15	CHNFUOWN(NO.OF OWN CHILDREN (<16) IN FAMILY UNIT)															Total
	.	-9	-8	-6	0	1	2	3	4	5	6	7	8	9	10	
-8	217	683	11	583	243	9	2	1	0	0	0	0	0	0	0	1749
0	8557	19621	109	5240	8425	213	64	14	3	0	0	1	0	0	0	42247
1	2	3	5	3	1	12843	150	29	3	0	0	1	0	0	0	13040
2	0	8	4	0	0	8	11936	83	16	2	0	0	0	0	0	12057
3	0	2	1	0	0	2	2	3704	22	1	1	0	0	0	0	3735
4	0	0	0	0	0	0	0	1	865	4	2	0	0	0	0	872
5	0	0	0	0	0	0	0	0	0	202	0	0	0	0	0	202
6	0	0	0	0	0	0	0	0	0	0	31	0	0	1	0	32
7	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	9
8	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	6
9	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	3
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Total	8776	20317	130	5826	8669	13075	12154	3832	909	209	34	11	6	4	1	73953

For further details and an assessment of the revised birth histories, see Ni Bhrolchain, Beaujouan and Murphy (2011) “Sources of error…”, Population Studies, 65 (3), 305-318.

ANNEX B QUALITY CHECKS AND EDITS CARRIED OUT ON MARRIAGE HISTORIES 1979-1998[6]

The type of information collected on current and retrospective partnerships has developed over the life time of the GHS, reflecting the increasing complexity of individuals' life course trajectories. This is shown in outline in Table 1. .

In summary the following information was collected:

1979-1980 information on current and previous marriages

1981-1988 duration of premarital cohabitation before the current marriage was added

1989-1998 premarital cohabitation prior to all previous marriages as well as the current one added

2000-2007 information on up to three periods of cohabitation that did not end in marriage added

The cleaning the marriage and partnership histories was carried out separately for the years 1979-1998 and 2000-2007. The checks and edits are therefore described in two documents, one for each of these periods.

This is the first of these two documents and describes the checks and edits carried out on the marriage and partnership histories for GHS rounds 1979-1998.

[6] Quality flags were generated for the 2008 and 2009 rounds when these were added but they have not been examined closely. The tabulations in this Annex do not include the 2008 and 2009 data.

Table 1: Summary of type of marriage and partnership information collected within the GHS from 1979 to 2007[1]
X represents information collected.

	Survey year (note no survey was conducted in 1997/98 or in 1999/00)														
Type of information	1979	1981	1983	1985	1987	1989	1991	1993	1995	1998	2000	2001	2003	2005	2007
Current and previous marriage dates	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Duration of current cohabitation	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Whether the respondent cohabited prior to their current/most recent marriage[2]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Length of premarital cohabitation before current marriage		X	X	X	X	X	X	X	X	X	X	X	X	X	X
Length of premarital cohabitation prior to current and previous marriages						X	X	X	X	X	X	X	X	X	X
Start and end dates for closed periods of cohabitation which did not lead to marriage											X	X	X	X	X

[1] Men were asked this information only from 1986
[2] This information was not collected in the 1980 round

Executive Summary

The consistency of dates in the marriage histories for the period 1979-1998 is excellent as seen in Table 2 which shows the numbers of marriages that are either without error or without error following editing. These are shown in four groups of years treated separately due to their differing data structures. In total 0.8% of the marriage histories had some sort of error in them which meant that information on one or more marriage spells is unusable.

Table 2 Number of marriage histories containing an error

	1979-82	1983-87	1988-1996	1998
No error	19115	32479	81180	7063
No error after edit	3	3	3	101
At least one erroneous spell	392	295	322	53
Total marriage histories	19510	32777	81505	7217

For further details, see Berrington et al (2011) Evaluation of the Partnership Histories in the Centre for Population Change GHS Time Series Dataset. CPC Working Paper 12, Centre for Population Change, University of Southampton

Checks on GHS Marriage Histories 1979-1998

Section 1 sets out the checks and edits applied to the marriage histories for the period 1979-1998. The tables that follow present the results of these checks for four blocks of years:
Section 2 1979-1982 (Table 79.1 - Table 79.13b)

Section 3 1983-1987 (Table 83.1- Table 83.11)

Section 4 1988-1996 (Table 88.1 - Table 88.13b)

Section 5 1998 (Table 98.1 – 98.13)

The checks and edits are organised in this way because the names and formats of the variables and the structure of the datasets changes through this period, and so results are presented for blocks of years within which the structure of the data was fairly homogeneous.

1 General principles

We count anomalies and list the frequency of anomalies, sometimes by survey year. This includes cases with a missing month but a valid year of marriage. A valid marriage date is defined as: year present and <= year of interview and >= year of respondent's birth (or after the year of birth when the date is unavailable).

No.	Check	Action
1	Marriage dates < date of interview?	T
2	Is there a valid set of marriage dates: i.e. first = valid, 2nd valid, 3rd valid or not: e.g some valid dates occur later in the sequence than an invalid date?	T
3	Is the chronological sequence of years of marriage dates correct? : i.e. marr date (1) <= marr date (2) <= marr date (3) etc.	T
4	Number of marriages reported = number of valid dates of marriage?	T
5	Frequency of missing months, (a) marriage dates (b) end of marriage dates, defined as the date of separation (where separated and/or divorced) or death. Where valid year, assign 6.5 to missing month. This incorporates a flag into the imputed month. In analysis, truncated months will be used.	T
6	Is the cause of end of marriage missing? Edits: (1) where cause missing if valid date of (a) widowhood, (b) divorce or (c) separation exists, assign widowhood, divorce or separation as cause; (2) where no valid date of end present, assign type of end from stated marital status if this = latest marriage and stated marital status = separated, divorced or widowed; otherwise (3) cause of end = not known (-8)	T
7	Is cause of end of marriage consistent with valid end dates? If cause = separation/divorce/widowhood, do valid dates of separation/separation and divorce/widowhood exist?	T
8	Where date of marriage (i+1) exists, does date of either divorce (i) or widowhood (i) exist?	T
9	For each marriage: date of marr (i) < date of sep (i) < date of div (i) < date marr (i+1) if exists OR date of marr (i) < date wid (i) < date marr (i+1) if exists	T
10	Dates of premarital cohabitation < date of interview	T
11	Frequency of invalid or incomplete dates or premarital cohabitation; missing/invalid month? Missing/invalid year?	T
12	For each marriage: date of start of premarital cohabitation (i) < date of marriage (i)	T
	At end of edits 1-12 (a) the quality of information for each marriage is classified in a set of flags: **marq1-marqn** and (b) the quality of the entire history is summarised in a flag named **marqual**	T by survey year

Key to Action column: L = list cases; T = tabulate frequency by type

2 GHS rounds 1979-1982: marriage history checks and edits

These findings relate to marriage histories for the years 1979-1982.
Note that in the tables that follow, persons who were not eligible for the Family Information section are coded -9.

Check 1
This check evaluates whether event dates are more recent than the date of interview, given a valid year of marriage.

Table 79.1

Marriage order	Number of marriage dates more recent than interview date	N (total cases)
1st marriage	3	19145
2nd marriage	0	1557
3rd marriage	0	43
4th marriage	0	3

The three erroneous dates are where the year of marriage is the same as the year of interview, and the month of marriage is greater than the month of interview. For these cases, the month of marriage has been edited to be the same as the month of interview. The month of marriage is flagged as erroneous dates with 0.5 added to it.

Check 2
This check examines the sequence of validity in marriage histories. There are 13 non-valid marriage dates for women married twice, all 13 being due to missing years of marriage (the month is present, but we are unable to impute).

Table 79.2

Marriage order	Missing valid marriage date 1	Missing valid marriage date 2	Missing valid marriage date 3	Missing valid marriage date 4
1	-	-	-	-
2	13	0	-	-
3	0	0	0	-
4	0	0	0	0

Check 3
This check evaluates whether the marriage start dates are in the correct temporal order. Error flags are defined as where the date of previous marriage (marriage i) occurs more recently than the marriage presented. There are no errors.

Table 79.3

Marriage order	Error flags	Valid cases
2	0	1557
3	0	43
4	0	3

Check 4

This check evaluates whether the number of reported marriages is consistent with the number of marriage start dates. Table 79.4a identifies the number of valid marriages associated with each marriage order (to identify over-reporting), while Table 79.4b identifies the number of missing valid marriage dates at each order (to identify underreporting). There is no over-reporting of the number of marriages. There are 352 people who were married only once and do not have a valid first marriage date. 13 respondents were married twice but do not have a valid first marriage date and 22 respondents who were married twice but have no valid 2nd marriage date. These missing dates are represented by -8 in the dataset and the marriage histories are coded as 3 in the **marqual** variable.

Table 79.4a

	Number of marriages (CPC variable **nmar**)				N (Number of valid marriages)
	1	2	3	4	
Number of valid 1st marriage dates	17579	1503	41	3	19125
Number of valid 2nd order marriage dates	-	1513	41	3	1557
Number of valid 3rd marriage dates	-	-	40	3	43
Number of valid 4th marriage dates	-	-	-	3	3
N (Number of marriage order **nmar**)	17931	1516	41	3	

Table 79.4b

	Number of marriages (CPC variable **nmar**)				N (number of valid marriages)
	1	2	3	4	
Number of missing 1st marriage dates	352	13	0	0	19125
Number of missing 2nd marriage dates	-	22	0	0	1557
Number of missing 3rd marriage dates	-	-	1	0	43
Number of missing 4th marriage dates	-	-	-	0	3
N (Number of marriage order **nmar**)	17931	1535	41	3	

Check 5

This check first examines the presence of months of marriage given the presence of a valid year of marriage. There are no errors as can be seen in Table 79.5a.

Table 79.5a

Marriage order	Number of missing marriage months	N (Number of cases)
1	0	19145
2	0	1557
3	0	43
4	0	3

The second part of the check examines the extent to which the month of marriage may have been imputed as June. When compared to an assumption of equal distribution, the excess of June months is presented in table 79.5b. This indicates that there is only a slight excess of June marriage dates, which probably reflects the true distribution rather than imputation.

Table 79.5b

Marriage	Excess June over uniform (% points)	N (total cases)
1st marriage	2.36	19298
2nd marriage	0.91	1560
3rd marriage	3.29	43
4th marriage	0.00	3

We also examine the number of missing end months. These are presented in 79.5c. We find no errors for this check.

Table 79.5c

Marriage order	Number of missing end months	N (Number of cases)
1	0	19145
2	0	1557
3	0	43
4	0	3

Check 6

This check examines the presence of end causes for marriages given a higher order marriage date. There are no errors for this check

Table 79.6

Marriage order	Number of missing end causes	N (Number of cases)
2	0	1557
3	0	43
4	0	3

Check 7

This check examines the consistency between the presence and type of end date and the type of end cause. There are no missing end dates for marriages that have been declared to have ended.

The only specific cause available with identifiable years is divorce (with independent dates), as there is a decree absolute available. Therefore, it is possible to identify excess dates of divorce, (i.e. a divorce date is presented for marriages ending in death). There are no errors for this check.

Table 79.7

Terminal marriage order	Number of excess divorce dates	Number of marriage with a decree absolute date present (equals number ending in divorce)
1	0	2353
2	0	108
3	0	5

Check 8.

This check ensures that given a valid start date for higher order marriage there is a valid end date for lower order marriages. There are 6 cases where there is a missing valid end date.

Table 79.8

Marriage order	Missing end date for marriage 1	Missing end date for marriage 2	Missing end date for marriage 3
2	6	-	-
3	0	0	-
4	0	0	0

Check 9.

This check ensures that the temporal ordering of the marriage history is correct. Note that erroneous dates of divorce are not counted as errors.

Table 79.9

Marriage order	Marriage start occurs before end of previous marriage	Marriage occurs before divorce for previous marriage	End of this marriage occurs before start of this marriage	Date of divorce for this marriage occurs before end of marriage	N (number of cases)
1	N/A	N/A	0	4	19145
2	0	4	1	0	1557
3	0	1	0	1	43
4	0	0	N/A	N/A	3

Checks 10-12

Note that dates of premarital cohabitation prior to all marriages was not collected over this period. Though we can recover the date of premarital cohabitation for the current marriage, these are derived from durations before the current marriage and therefore are by definition prior to the interview date and are non-negative and hence checks 10-12 are not carried out.

Check 13

This check presents the overall distribution of marriage history quality. There are 392 cases where marriage histories are partially OK, the majority of these occurring in 1982.

Table 79.13a

	Marriage quality	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	All OK	19115	15.7	98.0	98.0
	All OK after editing	3	.0	.0	98.0
	At least one usable spell	392	.3	2.0	100.0
	Total	19510	16.0	100.0	
Missing	-9.00	102208	84.0		
Total		121718	100.0		

Table 79.13b

		All OK	All OK after editing	At least one usable spell	
YEAR OF INTERVIEW	79	4894	0	35	4929
	80	4960	1	63	5024
	81	5156	2	68	5226
	82	4105	0	226	4331
Total		19115	3	392	19510

3 GHS rounds 1983-1987: Marriage history checks and edits

These are the results of the cleaning procedure agreed for the marital histories. These findings relate to histories extracted for the years 1983-87.

Check 1
This check evaluates whether event dates are more recent than the date of interview, given a valid year of marriage.

Table 83.1

Marriage order	Number of marriage dates more recent than interview date	N (total cases)
1st marriage	2	32504
2nd marriage	0	3258
3rd marriage	1	154
4th marriage	0	5

The two erroneous dates on marriage 1 and the one erroneous date on marriage 3 are due to an error on the month. The year of marriage and interview are the same.
Edit: These marriages have been flagged as edited dates (0.5 added to the month) and the month set at the month of the interview.

Check 2
This check examines the set of validity in marriage histories. There are no errors.

Table 83.2

Marriage order	Missing valid marriage date 1	Missing valid marriage date 2	Missing valid marriage date 3	Missing valid marriage date 4	N
2	0	0	-	-	3258
3	0	0	0	-	155
4	0	0	0	0	5

Check 3
This check evaluates whether the marriage start dates are in the correct temporal order. Error flags are defined as where the date of previous marriage (marriage i) occurs more recently than the marriage presented. There are no errors for this check.

Table 83.3

Marriage order	Error flags	Valid cases
2	0	3258
3	0	155
4	0	5

Check 4

This check ensures that the number of reported marriages is consistent with the number of marriage start dates. Table 83.4a identifies the number of valid marriages associated with each marriage order (to identify over reporting), while Table 83.4b identifies the number of missing valid marriage dates at each order (to identify underreporting). There are three cases where the entire marriage history appears to be missing.

Table 83.4a

	Number of marriages (CPC variable **nmar**)				
	1	2	3	4	Total no. with valid date in slot of this order
Number of valid 1st marriage dates	29227	3122	147	8	32099
Number of valid 2nd order marriage dates	-	3103	147	8	3218
Number of valid 3rd marriage dates	-	-	147	8	150
Number of valid 4th marriage dates	-	-	-	5	5
Total declaring this no. of marriages	28881	3122	147	8	

Table 83.4b

	Number of marriages (CPC variable **nmar**)				N
	1	2	3	4	Total no. with invalid date in slot of this order
Number of missing 1st marriage dates	273	0	0	0	3
Number of missing 2nd marriage dates	-	19	0	0	3
Number of missing 3rd marriage dates	-	-	0	0	3
Number of missing 4th marriage dates	-	-	-	3	3
Total declaring this no. of marriages	28881	3122	147	8	

Check 5

This check first examines the presence of months of marriage given the presence of a usable marriage year. There are no errors for this check.

Table 83.5a

Marriage order	Number of missing marriage months	N (Number of cases)
1	0	32504
2	0	3258
3	0	155
4	0	5

The second check examines the extent to which the month of marriage may have been imputed as the month of June. When compared to an assumption of equal distribution, the excess of June months is presented in table 83.5b. This indicates that there is only a slight excess of June marriage dates, which probably reflects true preference rather than imputation.

Table 83.5b

Marriage	Excess June over uniform (% points)	N (total cases)
1st marriage	2.5	32504
2nd marriage	2.1	3258
3rd marriage	1.3	155
4th marriage	-	5

We also examine the number of missing end months. There are no missing months.

Table 83.5c

Marriage order	Number of missing end causes	N (Number of cases)
1	0	6511
2	0	480
3	0	27

Check 6

This check examines the presence of end causes for marriages given a higher order marriage date. There are no errors for this check.

Table 83.6

Marriage order	Number of missing end causes	N (Number of cases)
2	0	3218
3	0	150
4	0	5

Check 7

This check examines the consistency between the presence and type of end date and the type of end cause. There are no missing end dates for marriages that have been declared to have ended.

The only specific cause available with identifiable years is divorce (with independent dates), as there is a decree absolute available. Therefore, it is possible to identify excess dates of divorce, (i.e. a divorce date is presented for marriages ending in death).

Table 83.7

Terminal marriage order	Number of excess divorce dates	Number of marriage with a decree absolute date present (equals number ending in divorce)
1	0	4979
2	0	310
3	0	13

Check 8.

This check ensures that given a valid start date for higher order marriage there is a valid end date for lower order marriages. There are no errors for this check.

Table 83.8

Marriage order	Missing end date for marriage 1	Missing end date for marriage 2	Missing end date for marriage 3	Ns
2	0	-	-	3258
3	0	0	-	155
4	0	0	0	5

Check 9.

This check ensures that the temporal ordering of the marriage history is correct. There is one erroneous case. Note that an erroneous divorce date is not counted as an error.

Table 83.9

Marriage order	Marriage start occurs before end of previous marriage	Marriage occurs before divorce for previous marriage	End of this marriage occurs before start of this marriage	Date of divorce for this marriage occurs before end of marriage	N (number of cases)
1	NA	NA	0	0	32504
2	1	14	0	0	3258
3	0	0	0	0	155
4	0	0	NA	NA	5

Checks 10-12.

Note that dates of premarital cohabitation prior to all marriages was not collected over this period. Though we can recover the date of premarital cohabitation for the current marriage, these are derived from durations before the current marriage and therefore are by definition prior to the interview date and are non-negative and hence checks 10-12 are not presented.

Check 13

Table 83.10

	Marqual	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-9.00	114195	77.7	77.7	77.7
	All OK	32479	22.1	22.1	99.8
	All OK after editing	3	.0	.0	99.8
	At least one usable spell	3	.0	.0	99.8
	Error (inconsistent number)	292	.2	.2	100.0
	Total (persons)	146972	100.0	100.0	

Table 83.11

Count	Marqual*intyear crosstabulation	Intyear					
		1983	1984	1985	1986[7]	1987	Total
marqual	-9.00	26070	23844	25088	19564	19629	114195
	All OK	4134	4080	4120	9982	10163	32479
	All OK after editing	3	0	0	0	0	3
	At least one usable spell	0	0	0	1	2	3
	Error (inconsistent number)	66	70	74	46	36	292
Total		30273	27994	29282	29593	29830	146972

[7] The rise in numbers here is due to men being asked a marriage history from 1986 onwards.

4 GHS rounds 1988-1996: marriage history checks and edits

These are the results of the cleaning procedure agreed for the marital histories. These findings relate to histories extracted for the years 1988-96.

Check 1

This check evaluates whether event dates are more recent than the date of interview.

Table 88.1

Marriage order	Number of marriage dates more recent than interview date	N (total cases)
1st marriage	1	81187
2nd marriage	0	10258
3rd marriage	0	756
4th marriage	0	40
5th marriage	0	2
6th marriage	0	1
7th marriage	0	1

We identify one error in this check. The date of marriage in this case appears to occur 10 months following the date of interview.

Edit: The date of marriage is corrected to the date of interview, with a flag of 0.5' on month.

Check 2

This check tabulates missing dates of marriage by marriage order. We identify 4 cases where there are errors, with missing first marriage dates.

Table 88.2

Marriage order	Missing valid marriage date 1	Missing valid marriage date 2	Missing valid marriage date 3	Missing valid marriage date 4	Missing valid marriage date 5	Missing valid marriage date 6	Cases
2	4	-	-	-	-	-	10258
3	0	0	-	-	-	-	756
4	0	0	0	-	-	-	40
5	0	0	0	0	-	-	2
6	0	0	0	0	0	-	1
7	0	0	0	0	0	0	1

Check 3

This check evaluates whether the marriage start dates are in the correct temporal order. There are no errors.

Table 88.3

Marriage order	Error flags	Valid cases
2	0	10258
3	0	756
4	0	40
5	0	2
6	0	1
7	0	1

Check4

This check evaluates the consistency between the number of reported marriages and the number of completed marital slots. Table 88.4a identifies the number of reported marriages compared with the marriage slots. This table is designed to identify over-reporting (i.e. there are reported slots higher than the reported marriages). We identify no over-reporting from this table. Table 88.4b identifies the number of missing marriage slots given a reported marriage. We identify 3 missing 2nd marriage dates and 4 missing first marriage dates for women married twice, and 55 missing marriage dates for women married once.

Table 88.4a

	1	2	3	4	5	6	7	Total no. with valid date in slot of this order
Number of valid 1st marriage dates	70930	9501	716	38	1	-	1	81187
Number of valid 2nd order marriage dates	-	9502	716	38	1	-	1	10258
Number of valid 3rd marriage dates	-	-	716	38	1	-	1	756
Number of valid 4th marriage dates	-	-	-	38	1	-	1	40
Number of valid 5th marriage dates	-	-	-	-	1	-	1	2
Number of valid 6th marriage dates	-	-	-	-	-	-	1	1
Number of valid 7th marriage dates	-	-	-	-	-	-	1	1
Total declaring this no. of marriages	70985	9505	716	38	1	0	1	

Table 88.4b

	1	2	3	4	5	6	7	Total no. with valid date in slot of this order
Number of missing 1st marriage dates	55	4	0	0	0	-	0	81187
Number of missing 2nd marriage dates	-	3	0	0	0	-	0	10258
Number of missing 3rd marriage dates	-	-	0	0	0	-	0	756
Number of missing 4th marriage dates	-	-	-	0	0	-	0	40
Number of missing 5th marriage dates	-	-	-	-	0	-	0	2
Number of missing 6th marriage dates	-	-	-	-	-	-	0	1
Number of missing 7th marriage dates	-	-	-	-	-	-	0	1
Total declaring this no. of marriages		9505	716	38	1	0	1	

Check 5

This check ensures the completeness of months of events, given the presence of a valid year of event. There is one missing month which is imputed.

Table 88.5.a

Marriage order	Number of missing marriage months	N (Number of cases)
1	1	81187
2	0	10258
3	0	756
4	0	40
5	0	2
6	0	1
7	0	1

The frequency of missing values is also evaluated for terminal events (divorce, separation, death). We find 3 missing divorce months, imputed at 6.5 months.

Table 88.5b

Terminal marriage order	Number of missing (valid) divorce months		Number of missing (valid) separation months		Number of missing (valid) death months	
1	3	16451	0	18861	0	2097
2	0	1668	0	2095	0	230
3	0	119	0	170	0	11
4	0	6	0	12	0	2
5	0	1	0	2	-	-
6	0	1	0	1	-	-
7	-	-	-	-	-	-

Check 6

This check evaluates the completeness of end causes for marriages with a valid terminal date. For example, where a date of divorce exists, a date of separation should also be recorded. We identify no errors in this check.

Table 88.6

Marriage order	Number of missing divorce causes (terminated divorces)		Number of missing death causes (terminated deaths)		Number of missing separation causes (terminated separations)	
1	0	16471	0	2097	0	18861
2	0	1668	0	230	0	2095
3	0	119	0	11	0	170
4	0	6	0	2	0	12
5	0	1	-	-	0	2
6	0	1	-	-	0	1

Check 7

This check ensures that the end dates reported are consistent with the end cause reported. We identify no excess end dates.

Table 88.7

Marriage order	Number of excess dates for divorces (i.e. death dates present)		Number of excess dates for death (i.e. sep. and divorce dates)		Number of excess dates for separations (i.e. divorce and death dates)	
1	0	16451	0	2097	0	18861
2	0	1668	0	230	0	2095
3	0	119	0	11	0	170
4	0	6	0	2	0	12
5	0	1	0	-	0	2
6	0	1	0	-	0	1

Check 8.

This check ensures that if a higher order marriage date exists, a valid end date exists for preceding marriages. There are 9 missing end dates for marriage 1 in total and 1 missing end date for marriage 2.

Table 88.8

Marriage order	Missing end date for marriage 1	Missing end date for marriage 2	Missing end date for marriage 3	Missing end date for marriage 4	Missing end date for marriage 5	Missing end date for marriage 6	N (Number of cases)
2	8	-	-	-	-	-	10258
3	1	1	-	-	-	-	756
4	0	0	0	-	-	-	40
5	0	0	0	0	-	-	2
6	0	0	0	0	0	-	1
7	0	0	0	0	0	0	1

Check 9.

This check ensures that the temporal ordering of the marriage history is correct. There are 4 ‘errors’ in divorce dates, but these are not flagged as errors because we use the date of separation as the end date of the marriage.

Table 88.9

Marriage order	Marriage start occurs before end of previous marriage	Marriage occurs before divorce for previous marriage	End of this marriage occurs before start of this marriage	Date of divorce for this marriage occurs before end of marriage	N (number of cases)
1	N/A	N/A	0	0	81187
2	0	4	0	0	10258
3	0	0	0	0	756
4	0	0	0	0	40
5	0	0	0	0	2
6	0	0	0	0	1
7	0	0	N/A	N/A	1

Check 11

This check ensures that the dates of premarital cohabitations are complete. We assess the completeness of month of cohabitation, year of cohabitation and month and year of cohabitation. The table is truncated at marriage 4 for brevity (no marriage above 6 has a premarital cohabitation).

Table 98.11

Marriage order	Year missing only	Month missing only	Year and month both missing	N (Number of reported premarital cohabitations)
1	12	190	28	15982
2	4	173	27	6660
3	1	18	2	30
4	0	0	1	24

Check 12

This check ensures that the dates of premarital cohabitation proceed the date of marriage. We identify on case where the premarital cohabitation starts after marriage, and this is corrected by imputation.

Table 88.12

Marriage order	Number of premarital cohabitations with start following marriage	N (Number of reported premarital cohabitations)
1	0	15982
2	1	6660
3	0	30
4	0	24

Check 13

This check summarises the overall quality of the marital histories. There are 58 histories for this period which are totally unusable, and 264 spells where there is at least one spell where there is an error on at least one spell.

Table 88.13a

		Frequency	Percent
Valid	-9	135330	62.4
	All OK	81180	37.4
	All OK after editing	3	.0
	At least one usable spell	264	.1
	Error	55	.0
	Error (inconsistent number)	3	.0
	Total	216835	100.0

Table 88.13b

		-9	All OK	All OK after editing	At least one usable spell	Error	Error (inconsistent number)	
Survey year	1988	15603	9637	0	52	55	3	25350
	1989	15501	9724	0	44	0	0	25269
	1990	14665	8971	0	27	0	0	23663
	1991	15158	9471	0	28	0	0	24657
	1992	15171	9335	0	29	0	0	24535
	1993	15157	8905	0	17	0	0	24079
	1994	14993	8592	3	34	0	0	23622
	1995	14784	8579	0	22	0	0	23385
	1996	14298	7966	0	11	0	0	22275
Total		135330	81180	3	264	55	3	216835

5 GHS 1998: marriage history checks and edits

These are the results of the cleaning procedure agreed for the marital histories. These findings relate to histories extracted for the years 1998.

Check 1

This check ensures that none of the reported marriages occur after the date of interview. In this survey year there were 6 missing interview dates**.** For the purposes of this check these cases have been ignored.

Table 98.1

Marriage order	Error flags	N (number of valid cases)
1	1	7201
2	0	1101
3	0	94
4	0	8
5	0	3

There is one marriage event identified where the interview occurred before the marriage. In this case the interview took place in January and the marriage reportedly of June 1998.

Edit: The marriage month is reassigned to the month of interview with 0.5 added to the year and month of marriage.

Check 2

This check examines the completeness of the marriage history, and the presence of preceding valid marriage dates given the existence of a valid marriage date. We identify 10 missing 1st marriage dates, a total of 3 missing 2nd marriage states and 1 missing 3rd marriage date.

Table 98.2

Marriage order	Missing marriage date 1	Missing marriage date 2	Missing marriage date 3	Missing marriage date 4	N
2	10	-	-	-	1101
3	0	2	-	-	94
4	0	1	1	-	8
5	0	0	0	0	3

Check 3

This check ensures that for a given marriage the preceding marriages started at an earlier date. Error flags represent that the previous marriage start date occurs after the current marriage start date. There are no errors.

Table 98.3

Marriage order	Number of error flags	N (number of valid cases)
2	0	1101
3	0	94
4	0	8
5	0	3

Check 4

This check ensures that the number of valid marriages dates and the number of reported marriages is consistent. Table 98.4a reports the number of valid marriage dates reported, to identify over-reporting. Table 98.4b identifies the number of missing marriage dates (underreporting). We identify some underreporting (17 histories), and no over-reporting.

Table 98.4a

Number of reported marriages	Number of valid 1st marriage date	Number of valid 2nd marriage date	Number of valid 3rd marriage date	Number of valid 4th marriage date	Number of valid 5th marriage date
1	5771	-	-	-	-
2	953	962	-	-	-
3	80	78	80	-	-
4	4	3	3	4	-
5	3	3	3	3	3
Missing	390	55	8	1	0

Table 98.4b

Number of reported marriages	Number of missing 1st marriage date	Number of missing 2nd marriage date	Number of missing 3rd marriage date	Number of missing 4th marriage date	Number of missing 5th marriage date
1	2	-	-	-	-
2	10	1	-	-	-
3	0	2	0	-	-
4	0	1	1	0	-
5	0	0	0	0	0
Missing	0	0	0	0	0

Check 5

This check examines the completeness of marriage months given a usable marriage year. There are 34 missing months which are imputed at the values of 6.5.

Table 98.5a

Marriage order	Number of missing marriage months	N (number of valid marriage years)
1	29	7201
2	5	1101
3	0	94
4	0	8
5	0	3

Table 98.5b examines the extent to which the month of marriage may have been influenced by imputation at the month of June. There is no significant preference for the month of June (6) except in the case of marriage orders over 3 where the preference probably reflects the relatively small numbers of cases with over 3 marriages.

Table 98.5b

Marriage order	Excess June over uniform (% points)	N (Number of cases)
1	3.61%	7174
2	3.24%	1097
3	8.69%	94
4	16.67%	8
5	N/A (no observations)	3

Missingness is also evaluated for the end causes of marriages, in table 98.5c. Note that there are no end causes for marriage 5 available.

Table 98.5c

Marriage order	Number of missing divorce months (total N)		Number of missing separation months (total N)		Number of missing death months (total N)	
1	131	(1823)	78	(2057)	0	(172)
2	15	(203)	12	(252)	0	(29)
3	2	(21)	1	(26)	0	(3)
4	0	(3)	0	(3)	0	(1)

Missing separation months are imputed at 6.5.

Check 6

This check ensures that the end cause of dissolved marriages is complete. We cannot impute these ends, since the dates for terminal event are not recorded in the dataset.

Table 98.6

Marriage order	Number of missing causes	N (number of dissolved marriages)
1	3	2349
2	1	306
3	0	31
4	0	4

Check 7

This check ensures that the dates presented for the end of marriage are consistent with the stated end causes. There are no errors.

Table 98.7

Marriage order	Number of marriages ending in divorce with excess end dates	Number of marriages ending in separation with excess end dates	Number of marriages ending in death with excess end dates	N (number of dissolved marriages)
1	0	0	0	2346
2	0	0	0	305
3	0	0	0	31
4	0	0	0	4

Check 8

This check ensures that where a valid higher order marriage date is present, lower marriage orders have a valid end date and cause. The majority of these are associated with marriage 2, with considerably smaller numbers for higher order marriages.

Table 98.8

Marriage order	Number of missing previous end dates	N (number of dissolved marriages)
2	38	1101
3	4	94
4	1	8
5	0	3

Check 9

This check examines the sequences of the entire marriage history. Error flags are presented for each marriage order. Since errors on date of divorce do not contribute to error flags, we identify two error flags. The first is in the second marriage where one marriage end date is earlier than the end and the second error results from the date of start of marriage 3 preceding the date of end of marriage 2. These are flagged as errors in **marqual**.

Table 98.9

Marriage order	Marriage start occurs before end of previous marriage	Marriage occurs before divorce for previous marriage	End of this marriage occurs before start of this marriage	Date of divorce for this marriage occurs before end of marriage	N (number of cases)
1	N/A	N/A	0	19	7201
2	0	9	1	9	1101
3	1	1	0	1	94
4	0	0	0	0	8
5	0	0	N/A	N/A	3

Check 10

This check ensures that the dates of premarital cohabitations do not occur after the date of interview. There are no errors.

Table 98.10

Marriage order	Number of premarital cohabitations after interview date	N (number of valid premarital cohabitation dates)
1	0	2185
2	0	803
3	0	72
4	0	8
5	0	2

Check 11

This check ensures that the dates of premarital cohabitations are complete. We assess the completeness of month of cohabitation, year of cohabitation and month and year of cohabitation. This missingness is conditional on reported premarital cohabitation.

Table 98.11

Marriage order	Year missing only	Month missing only	Year and month both missing	N (Number of reported premarital cohabitations)
1	0	37	11	2196
2	21	1	6	810
3	0	0	1	73
4	0	0	0	8
5	0	0	0	2

Check 12

This check ensures that the dates of premarital cohabitation proceed the date of marriage. There are two cases where this occurs, as evident form table 98.12. This is solved by imputing at 6.5, or at a random months (with an imputation flag) between January and marriage should the marriage occur before June.

Table 98.12

Marriage order	Number of premarital cohabitations with start following marriage	N (Number of reported premarital cohabitations)
1	0	2196
2	2	810
3	0	73
4	0	8
5	0	2

Check 13

The distribution of respondents according to the quality of their overall marriage history is shown in Table 98.13.

Table 98.13

	Marqual	
	Frequency	**Percentage**
-9.	13179	64.6
All OK	7063	34.6
All OK after editing	101	.5
At least one usable spell	48	.2
Error (all invalid)	4	.0
Out of order	1	.0
Total	20396	100.0

ANNEX C QUALITY CHECKS AND EDITS CARRIED OUT ON PARTNERSHIP HISTORIES FROM 2000-2007[8]

The type of information collected on current and retrospective partnerships has developed over the life time of the GHS, reflecting the increasing complexity of individuals' life course trajectories. This is shown in outline in the Table 1 below.

In summary the following information was collected:

1979-1980 information on current and previous marriages

1981-1988 duration of premarital cohabitation before the current marriage was added

1989-1998 premarital cohabitation prior to all previous marriages as well as the current one added

2000-2007 information on up to three periods of cohabitation that did not end in marriage added

The cleaning of the marriage and partnership histories was carried out separately for the years 1979-1998 and 2000-2007. The checks and edits are therefore described in two documents, one for each of these periods.

This is the second of these two documents and describes the checks and edits carried out on the marriage and partnership histories for GHS rounds 2000-2007.

[8] Quality flags were generated for the 2008 and 2009 rounds when these were added but they have not been examined closely. The tabulations in this Annex do not include the 2008 and 2009 data.

Table 1: Summary of type of marriage and partnership information collected within the GHS from 1979 to 2007[1]

X represents information collected.

	Survey year (note no survey was conducted in 1997/98 or in 1999/00)														
Type of information	1979	1981	1983	1985	1987	1989	1991	1993	1995	1998	2000	2001	2003	2005	2007
Current and previous marriage dates	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Duration of current cohabitation	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Whether the respondent cohabited prior to their current marriage	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Length of premarital cohabitation before current marriage		X	X	X	X	X	X	X	X	X	X	X	X	X	X
Length of premarital cohabitation prior to current and previous marriages						X	X	X	X	X	X	X	X	X	X
Start and end dates for closed periods of cohabitation which did not lead to marriage											X	X	X	X	X

[1] Men were asked this information only from 1986

Summary of key findings of cleaning of partnership and marriage histories, GHS 2000-2007.

1. Areas where the data quality are very good

1.1 Start and end dates of marriages

The reporting of start dates (year and month) of marriages, and their ordering relative to other marriages is very good. The main error is in 2004 where there are 116 cases with year of premarital cohabitation present but marriage year missing.

1.2 The reporting of marriage end dates is very good, at least in terms of year, although there are some missing months.

2. Areas where the data quality are reasonable

2.1 Number of partnerships[9] & marriages

- There are a sizeable number of respondents whose reported total number of partnership declared does not tally with the number of available start dates (Table PH4). However, this problem is proportionately small, especially for marriages (1.8% of respondents report more marriages than provide marriage start dates, whilst 7.4% of respondents report more cohabitations than valid dates of start of cohabitations).
- In 2000-01 a few hundred respondents in each survey reported marriage dates without reporting the marriage, whilst in the later surveys there was a tendency for the opposite – for respondents to report marriages without reporting valid dates.
- It must be noted however that some of the mis-match between the number of reported cohabitations and the number of dates is due to respondents only being asked to provide details of their first 3 free-standing cohabitations. For instance in 2000, 117 respondents declared a greater number of cohabitations than valid dates. In 45 cases the respondents were limited in the number of valid dates they could report by the number of slots available.
- To clarify two variables were created for number of partnerships/marriages: the number of marriages/partnerships that are declared by the respondent, and the number of usable marriage/partnership slots.

2.2. Ordering

The ordering of free-standing cohabitations was less good than the ordering of marriages (PH14). Out of sequence cohabitations have been re-ordered.

2.3 Overlapping unions

Between 3% (lower order unions) and 18% (small number of higher order unions) of partnerships, almost all of which are free standing cohabitations) are overlapping, i.e. the first one is reported to have ended after the second one has begun. This seems to us a reasonable level of error given the subject matter and structure of the questionnaire.

[9] The term "partnership" in this document includes, as well as marriages, free-standing cohabitations, that is spells of cohabitation that do not end in marriage to the cohabiting partner.

3. Areas where the data quality are less good

3.1 Missing months

There is a sizeable number of missing months in the dataset. This is particularly the case for marriage separation dates (especially higher order marriages) and cohabitation start and end dates (Table PH5). These have been imputed as detailed later in this document.

3.2 No information on number of partnerships

Since 2000, there is an apparent increase in the extent of missing declarations of numbers of marriages and partnerships (**npar** or **nmar**). The same period also displays a rise in missing values in relation to the fertility histories: that is numbers coded NA on variables in the fertility history among people who neither refuse the Family Information section nor are proxies. For example, in 2007, 899 respondents have missing data (8.3% of the sample): that is, 8.3% of respondents who were eligible for the Family Information section are coded missing (-8, NA) on all the key partnership history variables.[10]

4. Overall quality

After completion of the editing and re-ordering of the partnership histories is complete, quality indicators were generated for each marriage (**marq1-marq7**) and each partnership (**partq1-partq7**) of all respondents.

The overall quality is very good. For example, just 1000 of approximately 57,000 first marriages reported are not coded "ok" or "ok after editing" (and just 187 out of 8806 second marriages). Only about 1100 of around 70,000 reported first partnerships are not coded "ok" or "ok after editing"; the figure is 258 of 18,788 reported second partnerships.

Overall the quality of the reported histories is good. After completion of the edits and re-ordering of the partnership histories, an indicator representing the quality of the marriage history as a whole was generated (**marqual**) and an indicator of quality of the partnership history as a whole (**partqual**). Approximately 56,000 marriage histories are "ok" out of around 58,000 histories reported (Table MH21q). About 66,600 partnership histories are "ok" out of a total of around 70,000 reported partnership histories (Table PH21q).

[10] One possibility is that this may be due to the introduction of telephone interviewing, but the cause has not been established at the time of writing this document.

Cleaning of the partnership and marriage histories for 2000-2007

These are the results of the cleaning procedure agreed for the partnership and marital histories for the GHS years 2000-2007. Cleaning and edits to the marriage and partnership histories for GHS rounds 1979-1998 are described in the preceding Annex.

Definitions

Valid date defined as: year present and <= year of interview and >= date of birth of the respondent

Key to Action column:

L = list cases; T = tabulate number by type

Event:

- for marriage = premarital cohabitation, marriage, stop living together/death of the partner, divorce.
- for cohabitation = beginning of the cohabitation, end of the cohabitation (stop living together and/or end of the relationship)

Checks 1-8 are carried out on all partnerships. In some cases the checks are shown separately for marriages (which may have been preceded by cohabitation) and free-standing cohabitations.

Checks 11 to 19 are carried out in two series: first for all partnerships including those that involve a marriage (i.e. group 3 on page 15), and second only for partnerships that include a marriage (i.e. group 2 on page 14). The corresponding tables are labelled PH and MH respectively.

Through the document, "." and "missing" will be used interchangeably to identify missing values. Mostly we do not distinguish DNA and NA.

No.	Checks before re-ordering and corrections	Action
1	year of events < year of interview age at events > age 10	T
2	Is there a valid sequence of marriage/cohabitation years or not: i.e. 1st = valid, 2nd=valid, 3rd=valid; Otherwise, what is the problem: years not all valid, series out of sequence? Chronological sequence of years of marriage dates correct? : i.e. marr date (1) <= marr date (2) <= marr date (3) etc. If not classify as follows: (1) in chronological sequence; (2) out of sequence.	T
3	Where date of marriage (i+1) exists, does date of either divorce (i) or widowhood (i) exist? In that case, do we always have year end (i)<year beg (i+1)	T by year
4	Number of marriages/cohabitations reported =number of valid years of marriage/start cohabitation?	T by year
5	Number of missing months, for each type of event. Where valid year, assign 6.5 to missing month. This therefore constitutes a flag into the imputed month. In analysis, truncated months will be used. Cases with imputed months are checked for consistency re other dates and where problems occur e.g. zero durations of a partnership, further appropriate imputation is done.	T for each event and for each order of relationship

No.	Checks before re-ordering and corrections	Action
6	For each marriage: year of premar cohab (i) if exists < year of marr (i) < year of sep (i) < year of div (i) if exists OR year of marr (i) < year widow (i) if exists For each cohabitation: year of cohab (i) < year of sep (i)	T by year and order
7	Are the declarations of event consistent with valid end years? For instance, if cause = separation/divorce/widowhood, do valid dates of separation/separation and divorce/widowhood exist? Conversely, is the cause reported missing?	T by order
8	same by year	T by year

No.	Checks after re-ordering and corrections	Action
11	Comparison number of partnership declared and number of dates declared	T by year
12	Distribution of the types of beginning/end by year	T by year and order
13	Is there a valid date of beginning/end when a partnership is supposed to begin/end?	T by year and order
14	Check the ordering of beginnings, ends, marriages. Comparison before and after re-ordering	T by year
15	Valid dates of beginning/end? If not, other valid dates?	T by year and order
16	For each marriage: year of premarr (i) if exists < year of marr (i) < year of sep (i) < year of div (i) if exists OR year of marr (i) < year widow (i) if exists For each cohabitation: year of cohab (i) < year of sep (i)	T by year and order
17	Where date of marriage (i+1) exists, does date of beginning/end or divorce(i) exist? In that case, do we always have year event (i)<year beg (i+1)	T by year
18	Number of overlapping marriages/partnerships: once the marriages/partnerships are ordered by date at beginning, do some relationships overlap with the beginning of the next one?	T by year and order
19	Do some relationships have the same date of beginning, or end, or both?	T by year and order

No.	Quality and edit flags	Action
21	After the edits, classify each individual (see user guide): - *Indicator of quality specific to each marriage* **marq1-marq7** - *Indicator of quality specific to each cohabitation* **partq1-partq7**	Tabulate by year
22	After the edits, classify each individual (see user guide): - *Global indicator of quality of the marriage history* **marqual** - *Global indicator of quality of the relationship history* **partqual**	Tabulate by year

Detail of the checks

Check 1
This check evaluates whether

- event dates occur earlier than the date of interview
- events occur after age 10
- events occur after the birth of the eligible person

Notes:

- the absence, in some of these tables, of a column that appears in others denotes that no cases of the kind in question occur in that table.
- for selected years, a question was asked not only about the end of cohabitation but also about whether the end of cohabitation was the end of the relationship or just the end of the couple's living together; the date of whichever of these was not originally given was then asked. For that reason, three types of end of cohabitation are tabulated: the overall end of cohabitation, the end of the relationship, and the end of living together.

Tables PH1

year premarital cohabitation			
year	no problem	before age 10	Total
2000	10233	0	10233
2001	11338	0	11338
2002	10934	0	10934
2003	12792	0	12792
2004	11057	0	11057
2005	12975	0	12975
2006	11708	1	11709
2007	10836	0	10836
Total	91873	1	91874

year marriage					
year	no problem	before birth	before age 10	after interview	Total
2000	10232	0	1	0	10233
2001	11338	0	0	0	11338
2002	10932	0	0	2	10934
2003	12792	0	0	0	12792
2004	11057	0	0	0	11057
2005	12974	0	1	0	12975
2006	11703	1	5	0	11709
2007	10832	2	2	0	10836
Total	91860	3	9	2	91874

year death husb/wife		
year	no problem	Total
2000	10233	10233
2001	11338	11338
2002	10934	10934
2003	12792	12792
2004	11057	11057
2005	12975	12975
2006	11709	11709
2007	10836	10836
Total	91874	91874

year separation of marriage			
year	no problem	before birth	Total
2000	10233	0	10233
2001	11338	0	11338
2002	10934	0	10934
2003	12792	0	12792
2004	11057	0	11057
2005	12975	0	12975
2006	11708	1	11709
2007	10836	0	10836
Total	91873	1	91874

year divorce			
year	no problem	after interview	Total
2000	10233	0	10233
2001	11338	0	11338
2002	10934	0	10934
2003	12791	1	12792
2004	11057	0	11057
2005	12975	0	12975
2006	11709	0	11709
2007	10836	0	10836
Total	91873	1	91874

year beginning cohabitation					
year	no problem	before birth	before age 10	after interview	Total
2000	10232	0	1	0	10233
2001	11337	0	0	1	11338
2002	10934	0	0	0	10934
2003	12790	0	2	0	12792
2004	11056	0	1	0	11057
2005	12974	0	1	0	12975
2006	11703	1	5	0	11709
2007	10835	0	1	0	10836
Total	91861	1	11	1	91874

year end cohabitation					
year	no problem	before birth	before age 10	after interview	Total
2000	10231	0	1	1	10233
2001	11335	0	0	3	11338
2002	10934	0	0	0	10934
2003	12785	0	0	7	12792
2004	11053	0	0	4	11057
2005	12975	0	0	0	12975
2006	11705	1	3	0	11709
2007	10836	0	0	0	10836
Total	91854	1	4	15	91874

year end living together			
year	no problem	after interview	Total
2000	10233	0	10233
2001	11338	0	11338
2002	10934	0	10934
2003	12792	0	12792
2004	11056	1	11057
2005	12975	0	12975
2006	11709	0	11709
2007	10836	0	10836
Total	91873	1	91874

year end relationship				
year	no problem	before age 10	after interview	Total
2000	10233	0	0	10233
2001	11337	0	1	11338
2002	10934	0	0	10934
2003	12792	0	0	12792
2004	11054	0	3	11057
2005	12972	0	3	12975
2006	11708	1	0	11709
2007	10836	0	0	10836
Total	91866	1	7	91874

A few small edits have been applied manually to these deficient dates. The tables above do not reflect these minor edits.

Check 2

This check evaluates whether the marriage and cohabitation (without marriage) start dates are in the correct temporal order (column “ordered no problem”), only when there is no missing in the middle of the series of start dates (column “missings in the middle”). Respondents with only one marriage or cohabitation are included in the tables.

We define error flags in this check as where the start date of previous marriage/cohabitation (start marriage/cohabitation i) occurs at a later date than the marriage / cohabitation that follows it (start marriage/cohabitation i+1).

Tables PH2

beginning of marriages					
year	.	ordered no problem	missings in the middle	out of sequence if no missing	Total
2000	3459	6757	16	1	10233
2001	3973	7336	29	0	11338
2002	3999	6908	27	0	10934
2003	4759	8008	25	0	12792
2004	4381	6653	23	0	11057
2005	4900	8051	24	0	12975
2006	4675	7012	20	2	11709
2007	4672	6149	14	1	10836
Total	34818	56874	178	4	91874

beginning of cohabitations					
year	.	ordered no problem	missings in the middle	out of sequence if no missing	Total
2000	8943	1269	13	8	10233
2001	9996	1321	12	9	11338
2002	9467	1448	13	6	10934
2003	10913	1827	33	19	12792
2004	9518	1513	10	16	11057
2005	11069	1880	9	17	12975
2006	10149	1515	11	34	11709
2007	9493	1308	7	28	10836
Total	79548	12081	108	137	91874

The edit following this check is to reorder when the series is out of sequence and this is done later in the process. There is no action when there are missing values in the middle.

Check 3

This check evaluates whether the next marriage begins after the end date of the previous one, and does the same for free-standing cohabitations. This check is done before re-ordering the out of order sequences that are identified in check 2. Additionally, it includes overlapping slots, parts of which are edited manually afterwards.

Tables PH3

end/beg marriages					
year	.	ordered no problem	missings in the middle	end after beginning of next rel	Total
2000	3459	6747	19	8	10233
2001	3973	7328	29	8	11338
2002	3999	6892	43	0	10934
2003	4759	8012	21	0	12792
2004	4381	6655	18	3	11057
2005	4900	8046	28	1	12975
2006	4675	7001	29	4	11709
2007	4672	6132	29	3	10836
Total	34818	56813	216	27	91874

end/beg cohabitations					
year	.	ordered no problem	missings in the middle	end after beginning of next rel	Total
2000	8943	1252	12	26	10233
2001	9996	1303	11	28	11338
2002	9467	1431	13	23	10934
2003	10913	1809	20	50	12792
2004	9518	1497	8	34	11057
2005	11069	1850	7	49	12975
2006	10149	1493	8	59	11709
2007	9493	1279	7	57	10836
Total	79548	11914	86	326	91874

Check 4

This check examines the consistency between the number of reported marriages/cohabitations and the number of valid beginning dates.

Tables PH4

number of marriages declared vs number valid years of marriage						
year	.	no problem	not enough valid marriage years	not enough marriages declared	missing no. marriage but valid years	Total
2000	3448	6380	31	0	374	10233
2001	3964	6961	41	0	372	11338
2002	3982	6820	48	0	84	10934
2003	4709	7906	85	0	92	12792
2004	4198	6622	216	0	21	11057
2005	4805	8038	128	4	0	12975
2006	4616	6880	191	2	20	11709
2007	4414	6069	333	2	18	10836
Total	34136	55676	1073	8	981	91874

number of cohab declared vs number valid years of cohabitations				
year	.	no problem	not enough valid cohab years	Total
2000	8873	1243	117	10233
2001	9923	1299	116	11338
2002	9390	1434	110	10934
2003	10825	1798	169	12792
2004	9463	1495	99	11057
2005	11014	1850	111	12975
2006	10082	1506	121	11709
2007	9417	1299	120	10836
Total	78987	11924	963	91874

Marriages of people who have a missing number of marriages but give valid years are accepted as proper marriages of the individual.

As seen below, the number of valid years of cohabitation is partly limited by the number of slots available for cohabitation dates (limited to 3). The maximum number of marriages occurring does not exceed the number of slots available.

Tables PH4b

year	ncoh				Total
	-8	0	1-3	4+	
2000	160	7802	2226	45	10233
2001	324	8540	2432	42	11338
2002	432	7978	2494	30	10934
2003	543	9081	3110	58	12792
2004	510	7942	2562	43	11057
2005	578	9167	3167	63	12975
2006	1186	7796	2677	50	11709
2007	1277	7106	2410	43	10836
Total	5010	65412	21078	374	91874

For instance in 2000, 117 persons declare too many cohabitation dates, but among them 45 were limited by the number of slots available.

Check 5

This checks the completeness of months of events, given the presence of a valid year of event. It also gives the total number of valid dates relating to relationships, assuming that missing months will be attributed.

Tables PH5

	premarital cohabitation				
	order 1	order 2	order 3	order 4	order 5
valid year and month	21261	6305	594	46	2
valid year but month missing	885	345	36	4	0
total valid years	22146	6650	630	50	2

	year marriage				
	order 1	order 2	order 3	order 4	order 5
valid year and month	55640	8739	782	59	3
valid year but month missing	1260	93	12	0	0
total valid years	56900	8832	794	59	3

	year death partner			
	order 1	order 2	order 3	order 4
valid year and month	1079	174	19	1
valid year but month missing	16	6	1	0
total valid years	1095	180	20	1

	separation of marriage				
	order 1	order 2	order 3	order 4	order 5
valid year and month	15469	1909	170	12	2
valid year but month missing	916	154	19	3	0
total valid years	16385	2063	189	15	2

	divorce				
	order 1	order 2	order 3	order 4	order 5
valid year and month	12490	1411	118	10	2
valid year but month missing	1519	212	26	1	0
total valid years	14009	1623	144	11	2

	beginning cohabitation		
	order 1	order 2	order 3
valid year and month	11305	2818	713
valid year but month missing	917	217	49
total valid years	12222	3035	762

	end cohabitation		
	order 1	order 2	order 3
valid year and month	11314	2817	713
valid year but month missing	916	216	52
total valid years	12230	3033	765

	end living together		
	order 1	order 2	order 3
valid year and month	2235	511	144
valid year but month missing	139	29	7
total valid years	2374	540	151

	end relationship		
	order 1	order 2	order 3
valid year and month	1687	408	107
valid year but month missing	92	17	5
total valid years	1779	425	112

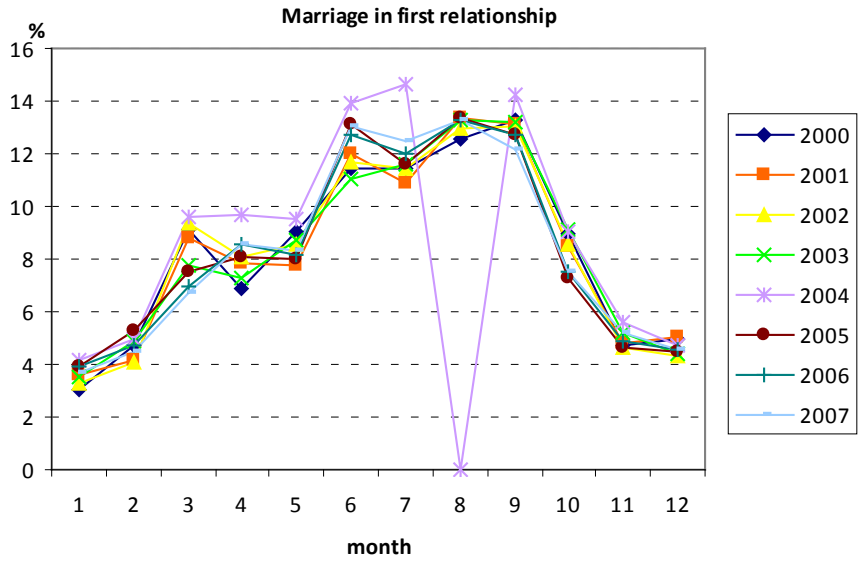
The check below examines the extent to which the month of marriage may have been influenced by imputation of the month of June. When compared to an assumption of equal distribution, there is strong evidence of an excess of June months (6), apart from marriage months. Additionally, in 2004 it appears there is an error in the registration

of the month of marriage, since nobody has a month of August (8). It seems those August marriage months were miscoded -8 and so where the year of marriage is present in 2004, these -8 values on month have been recoded to 8. (The sample sizes of the right order of magnitude).

Table PH5b: Distribution of the months of each event

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
Start of 1st rel.	6.6	5.9	8.7	8.0	7.8	13.7	9.6	9.7	11.1	7.8	5.8	5.4
Start of 2nd rel.	9.7	7.0	7.3	8.1	8.4	13.4	8.2	8.7	8.7	7.5	6.8	6.3
End of 1st rel.	9.2	7.0	8.3	7.3	7.5	16.5	7.8	8.3	8.0	6.9	6.2	6.8
End of 2nd rel.	9.4	7.7	8.2	7.7	7.3	13.8	8.3	8.5	7.7	7.2	6.5	7.7
Marriage in 1st rel.	3.6	4.7	8.2	8.1	8.5	12.3	11.9	11.8	13.0	8.3	4.9	4.6
Marriage in 2nd rel.	3.7	4.7	7.0	8.0	9.6	11.7	11.0	12.0	11.9	8.6	5.7	6.0
Divorce in 1st rel.	7.3	6.3	7.7	7.6	6.7	19.5	7.9	7.6	8.6	7.3	7.2	6.5
Divorce in 2nd rel.	6.8	7.7	8.0	6.5	6.9	17.0	7.9	7.9	8.7	6.6	7.5	8.5

Figure PH5b: distribution of the months of marriage in first relationship



Check 6

This checks whether the temporal ordering of events within the marriage/cohabitation slots is correct.

Tables PH6

order within marriage 1						
year	.	year of divorce but not of marriage	ordered no problem	beginning and end permuted	divorce before separation	Total
2000	3466	7	6746	2	10	2
2001	3990	8	7326	0	14	0
2002	4016	7	6900	2	9	0
2003	4765	16	7991	1	19	0
2004	4373	29	6640	1	14	0
2005	4909	11	8043	1	11	0
2006	4679	14	6994	9	13	0
2007	4679	5	6138	6	8	0
Total	34877	97	56778	22	98	2

order within marriage 2						
year	.	year of divorce but not of marriage	ordered no problem	beginning and end permuted	divorce before separation	Total
2000	9205	3	1023	0	2	0
2001	10194	5	1137	1	1	0
2002	9823	0	1110	0	1	0
2003	11584	7	1196	3	2	0
2004	9945	6	1106	0	0	0
2005	11720	6	1247	0	2	0
2006	10631	1	1073	0	3	1
2007	9908	4	920	1	2	1
Total	83010	32	8812	5	13	2

order within marriage 3				
year	.	year of divorce but not of marriage	ordered no problem	Total
2000	10148	0	85	0
2001	11231	0	106	1
2002	10841	0	93	0
2003	12677	0	115	0
2004	10957	1	99	0
2005	12860	0	115	0
2006	11618	0	91	0
2007	10747	0	89	0
Total	91079	1	793	1

order within marriage 4			
year	.	year of divorce but not of marriage	Total
2000	10226	0	7
2001	11327	1	10
2002	10925	0	9
2003	12786	0	6
2004	11047	0	10
2005	12970	0	5
2006	11704	0	5
2007	10829	0	7
Total	91814	1	59

order within marriage 5			
year	.	ordered no problem	Total
2000	10233	0	10233
2001	11337	1	11338
2002	10933	1	10934
2003	12792	0	12792
2004	11056	1	11057
2005	12975	0	12975
2006	11709	0	11709
2007	10836	0	10836
Total	91871	3	91874

order within cohabitation 1				
year	.	ordered no problem	beginning and end permuted	Total
2000	8955	1275	3	10233
2001	10008	1325	5	11338
2002	9481	1451	2	10934
2003	10944	1841	7	12792
2004	9526	1531	0	11057
2005	11078	1893	4	12975
2006	10160	1546	3	11709
2007	9500	1336	0	10836
Total	79652	12198	24	91874

order within cohabitation 2				
year	.	ordered no problem	beginning and end permuted	Total
2000	9937	296	0	10233
2001	11018	317	3	11338
2002	10592	342	0	10934
2003	12299	491	2	12792
2004	10705	350	2	11057
2005	12479	495	1	12975
2006	11303	404	2	11709
2007	10506	329	1	10836
Total	88839	3024	11	91874

order within cohabitation 3				
year	.	ordered no problem	beginning and end permuted	Total
2000	10151	81	1	10233
2001	11258	80	0	11338
2002	10865	69	0	10934
2003	12674	118	0	12792
2004	10969	88	0	11057
2005	12838	137	0	12975
2006	11607	102	0	11709
2007	10750	85	1	10836
Total	91112	760	2	91874

Only obvious errors have been corrected manually, but beginning and end have not been reversed routinely when they were found in incorrect order. It would be making a strong assumption that the dates have actually been permuted, while the error could result from one or both dates being mistaken.

Check 7

This evaluates the consistency between the reported end cause and the type of end date present.

For premarital cohabitation, when a person answers yes to the question “Before getting married, did you and your husband/wife live together as a couple?”, we consider they declare a premarital cohabitation. Thus if they do not have a valid date of premarital cohabitation for that marriage, we consider the date is missing (in the “declaration but no valid year” row). Similarly, we use the answer to the question on the type of end of marriage to see whether the marriage ended and how, and then to see whether the end date is missing or not.
For marriage and cohabitation there is no direct question, but a question on the number of marriages/cohabitations the person has had. Then we use this at each order to suppose whether the person had or not a marriage/cohabitation and whether there is a valid date that corresponds.

Tables PH7

	premarital cohabitation				
	order 1	order 2	order 3	order 4	order 5
declar. and year coincide	22143	6649	630	50	2
declar. but no valid year	264	111	12	1	0
valid year but no declar.	3	1	0	0	0

	marriage				
	order 1	order 2	order 3	order 4	order 5
declar. and year coincide	55922	8680	774	56	3
declar. but no valid year	811	362	71	11	0
valid year but no declar.	978	152	20	3	0

	death husband			
	order 1	order 2	order 3	order 4
declar. and year coincide	1095	180	20	1
declar. but no valid year	0	0	0	0
valid year but no declar.	13	2	0	0

	separation of marriage				
	order 1	order 2	order 3	order 4	order 5
declar. and year coincide	16385	2063	189	15	2
declar. but no valid year	779	168	22	1	0
valid year but no declar.	0	0	0	0	0

	divorce				
	order 1	order 2	order 3	order 4	order 5
declar. and year coincide	14009	1623	144	11	2
declar. but no valid year	963	183	12	0	0
valid year but no declar.	0	0	0	0	0

	beginning cohabitation		
	order 1	order 2	order 3
declar. and year coincide	12222	3035	762
declar. but no valid year	665	225	108
valid year but no declar.	0	0	0

	end cohabitation		
	order 1	order 2	order 3
declar. and year coincide	12213	3019	761
declar. but no valid year	444	128	63
valid year but no declar.	12657	14	4

	end living together		
	order 1	order 2	order 3
declar. and year coincide	2373	540	151
declar. but no valid year	161	48	23
valid year but no declar.	2534	0	0

	end of the relationship		
	order 1	order 2	order 3
declar. and year coincide	1779	425	112
declar. but no valid year	140	25	13
valid year but no declar.	0	0	0

Edits
(1) where cause is missing: if valid date of (a) widowhood, (b) divorce or (c) separation exists, assign widowhood, divorce or separation as cause;
(2) where no valid date of end present, assign type of end from stated marital status if this = latest marriage and stated marital status = separated, divorced or widowed; otherwise (3) cause of end = not known (-8)

Check 8

Shows the consistency of end causes declared and the presence of end dates.

Tables PH8

marriage ending in death			
year	no problem	declar. but no valid year	Total
2000	10232	1	10233
2001	11337	1	11338
2002	10932	2	10934
2003	12786	6	12792
2004	11056	1	11057
2005	12972	3	12975
2006	11709	0	11709
2007	10835	1	10836
Total	91859	15	91874

divorce			
year	no problem	declar. but no valid year	Total
2000	10117	116	10233
2001	11173	165	11338
2002	10764	170	10934
2003	12661	131	12792
2004	10943	114	11057
2005	12836	139	12975
2006	11601	108	11709
2007	10742	94	10836
Total	90837	1037	91874

separation of marriage			
year	no problem	declar. but no valid year	Total
2000	10069	164	10233
2001	11156	182	11338
2002	10817	117	10934
2003	12660	132	12792
2004	11000	57	11057
2005	12871	104	12975
2006	11625	84	11709
2007	10758	78	10836
Total	90956	918	91874

partners stopped living together				
year	no problem	declar. but no valid year	year but no declar.	Total
2000	9610	288	335	10233
2001	10732	254	352	11338
2002	10236	275	423	10934
2003	11896	355	541	12792
2004	10594	246	217	11057
2005	12460	286	229	12975
2006	11248	238	223	11709
2007	10433	203	200	10836
Total	87209	2145	2520	91874

relationship ended				
year	no problem	declar. but no valid year	year but no declar.	Total
2000	9594	369	270	10233
2001	10728	373	237	11338
2002	10215	457	262	10934
2003	11893	559	340	12792
2004	10607	227	223	11057
2005	12472	241	262	12975
2006	11258	237	214	11709
2007	10441	215	180	10836
Total	87208	2678	1988	91874

global end cohabitation				
year	no problem	declar. but no valid year	year but no declar.	Total
2000	10173	60	0	10233
2001	11282	55	1	11338
2002	10869	65	0	10934
2003	12725	59	8	12792
2004	11011	42	4	11057
2005	12925	40	10	12975
2006	11646	57	6	11709
2007	10774	57	5	10836
Total	91405	435	34	91874

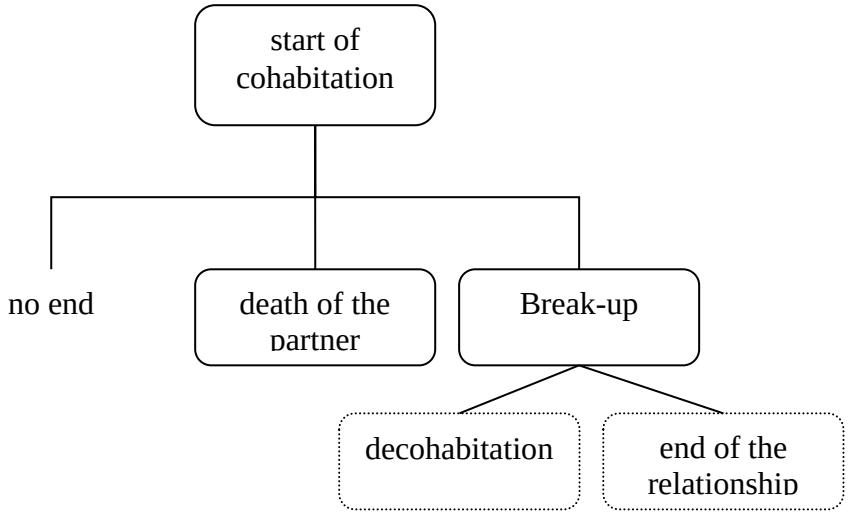
Reordering the partnership histories

Following the above checks, the partnership histories were reordered, and a further set of checks was run.

In all cases, reordering is based on the earliest year within any set of partnership dates.

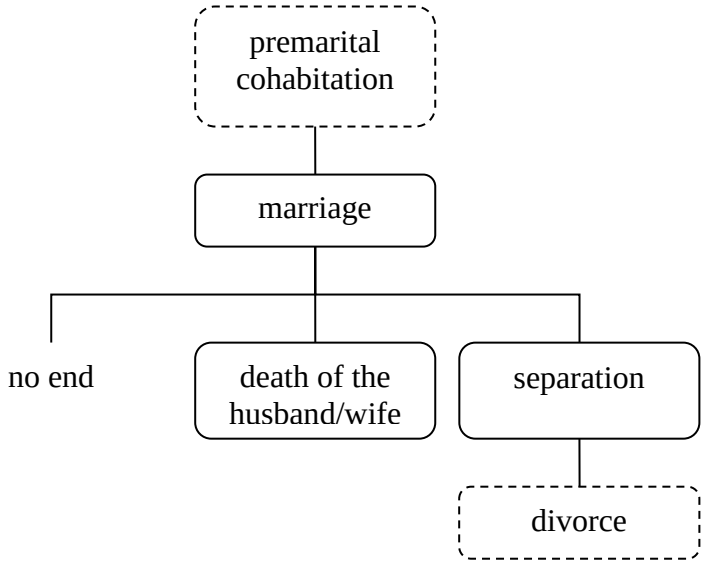
General process

1- Vector of dates concerning non married relationships

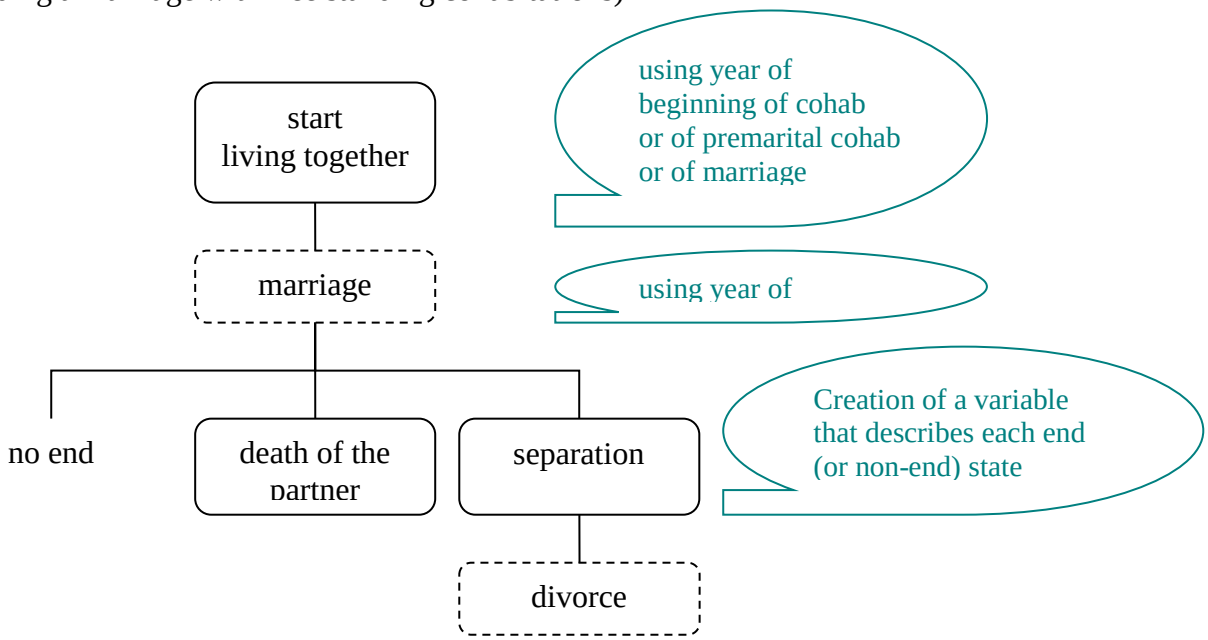


As described in check 1, the date of end of relationship is asked in two steps in the questionnaires for the years under study. If the partnership ends with a breakup, we choose the date when the individual actually stops living with the partner in the same household (decohabitation). The number of events is given in check 6.

2- Vector of dates concerning relationships that include a marriage.



3- Vector created for the combined partnership histories (combining relationships including a marriage with free standing cohabitations)



Remarks
To re-order this vector, we take the minimum (earliest) of the dates present in a slot and re-order by that minimum date
In general the date of separation is used to identify the end of a marriage. If year of separation is absent but divorce year is present, we use divorce year + month (see postcheck 11).

Verification of the combined partnership history

Postcheck 11

This check compares the number of partnerships declared and the number of valid partnership dates.
Number of partnerships declared (**npart**) is a CPC-derived variable generated by adding the original GHS variables that give the number of cohabitations and marriages (renamed **nmar** and **ncoh**). Two other variables have been created in the construction of the dataset, namely **npartu** and **nmaru**, which give the number of valid marriage and partnership dates, as described in the flag variables **parqn** and **marqn**.

The missing column in the table below refers to cases where either **nmar** or **ncoh** is missing. Note that the -8s in PH11nc increase from the year 2000 to reach 1,200 cases in 2007. Note that **npart** is equal to the minimum number of partnerships reported in that it reflects the sum of **ncoh** and **nmar** OR one of these where the other is missing. Missing values on **npart** will therefore reflect people missing on both variables.

Table PH11

year	checknbp					Total
	.	same number partn and valid dates	number partn>0 but no valid date	number partn<number valid dates	number partn>number valid dates	
2000	135	9910	54	2	132	10233
2001	304	10827	62	1	144	11338
2002	411	10292	71	15	145	10934
2003	518	11965	107	15	187	12792
2004	507	10286	123	11	130	11057
2005	582	12071	128	22	172	12975
2006	762	10458	88	38	363	11709
2007	838	9417	233	18	330	10836
Total	4057	85226	866	122	1603	91874

Table PH11nc: Number of cohabitations declared

year	ncoh										Total
	-8	0	1	2	3	4	5	6	7	8	
2000	155	7802	1780	369	82	23	12	6	4	0	10233
2001	322	8540	1913	418	103	22	9	2	9	0	11338
2002	431	7978	1958	441	96	18	6	5	1	0	10934
2003	541	9081	2344	620	148	32	15	5	6	0	12792
2004	508	7942	1997	458	109	27	9	3	2	2	11057
2005	571	9167	2443	572	159	40	13	4	5	1	12975
2006	1087	7796	2061	580	135	36	7	4	3	0	11709
2007	1203	7106	1885	484	115	31	4	5	2	1	10836
Total	4818	65412	16381	3942	947	229	75	34	32	4	91874

Table PH11nm: Number of marriages declared

year	nmar							Total
	-8	0	1	2	3	4	5	
2000	158	3286	5750	951	81	7	0	10233
2001	324	3633	6230	1042	98	10	1	11338
2002	428	3552	5830	1028	86	9	1	10934
2003	530	4174	6868	1105	109	6	0	12792
2004	530	3667	5740	1019	90	10	1	11057
2005	570	4233	6892	1159	114	7	0	12975
2006	817	3815	5873	1084	112	8	0	11709
2007	899	3552	5339	933	104	9	0	10836
Total	4256	29912	48522	8321	794	66	3	91874

Table PH11np: Number of partnerships declared

year	npart													Total
	-8	0	1	2	3	4	5	6	7	8	9	10	11	
2000	134	1970	5985	1688	345	75	22	5	6	2	1	0	0	10233
2001	304	2114	6608	1802	397	78	22	3	5	4	1	0	0	11338
2002	410	2008	6230	1781	410	69	14	5	5	1	0	0	1	10934
2003	516	2262	7222	2113	513	118	31	7	6	3	0	1	0	12792
2004	490	2143	6032	1854	419	84	22	7	2	1	3	0	0	11057
2005	546	2328	7213	2171	564	107	29	8	4	5	0	0	0	12975
2006	721	2176	6170	2029	478	98	25	7	3	1	1	0	0	11709
2007	802	2086	5656	1759	411	89	22	7	1	2	1	0	0	10836
Total	3923	17087	51116	15197	3537	718	187	49	32	19	7	1	1	91874

Table PH11nbp: Total number of partnerships with a valid beginning date

year	npartu								Total
	0	1	2	3	4	5	6	7	
2000	2157	6032	1630	342	62	10	0	0	10233
2001	2480	6659	1748	371	70	10	0	0	11338
2002	2488	6278	1726	373	57	10	1	1	10934
2003	2883	7239	2055	494	106	13	2	0	12792
2004	2739	6019	1800	414	72	11	2	0	11057
2005	2966	7246	2113	548	90	12	0	0	12975
2006	2937	6438	1797	444	81	11	1	0	11709
2007	3080	5761	1546	361	79	8	1	0	10836
Total	21730	51672	14415	3347	617	85	7	1	91874

Postcheck 12
Distribution of the type of relationship and of the type of end.

Checks are given for the 6 first relationships, since there are too few later ones to justify additional tables.

Note that the type of partnership, i.e. whether it was a free-standing cohabitation, a marriage preceded by premarital cohabitation or a direct marriage are derived from putting together information from the separate marriage histories and cohabitation histories. In other words, respondents are not asked directly, for each of their partnerships, about its type. For the end type however, we use the direct answer to the question on the type of end of the relationship (when it ended).

The missing column includes people who either did not have a first, second, etc. partnership, or for whom first, second etc. partnership information is missing.

Note in table PH12b, the category ‘No actual separation, just decohab.’ refers to people who continued in their relationship but stopped cohabiting, i.e. became ‘Living Apart Together’. There are so few of these, and the question was asked for so few years, that the coding is not retained in the final database.

Tables PH12a

year	cohmar1				Total
	.	marriage with premar cohab	direct marriage	cohabitation	
2000	2138	1973	4417	1705	10233
2001	2440	2189	4788	1921	11338
2002	2447	2122	4384	1981	10934
2003	2837	2726	4806	2423	12792
2004	2664	2357	4018	2018	11057
2005	2925	2920	4606	2524	12975
2006	2911	2629	3935	2234	11709
2007	3050	2406	3348	2032	10836
Total	21412	19322	34302	16838	91874

year	cohmar2				Total
	.	marriage with premar cohab	direct marriage	cohabitation	
2000	8159	921	320	833	10233
2001	9105	1012	358	863	11338
2002	8717	1031	320	866	10934
2003	10073	1159	345	1215	12792
2004	8720	1064	334	939	11057
2005	10172	1266	307	1230	12975
2006	9335	1099	251	1024	11709
2007	8805	944	227	860	10836
Total	73086	8496	2462	7830	91874

year	cohmar3				Total
	.	marriage with premar cohab	direct marriage	cohabitation	
2000	9816	163	45	209	10233
2001	10875	172	52	239	11338
2002	10482	196	40	216	10934
2003	12152	237	37	366	12792
2004	10546	191	34	286	11057
2005	12307	249	45	374	12975
2006	11161	208	33	307	11709
2007	10375	180	33	248	10836
Total	87714	1596	319	2245	91874

year	cohmar4				Total
	.	marriage with premar cohab	direct marriage	cohabitation	
2000	10159	34	3	37	10233
2001	11257	26	9	46	11338
2002	10862	25	12	35	10934
2003	12665	48	5	74	12792
2004	10966	37	7	47	11057
2005	12869	52	5	49	12975
2006	11611	47	4	47	11709
2007	10747	49	1	39	10836
Total	91136	318	46	374	91874

year	cohmar5				Total
	.	marriage with premar cohab	direct marriage	cohabitation	
2000	10223	5	0	5	10233
2001	11327	0	2	9	11338
2002	10922	3	2	7	10934
2003	12775	8	0	9	12792
2004	11044	8	0	5	11057
2005	12963	4	1	7	12975
2006	11697	2	1	9	11709
2007	10826	3	1	6	10836
Total	91777	33	7	57	91874

year	cohmar6				Total
	.	marriage with premar cohab	direct marriage	cohabitation	
2000	10233	0	0	0	10233
2001	11338	0	0	0	11338
2002	10932	0	1	1	10934
2003	12790	0	0	2	12792
2004	11055	1	0	1	11057
2005	12975	0	0	0	12975
2006	11708	1	0	0	11709
2007	10835	1	0	0	10836
Total	91866	3	1	4	91874

Tables PH12b

year	causend1					Total
	missing	no end	stopped living tgth	partner died	no actual separation, just decohab	
2000	2118	4959	2980	141	35	10233
2001	2422	5501	3206	175	34	11338
2002	2426	5128	3193	149	38	10934
2003	2811	5927	3841	160	53	12792
2004	2656	5015	3232	154	0	11057
2005	2952	5918	3946	159	0	12975
2006	3001	5313	3276	119	0	11709
2007	3104	4771	2863	98	0	10836
Total	21490	42532	26537	1155	160	91874

year	causend2					Total
	missing	no end	stopped living tgth	partner died	no actual separation, just decohab	
2000	8133	1322	738	25	15	10233
2001	9083	1446	758	35	16	11338
2002	8707	1389	793	35	10	10934
2003	10065	1619	1036	38	34	12792
2004	8715	1466	841	35	0	11057
2005	10170	1682	1092	31	0	12975
2006	9256	1533	893	27	0	11709
2007	8763	1293	751	29	0	10836
Total	72892	11750	6902	255	75	91874

year	causend3					Total
	missing	no end	stopped living tgth	partner died	no actual separation, just decohab	
2000	9811	230	176	6	10	10233
2001	10867	278	180	8	5	11338
2002	10472	283	170	5	4	10934
2003	12151	363	260	9	9	12792
2004	10545	307	200	5	0	11057
2005	12303	391	272	9	0	12975
2006	11157	315	232	5	0	11709
2007	10369	257	208	2	0	10836
Total	87675	2424	1698	49	28	91874

year	causend4					Total
	missing	no end	stopped living tgth	partner died	no actual separation, just decohab	
2000	10113	93	27	0	0	10233
2001	11209	97	31	0	1	11338
2002	10827	82	25	0	0	10934
2003	12609	136	46	0	1	12792
2004	10914	105	35	3	0	11057
2005	12784	153	38	0	0	12975
2006	11549	124	36	0	0	11709
2007	10702	104	29	1	0	10836
Total	90707	894	267	4	2	91874

year	causend5				Total
	missing	no end	stopped living tgth	no actual separation, just decohab	
2000	10205	27	1	0	10233
2001	11309	29	0	0	11338
2002	10907	24	3	0	10934
2003	12747	40	4	1	12792
2004	11025	28	4	0	11057
2005	12933	41	1	0	12975
2006	11678	30	1	0	11709
2007	10804	30	2	0	10836
Total	91608	249	16	1	91874

year	causend6			Total
	missing	no end	stopped living tgth	
2000	10228	4	1	10233
2001	11337	1	0	11338
2002	10928	5	1	10934
2003	12782	10	0	12792
2004	11052	5	0	11057
2005	12970	5	0	12975
2006	11705	4	0	11709
2007	10832	4	0	10836
Total	91834	38	2	91874

Postcheck 13

This check looks at whether there is a valid date of beginning/end of partnerships that we believe to exist (because at least one date is declared within the relationship) or an end is declared (in the variable that gives the end cause).

Respondents who don't declare a partnership and those with a missing code on whether the partnership has started/ended are in the 'missing' column.

Tables PH13a

year	chkbeg1			Total
	missing	beg. supposed and valid date beg.	no valid date beg. while supposed beg.	
2000	2138	8060	35	10233
2001	2440	8836	62	11338
2002	2447	8417	70	10934
2003	2837	9865	90	12792
2004	2664	8293	100	11057
2005	2925	9987	63	12975
2006	2911	8749	49	11709
2007	3050	7736	50	10836
Total	21412	69943	519	91874

year	chkbeg.2			Total
	missing	beg. supposed and valid date beg.	no valid date beg. while supposed beg.	
2000	8159	2056	18	10233
2001	9105	2211	22	11338
2002	8717	2187	30	10934
2003	10073	2689	30	12792
2004	8720	2312	25	11057
2005	10172	2770	33	12975
2006	9335	2348	26	11709
2007	8805	2004	27	10836
Total	73086	18577	211	91874

year	chkbeg.3			Total
	missing	beg. supposed and valid date beg.	no valid date beg. while supposed beg.	
2000	9816	417	0	10233
2001	10875	460	3	11338
2002	10482	449	3	10934
2003	12152	632	8	12792
2004	10546	506	5	11057
2005	12307	661	7	12975
2006	11161	542	6	11709
2007	10375	459	2	10836
Total	87714	4126	34	91874

year	chkbeg.4			Total
	missing	beg. supposed and valid date beg.	no valid date beg. while supposed beg.	
2000	10159	73	1	10233
2001	11257	80	1	11338
2002	10862	72	0	10934
2003	12665	127	0	12792
2004	10966	90	1	11057
2005	12869	106	0	12975
2006	11611	97	1	11709
2007	10747	89	0	10836
Total	91136	734	4	91874

year	chkbeg.5			Total
	missing	beg. supposed and valid date beg.	no valid date beg. while supposed beg.	
2000	10223	10	0	10233
2001	11327	11	0	11338
2002	10922	12	0	10934
2003	12775	17	0	12792
2004	11044	13	0	11057
2005	12963	12	0	12975
2006	11697	12	0	11709
2007	10826	9	1	10836
Total	91777	96	1	91874

Tables PH13b

year	chkend1				Total
	missing	end declared and valid date end	no valid date end while declared end	no end declared but valid end date	
2000	7078	2990	131	34	10233
2001	7923	3231	150	34	11338
2002	7556	3230	112	36	10934
2003	8740	3885	114	53	12792
2004	7671	3316	70	0	11057
2005	8866	4011	94	4	12975
2006	8312	3317	78	2	11709
2007	7872	2877	84	3	10836
Total	64018	26857	833	166	91874

year	chkend2				Total
	missing	end declared and valid date end	no valid date end while declared end	no end declared but valid end date	
2000	9455	698	65	15	10233
2001	10529	742	51	16	11338
2002	10096	788	40	10	10934
2003	11687	1025	49	31	12792
2004	10179	856	20	2	11057
2005	11850	1092	30	3	12975
2006	10785	882	38	4	11709
2007	10055	751	29	1	10836
Total	84636	6834	322	82	91874

year	chkend3				Total
	missing	end declared and valid date end	no valid date end while declared end	no end declared but valid end date	
2000	10041	169	13	10	10233
2001	11146	176	12	4	11338
2002	10755	162	13	4	10934
2003	12512	256	13	11	12792
2004	10851	199	6	1	11057
2005	12692	269	12	2	12975
2006	11472	221	16	0	11709
2007	10625	192	18	1	10836
Total	90094	1644	103	33	91874

year	chkend4				Total
	missing	end declared and valid date end	no valid date end while declared end	no end declared but valid end date	
2000	10206	24	3	0	10233
2001	11306	27	4	1	11338
2002	10909	24	1	0	10934
2003	12745	43	3	1	12792
2004	11019	34	4	0	11057
2005	12937	37	1	0	12975
2006	11673	34	2	0	11709
2007	10806	27	3	0	10836
Total	91601	250	21	2	91874

year	chkend5			Total
	missing	end declared and valid date end	no end declared but valid end date	
2000	10232	1	0	10233
2001	11338	0	0	11338
2002	10931	3	0	10934
2003	12787	4	1	12792
2004	11053	4	0	11057
2005	12974	1	0	12975
2006	11708	1	0	11709
2007	10834	2	0	10836
Total	91857	16	1	91874

Postcheck 14

Check the order of the component dates of the relationships within the respondent's entire relationship history (including marriages and free-standing cohabitations).

- Table PH14 gives the data before re-ordering in which **ordstrtv, ordmav** and **ordendlvv** represent the start dates, marriage dates and end dates prior to re-ordering.
- Table PH14c gives the corresponding data following reordering, with **ordstrt, ordma and ordendlv** representing start date, marriage dates and end dates after re-ordering.

Note that the numbers "out of sequence" or "missing in the middle" vary according to whether the consistency of the start, or end, or marriage date is tabulated.

Because we have ordered these according to the minimum date within each history, those with a missing date can be identified as in the middle by *some* date within that partnership slot.

The number of "missings in the middle" in **ordma** of table PH14c is large, because some of these partnerships are not marriages, and so the marriage dates are absent.

Note that end dates for marriages are based on date of separation rather than date of divorce, as divorce dates appear to have more errors associated with them.

The "missing" column includes the persons who have not declared any date of start of relationship, marriage, etc.

Tables PH14

year	ordstrtv				Total
	missing	ordered no problem	missings in the middle	none missing but out of sequence	
2000	2157	7605	21	450	10233
2001	2480	8382	24	452	11338
2002	2488	7911	36	499	10934
2003	2883	9250	53	606	12792
2004	2739	7745	32	541	11057
2005	2966	9305	30	674	12975
2006	2937	8158	30	584	11709
2007	3080	7217	19	520	10836
Total	21730	65573	245	4326	91874

year	ordmav				Total
	missing	ordered no problem	missings in the middle	none missing but out of sequence	
2000	3470	6758	4	1	10233
2001	3989	7338	11	0	11338
2002	4018	6910	6	0	10934
2003	4772	8008	12	0	12792
2004	4393	6654	10	0	11057
2005	4915	8052	8	0	12975
2006	4688	7013	6	2	11709
2007	4682	6150	3	1	10836
Total	34927	56883	60	4	91874

year	ordendlvv				Total
	missing	ordered no problem	missings in the middle	none missing but out of sequence	
2000	7191	2580	380	82	10233
2001	8059	2806	385	88	11338
2002	7660	2764	421	89	10934
2003	8841	3325	509	117	12792
2004	7723	2771	463	100	11057
2005	8945	3349	575	106	12975
2006	8365	2738	492	114	11709
2007	7931	2368	436	101	10836
Total	64715	22701	3661	797	91874

Tables PH14c

year	ordstrt				Total
	missing	ordered no problem	missings in the middle	none missing but out of sequence	
2000	2157	8056	19	1	10233
2001	2480	8829	29	0	11338
2002	2488	8408	37	1	10934
2003	2883	9850	59	0	12792
2004	2739	8286	31	1	11057
2005	2966	9977	31	1	12975
2006	2937	8745	26	1	11709
2007	3080	7738	17	1	10836
Total	21730	69889	249	6	91874

year	ordma				Total
	missing	ordered no problem	missings in the middle	none missing but out of sequence	
2000	3470	6305	457	1	10233
2001	3989	6876	471	2	11338
2002	4018	6406	508	2	10934
2003	4772	7404	614	2	12792
2004	4393	6139	523	2	11057
2005	4915	7387	670	3	12975
2006	4688	6447	572	2	11709
2007	4682	5649	503	2	10836
Total	34927	52613	4318	16	91874

year	ordendlv				Total
	missing	ordered no problem	missings in the middle	none missing but out of sequence	
2000	7191	3010	22	10	10233
2001	8059	3244	18	17	11338
2002	7660	3235	13	26	10934
2003	8841	3889	33	29	12792
2004	7723	3280	21	33	11057
2005	8945	3969	29	32	12975
2006	8365	3289	36	19	11709
2007	7931	2857	31	17	10836
Total	64715	26773	203	183	91874

Postcheck 15

We look here at three relationships only as the tables are confined to cohabitations. The corresponding tables for the relationships including a marriage are in the MH series of tables, below.
This check looks to see whether there is a valid year of start of cohabitation and where this is not the case, whether some end date for this relationship is present.
"Missing" is used when there is neither a start date, nor an end date.

Tables PH15

year	varcoh1				Total
	missing	year cohabitation present	no yr cohab but yr end rel	no yr cohab but yr stop liv tgth	
2000	8947	1278	4	4	10233
2001	9998	1330	8	2	11338
2002	9478	1453	3	0	10934
2003	10902	1848	13	29	12792
2004	9518	1531	2	6	11057
2005	11076	1897	1	1	12975
2006	10156	1549	2	2	11709
2007	9495	1336	4	1	10836
Total	79570	12222	37	45	91874

year	varcoh2				Total
	missing	year cohabitation present	no yr cohab but yr end rel	no yr cohab but yr stop liv tgth	
2000	9937	296	0	0	10233
2001	11018	320	0	0	11338
2002	10592	342	0	0	10934
2003	12295	493	0	4	12792
2004	10703	352	1	1	11057
2005	12478	496	0	1	12975
2006	11300	406	2	1	11709
2007	10504	330	2	0	10836
Total	88827	3035	5	7	91874

year	varcoh3				Total
	missing	year cohabitation present	no yr cohab but yr end rel	no yr cohab but yr stop liv tgth	
2000	10151	82	0	0	10233
2001	11255	80	3	0	11338
2002	10865	69	0	0	10934
2003	12674	118	0	0	12792
2004	10968	88	1	0	11057
2005	12836	137	0	2	12975
2006	11604	102	0	3	11709
2007	10750	86	0	0	10836
Total	91103	762	4	5	91874

Postcheck 16

Check the order of the beginning and end (if exists) within each relationship after the corrections, and if they exist, of premarital cohabitation and marriage, and of separation and divorce.

Note that the column giving the end date greater than divorce date is provided here for information but these are not flagged as errors, because we do not place much reliance on the divorce dates.

The column ‘missing’ includes those who did not report any start date for this relationship order.

The cases identified in the ‘beg date>=end date’ column are not edited further since it seems likely that one or other date was mistaken.

We note that relatively high proportions of partnerships of order 3 and 4 have an end date earlier than the beginning date.

Tables PH16

year	chkin1					Total
	missing	no problem	beg date>=end date	pre mar coh date>mar date	end date>div date	
2000	2171	8015	20	1	26	10233
2001	2502	8791	21	0	24	11338
2002	2517	8369	27	0	21	10934
2003	2926	9785	30	0	51	12792
2004	2763	8235	24	0	35	11057
2005	2988	9926	29	0	32	12975
2006	2960	8691	28	0	30	11709
2007	3100	7692	21	0	23	10836
Total	21927	69504	200	1	242	91874

year	chkin2					Total
	missing	no problem	beg date>=end date	pre mar coh date>mar date	end date>div date	
2000	8177	2044	7	1	4	10233
2001	9127	2199	10	0	2	11338
2002	8747	2173	11	0	3	10934
2003	10103	2661	21	0	7	12792
2004	8745	2299	12	0	1	11057
2005	10205	2754	11	0	5	12975
2006	9361	2328	11	1	8	11709
2007	8832	1988	11	1	4	10836
Total	73297	18446	94	3	34	91874

year	chkin3					Total
	missing	no problem	beg date>=end date	pre mar coh date>mar date	end date>div date	
2000	9816	410	6	1	0	10233
2001	10878	457	1	0	2	11338
2002	10485	446	3	0	0	10934
2003	12160	622	10	0	0	12792
2004	10551	501	4	0	1	11057
2005	12314	655	6	0	0	12975
2006	11167	539	3	0	0	11709
2007	10377	452	7	0	0	10836
Total	87748	4082	40	1	3	91874

year	chkin4			Total
	missing	no problem	beg date>=end date	
2000	10160	73	0	10233
2001	11258	80	0	11338
2002	10862	72	0	10934
2003	12665	127	0	12792
2004	10967	88	2	11057
2005	12869	106	0	12975
2006	11612	97	0	11709
2007	10747	89	0	10836
Total	91140	732	2	91874

Tables PH16d: Proportion with beg date>=end date, by partnership order (in %)

	Partnership order			
year	1	2	3	4
2000	0.67	0.99	3.35	0.00
2001	0.65	1.33	0.56	0.00
2002	0.83	1.39	1.82	0.00
2003	0.77	2.02	3.80	0.00
2004	0.73	1.42	2.04	6.06
2005	0.73	1.02	2.25	0.00
2006	0.91	1.25	1.38	0.00
2007	0.73	1.48	3.63	0.00

Tables PH16d: Proportion with pre mar coh date>mar date, by partnership order (in %)

	Partnership order			
year	1	2	3	4
2000	0.02	0.08	0.48	0.00
2001	0.00	0.00	0.00	0.00
2002	0.00	0.00	0.00	0.00
2003	0.00	0.00	0.00	0.00
2004	0.00	0.00	0.00	0.00
2005	0.00	0.00	0.00	0.00
2006	0.00	0.08	0.00	0.00
2007	0.00	0.09	0.00	0.00

Tables PH16d: Proportion with end date>div date, by partnership order (in %)

	Partnership order			
year	1	2	3	4
2000	1.68	1.83	0.00	0.00
2001	1.44	0.87	4.76	0.00
2002	1.35	1.30	0.00	0.00
2003	2.68	2.56	0.00	0.00
2004	2.13	0.41	4.00	0.00
2005	1.64	1.78	0.00	0.00
2006	1.83	3.38	0.00	0.00
2007	1.58	2.01	0.00	0.00

Postcheck 17

Note that to appear in anything other than the missing column of table PH17, a person must have had at least two relationships.
In these cases, we ask whether we have the elements of the i^{th} relationship and did they all happen before the beginning of the $i+1^{th}$?
Note that the column on divorce date is provided here for information but these are not flagged as errors, because we do not place much reliance on the divorce dates.
At present, we do not see that there is a straightforward defensible edit that can be applied to the cases that have no end date for i but a start date for i+1.
In relation to end date i greater than start date i+1, these would be considered in the overlap checks PH18 etc.

Tables PH17

year	chkout1					Total
	missing	start date for i+1, no problem	start date for i+1 but not for i	start date for i+1 but no end date for i	end date i>start date i+1	
2000	8177	1920	14	42	80	10233
2001	9127	2071	17	41	82	11338
2002	8747	2037	24	37	89	10934
2003	10103	2501	41	40	107	12792
2004	8745	2159	20	34	99	11057
2005	10205	2583	16	37	134	12975
2006	9361	2157	19	59	113	11709
2007	8832	1831	9	58	106	10836
Total	73297	17259	160	348	810	91874

year	chkout2					Total
	missing	start date for i+1, no problem	start date for i+1 but not for i	start date for i+1 but no end date for i	end date i>start date i+1	
2000	9816	382	2	8	25	10233
2001	10878	418	6	5	31	11338
2002	10485	401	5	10	33	10934
2003	12160	557	11	23	41	12792
2004	10551	452	6	10	38	11057
2005	12314	585	6	16	54	12975
2006	11167	477	1	22	42	11709
2007	10377	411	3	19	26	10836
Total	87748	3683	40	113	290	91874

year	chkout3					Total
	missing	start date for i+1, no problem	start date for i+1 but not for i	start date for i+1 but no end date for i	end date i>start date i+1	
2000	10160	66	0	1	6	10233
2001	11258	74	0	2	4	11338
2002	10862	65	1	1	5	10934
2003	12665	106	3	8	10	12792
2004	10967	76	2	4	8	11057
2005	12869	93	3	0	10	12975
2006	11612	82	3	5	7	11709
2007	10747	77	0	3	9	10836
Total	91140	639	12	24	59	91874

year	chkout4					Total
	missing	start date for i+1, no problem	start date for i+1 but not for i	start date for i+1 but no end date for i	end date i>start date i+1	
2000	10223	10	0	0	0	10233
2001	11327	9	1	1	0	11338
2002	10922	12	0	0	0	10934
2003	12775	14	0	0	3	12792
2004	11044	13	0	0	0	11057
2005	12963	10	0	1	1	12975
2006	11697	11	0	0	1	11709
2007	10827	8	0	0	1	10836
Total	91778	87	1	2	6	91874

Postcheck 18

These tables show the number of partnerships that are overlapping each other before/after correction. The corrections are small edits based on the visual inspection of cases with errors. PH18 gives the figures prior to these edits whilst PH18c gives figures following those minor edits.

Tables PH18 before corrections

Overlapping union 1 and 2			
year	no	yes	% yes/no
2000	1934	88	4.4%
2001	2088	84	3.9%
2002	2061	91	4.2%
2003	2542	111	4.2%
2004	2179	103	4.5%
2005	2599	145	5.3%
2006	2177	117	5.1%
2007	1841	117	6.0%
Total	17421	856	4.7%

Overlapping union 2 and 3			
year	no	yes	% yes/no
2000	384	25	6.1%
2001	424	31	6.8%
2002	406	34	7.7%
2003	568	41	6.7%
2004	458	38	7.7%
2005	593	54	8.3%
2006	478	43	8.3%
2007	414	26	5.9%
Total	3725	292	7.3%

Overlapping union 3 and 4			
year	no	yes	% yes/no
2000	66	6	8.3%
2001	74	4	5.1%
2002	66	5	7.0%
2003	109	10	8.4%
2004	78	8	9.3%
2005	96	10	9.4%
2006	85	7	7.6%
2007	77	9	10.5%
Total	651	59	8.3%

Overlapping union 4 and 5			
year	no	yes	% yes/no
2000	10	0	0.0%
2001	10	0	0.0%
2002	12	0	0.0%
2003	14	3	17.6%
2004	13	0	0.0%
2005	10	1	9.1%
2006	11	1	8.3%
2007	8	2	20.0%
Total	88	7	7.4%

Tables PH18c after correction

Overlapping partnership 1 and 2			
year	no	yes	% yes
2000	1934	80	4.0%
2001	2088	82	3.8%
2002	2061	89	4.1%
2003	2542	107	4.0%
2004	2179	99	4.3%
2005	2599	134	4.9%
2006	2176	113	4.9%
2007	1840	106	5.4%
Total	17419	810	4.4%

Overlapping partnership 2 and 3			
year	no	yes	% yes
2000	384	25	6.1%
2001	424	31	6.8%
2002	406	33	7.5%
2003	568	41	6.7%
2004	458	38	7.7%
2005	591	54	8.4%
2006	478	42	8.1%
2007	414	26	5.9%
Total	3723	290	7.2%

Overlapping partnership 3 and 4			
year	no	yes	% yes
2000	66	6	8.3%
2001	74	4	5.1%
2002	66	5	7.0%
2003	109	10	8.4%
2004	78	8	9.3%
2005	96	10	9.4%
2006	85	7	7.6%
2007	77	9	10.5%
Total	651	59	8.3%

Overlapping partnership 4 and 5			
year	no	yes	% yes
2000	10	0	0.0%
2001	10	0	0.0%
2002	12	0	0.0%
2003	14	3	17.6%
2004	13	0	0.0%
2005	10	1	9.1%
2006	11	1	8.3%
2007	8	1	11.1%
Total	88	6	6.4%

Postcheck 19

These tables show whether the person has declared two partnerships with the same beginning or/and end dates.
The following tables include only individuals who have at least two partnerships.

Before the construction of these tables, the following edit was made for respondents who have exactly the same beginning and end date for their first relationship and another one:

- where the relationships involved are two cohabitations: we delete one of these relationships.
- where one of the relationships is a marriage no action is taken.

As a matter of fact, the “other” relationship is always of order 2.

The five cases remaining with a same beginning and end dates involve a cohabitation with a marriage.

Tables PH19 before correction

Same beginning and/or end in partnership 1 compared with others				
year	different beg and ends	same beg	same end	same beg and end
2000	8069	10	4	0
2001	8863	2	0	0
2002	8439	5	4	0
2003	9909	11	7	0
2004	8317	7	2	1
2005	9994	15	3	0
2006	8765	9	3	1
2007	7737	12	5	5
Total	70093	71	28	7

Same beginning and/or end in partnership 2 compared with others			
year	different beg and ends	same beg	same end
2000	2068	2	0
2001	2219	1	1
2002	2193	2	2
2003	2699	10	2
2004	2325	2	2
2005	2781	9	3
2006	2354	6	1
2007	2014	5	0
Total	18653	37	11

Same beginning and/or end in partnership 3 compared with others			
year	different beg and ends	same beg	same end
2000	417	0	0
2001	461	0	1
2002	450	1	0
2003	637	2	1
2004	509	1	0
2005	665	2	0
2006	544	2	0
2007	458	1	0
Total	4141	9	2

Tables PH19c after correction

Same beginning and/or end in partnership 1 compared with others			
year	different beg and ends	same beg	same end
2000	8077	2	4
2001	8865	0	0
2002	8441	3	4
2003	9913	7	7
2004	8321	3	2
2005	10005	4	3
2006	8770	4	3
2007	7744	5	5
Total	70136	28	28

Same beginning and/or end in partnership 2 compared with others			
year	different beg and ends	same beg	same end
2000	2060	2	0
2001	2217	1	1
2002	2192	1	2
2003	2699	6	2
2004	2321	2	2
2005	2772	7	3
2006	2350	5	1
2007	2007	5	0
Total	18618	29	11

Same beginning and/or end in partnership 3 compared with others			
year	different beg and ends	same beg	same end
2000	417	0	0
2001	461	0	1
2002	449	1	0
2003	633	2	1
2004	509	1	0
2005	663	2	0
2006	543	2	0
2007	458	1	0
Total	4133	9	2

Classification of errors 21

This check examines the frequencies of erroneous partnership histories. **partq1-partqn** given the checks internal to each partnership spell of each individual. **partqual** corresponds to the overall quality of the partnership history: only one such variable is derived for each individual.

The check only looks at beginning and end of partnerships, and does not check any intermediate or following event like marriage and divorce.

code	meaning	date in spell n	beg date	end date	order ok
0	no date present in the spell	no			
1	no problem	yes	yes	(yes)	yes
2	OK after editing	yes	yes edited	yes edited	yes edited
3	partially OK	yes	yes	no	yes
4	error	yes	no		OR no
-9	DNA: not eligible FI, proxy				

Explanation:

Code 1 "no problem": when the partnership has at least a beginning date, an end date (if this applies) and that the end date>beginning date.

Code 2: where code 1 applies but that at least one of the dates present has been edited

Code 3: where a beginning date is present but an end date which should be present is missing

Code 4: missing beginning date or beginning and end date out of sequence

Tables PH21

year	partq1					Total
	no date present in the spell	no problem	ok after editing	partially ok	error	
2000	2138	7947	10	104	34	10233
2001	2440	8715	5	120	58	11338
2002	2447	8340	8	69	70	10934
2003	2837	9771	14	85	85	12792
2004	2664	8257	6	32	98	11057
2005	2925	9907	12	66	65	12975
2006	2911	8684	10	50	54	11709
2007	3050	7684	10	44	48	10836
Total	21412	69305	75	570	512	91874

year	partq2					Total
	no date present in the spell	no problem	ok after editing	partially ok	error	
2000	8159	2028	9	29	8	10233
2001	9105	2182	9	23	19	11338
2002	8717	2181	1	8	27	10934
2003	10073	2664	12	18	25	12792
2004	8720	2302	7	7	21	11057
2005	10172	2777	2	7	17	12975
2006	9335	2344	5	9	16	11709
2007	8805	2000	7	4	20	10836
Total	73086	18478	52	105	153	91874

year	partq3					Total
	no date present in the spell	no problem	ok after editing	partially ok	error	
2000	9816	411	1	4	1	10233
2001	10875	454	2	4	3	11338
2002	10482	450	0	0	2	10934
2003	12152	631	5	2	2	12792
2004	10546	504	2	1	4	11057
2005	12307	663	2	2	1	12975
2006	11161	544	0	2	2	11709
2007	10375	457	1	2	1	10836
Total	87714	4114	13	17	16	91874

year	partq4					Total
	no date present in the spell	no problem	ok after editing	partially ok	error	
2000	10159	74	0	0	0	10233
2001	11257	81	0	0	0	11338
2002	10862	72	0	0	0	10934
2003	12665	123	3	1	0	12792
2004	10966	89	0	1	1	11057
2005	12869	104	2	0	0	12975
2006	11611	96	2	0	0	11709
2007	10747	88	1	0	0	10836
Total	91136	727	8	2	1	91874

year	partq5		Total
	no date present in the spell	no problem	
2000	10223	10	10233
2001	11327	11	11338
2002	10922	12	10934
2003	12775	17	12792
2004	11044	13	11057
2005	12963	12	12975
2006	11697	12	11709
2007	10826	10	10836
Total	91777	97	91874

year	partq6		Total
	no date present in the spell	no problem	
2000	10233	0	10233
2001	11338	0	11338
2002	10932	2	10934
2003	12790	2	12792
2004	11055	2	11057
2005	12975	0	12975
2006	11708	1	11709
2007	10835	1	10836
Total	91866	8	91874

Construction of **partqual**

0	No partnership
1	All OK
2	OK after editing
3	Partially OK
4	Unusable (decomposed 41 42 43)
-9	DNA: not eligible FI, proxy

partqual=0 for people who declared they didn't experience any partnership.
If they declare at least one partnership but don't have the right number of valid dates of beginning of partnership (after editing), then **partqual**=41.
If their number of start dates is the same as the number of partnerships declared, then we do additional checks:
If there is no problem at all, then **partqual**=1.
If at least two partnerships are overlapping, then **partqual**=42.
If there are neither overlaps nor other problems but some dates have been edited, then **partqual**=2.
If all the start dates are present but some end dates are missing, the series is considered as partially valid (**partqual**=3).
If there is another problem that makes the series unusable, for instance that the start and end dates are out of sequence, then **partqual**=43
-8 (missing) is coded when the no. of partnerships is missing

These coding conventions are summarized in the table below:

		number of dates	partn ok	partn edited	partn partially ok	partn unusable	overlap
0	No partnership declared	0					
1	no problem	>=1	yes	no	no	no	no
2	OK after editing	>=1		yes	no	no	OR yes edited
3	Partially OK	>=1			yes	no	no
4	error: inconsistent number	incons with decl					
5	error: other	>=1				yes	OR yes
-9	DNA: not elig FI, proxy						
-8	Missing						

Tables PH21q

year	partqual							Total
	missing	no partn declared	no problem	ok after editing	partially ok	error: inconsistent number	error: other	
2000	134	1970	7702	14	124	134	155	10233
2001	304	2114	8462	4	136	145	173	11338
2002	410	2008	8099	5	64	160	188	10934
2003	516	2262	9453	15	91	202	253	12792
2004	490	2143	7993	5	31	141	254	11057
2005	546	2328	9538	3	65	194	301	12975
2006	721	2176	8118	4	53	401	236	11709
2007	802	2086	7194	7	42	348	357	10836
Total	3923	17087	66559	57	606	1725	1917	91874

Verification of the marriage history

Postcheck marriage 11

This check compares the number of marriages declared and the number of valid marriage dates.

Tables MH11

year	checknbm					Total
	missing	same number marr and valid dates	number marr>0 but no valid date	number marr<number valid dates	number marr>number valid dates	
2000	158	10038	25	0	12	10233
2001	324	10965	30	0	19	11338
2002	428	10455	38	0	13	10934
2003	530	12172	66	0	24	12792
2004	530	10309	194	0	24	11057
2005	570	12271	112	4	18	12975
2006	821	10695	70	2	121	11709
2007	899	9601	244	2	90	10836
Total	4260	86506	779	8	321	91874

Postcheck marriage 12

Distribution of the type of marriage and of the type of end. The missing column includes people with no marriages or an undeclared number of marriages.

Tables MH12a

year	typmar1			Total
	missing	marriage with premar cohab	direct marriage	
2000	3456	2289	4488	10233
2001	3966	2507	4865	11338
2002	3993	2477	4464	10934
2003	4738	3156	4898	12792
2004	4217	2730	4110	11057
2005	4881	3408	4686	12975
2006	4662	3055	3992	11709
2007	4659	2786	3391	10836
Total	34572	22408	34894	91874

year	typmar2			Total
	missing	marriage with premar cohab	direct marriage	
2000	9214	737	282	10233
2001	10206	811	321	11338
2002	9836	822	276	10934
2003	11586	923	283	12792
2004	9949	845	263	11057
2005	11727	992	256	12975
2006	10638	854	217	11709
2007	9912	717	207	10836
Total	83068	6701	2105	91874

year	typmar3			Total
	missing	marriage with premar cohab	direct marriage	
2000	10149	66	18	10233
2001	11237	76	25	11338
2002	10843	72	19	10934
2003	12679	96	17	12792
2004	10961	77	19	11057
2005	12864	86	25	12975
2006	11620	72	17	11709
2007	10748	74	14	10836
Total	91101	619	154	91874

year	typmar4			Total
	missing	marriage with premar cohab	direct marriage	
2000	10227	6	0	10233
2001	11328	6	4	11338
2002	10926	7	1	10934
2003	12786	5	1	12792
2004	11047	7	3	11057
2005	12970	5	0	12975
2006	11704	5	0	11709
2007	10829	7	0	10836
Total	91817	48	9	91874

year	typmar5			Total
	missing	marriage with premar cohab	direct marriage	
2000	10233	0	0	10233
2001	11337	0	1	11338
2002	10934	0	0	10934
2003	12792	0	0	12792
2004	11056	1	0	11057
2005	12975	0	0	12975
2006	11709	0	0	11709
2007	10836	0	0	10836
Total	91872	1	1	91874

Tables MH12b

year	causendm1					Total
	missing	current marriage	spouse died	divorce	stopped living tgth	
2000	3449	4635	136	1735	278	10233
2001	3958	5043	171	1880	286	11338
2002	3984	4728	145	1801	276	10934
2003	4728	5451	151	2106	356	12792
2004	4206	4642	147	1838	224	11057
2005	4878	5452	156	2176	313	12975
2006	4727	4801	112	1828	241	11709
2007	4692	4235	90	1601	218	10836
Total	34622	38987	1108	14965	2192	91874

year	causendm2					Total
	missing	current marriage	spouse died	divorce	stopped living tgth	
2000	9210	740	19	206	58	10233
2001	10204	817	25	231	61	11338
2002	9834	820	23	210	47	10934
2003	11585	868	24	258	57	12792
2004	9950	829	27	203	48	11057
2005	11727	910	22	260	56	12975
2006	10554	867	21	218	49	11709
2007	9871	711	20	189	45	10836
Total	82935	6562	181	1775	421	91874

year	causendm3					Total
	missing	current marriage	spouse died	divorce	stopped living tgth	
2000	10149	54	1	19	10	10233
2001	11237	71	2	22	6	11338
2002	10843	64	1	21	5	10934
2003	12679	78	5	23	7	12792
2004	10961	71	4	15	6	11057
2005	12864	85	3	17	6	12975
2006	11620	65	1	17	6	11709
2007	10749	60	2	17	8	10836
Total	91102	548	19	151	54	91874

year	causendm4					Total
	missing	current marriage	spouse died	divorce	stopped living tgth	
2000	10227	4	0	1	1	10233
2001	11328	7	0	3	0	11338
2002	10926	8	0	0	0	10934
2003	12786	5	0	0	1	12792
2004	11047	5	1	2	2	11057
2005	12970	4	0	1	0	12975
2006	11704	3	0	2	0	11709
2007	10829	6	0	1	0	10836
Total	91817	42	1	10	4	91874

year	causendm5			Total
	missing	current marriage	divorce	
2000	10233	0	0	10233
2001	11337	1	0	11338
2002	10934	0	0	10934
2003	12792	0	0	12792
2004	11056	0	1	11057
2005	12975	0	0	12975
2006	11709	0	0	11709
2007	10836	0	0	10836
Total	91872	1	1	91874

Postcheck marriage 13

When a marriage is believed to exist (because at least one date is declared within the marriage slot) or when an end is declared (in the variable that gives the end cause), is there a valid date?

The date checked here is the date of marriage rather than of any premarital cohabitation.
A respondent with a missing code on whether the marriage had ended is in the 'missing' column.

Tables MH13a

year	chkbeg1			Total
	missing	mar supposed and valid date mar	no valid date mar while supposed mar	
2000	3456	6760	17	10233
2001	3966	7340	32	11338
2002	3993	6911	30	10934
2003	4738	8012	42	12792
2004	4217	6656	184	11057
2005	4881	8055	39	12975
2006	4662	7016	31	11709
2007	4659	6151	26	10836
Total	34572	56901	401	91874

year	chkbeg2			Total
	missing	mar supposed and valid date mar	no valid date mar while supposed mar	
2000	9214	1017	2	10233
2001	10206	1126	6	11338
2002	9836	1094	4	10934
2003	11586	1191	15	12792
2004	9949	1097	11	11057
2005	11727	1237	11	12975
2006	10638	1065	6	11709
2007	9912	916	8	10836
Total	83068	8743	63	91874

year	chkbeg3			Total
	missing	mar supposed and valid date mar	no valid date mar while supposed mar	
2000	10149	84	0	10233
2001	11237	101	0	11338
2002	10843	91	0	10934
2003	12679	113	0	12792
2004	10961	95	1	11057
2005	12864	111	0	12975
2006	11620	89	0	11709
2007	10748	88	0	10836
Total	91101	772	1	91874

year	chkbeg4			Total
	missing	mar supposed and valid date mar	no valid date mar while supposed mar	
2000	10227	6	0	10233
2001	11328	10	0	11338
2002	10926	8	0	10934
2003	12786	6	0	12792
2004	11047	9	1	11057
2005	12970	5	0	12975
2006	11704	5	0	11709
2007	10829	7	0	10836
Total	91817	56	1	91874

Tables MH13b

year	chkend1			Total
	missing	end declared and valid date end	no valid date end while declared end	
2000	8084	2024	125	10233
2001	9001	2185	152	11338
2002	8712	2112	110	10934
2003	10179	2492	121	12792
2004	8848	2157	52	11057
2005	10330	2548	97	12975
2006	9528	2113	68	11709
2007	8927	1844	65	10836
Total	73609	17475	790	91874

year	chkend2			Total
	missing	end declared and valid date end	no valid date end while declared end	
2000	9950	244	39	10233
2001	11021	287	30	11338
2002	10654	267	13	10934
2003	12453	321	18	12792
2004	10779	270	8	11057
2005	12637	326	12	12975
2006	11421	273	15	11709
2007	10582	241	13	10836
Total	89497	2229	148	91874

year	chkend3			Total
	missing	end declared and valid date end	no valid date end while declared end	
2000	10203	26	4	10233
2001	11308	28	2	11338
2002	10907	27	0	10934
2003	12757	32	3	12792
2004	11032	25	0	11057
2005	12949	24	2	12975
2006	11685	22	2	11709
2007	10809	25	2	10836
Total	91650	209	15	91874

year	chkend4			Total
	missing	end declared and valid date end	no valid date end while declared end	
2000	10231	2	0	10233
2001	11335	3	0	11338
2002	10934	0	0	10934
2003	12791	1	0	12792
2004	11052	4	1	11057
2005	12974	1	0	12975
2006	11707	2	0	11709
2007	10835	1	0	10836
Total	91859	14	1	91874

Postcheck marriage 14

Check the order of the component dates of the relationships

- Table PH14 gives the data before re-ordering in which **ordstrtv, ordmav** and **ordendlvv** represent the start dates, marriage dates and end dates prior to re-ordering.
- Table PH14c gives the corresponding data following reordering, with **ordstrt, ordma** and **ordendlv** representing start date, marriage dates, end dates after re-ordering.

The "missing" column includes people who have not declared any date in the concerned component date.

Tables MH14

year	ordstrtv				Total
	missing	ordered no problem	missings in the middle	none missing but out of sequence	
2000	3482	6738	11	2	10233
2001	4013	7306	16	3	11338
2002	4043	6865	26	0	10934
2003	4784	7991	17	0	12792
2004	4301	6734	20	2	11057
2005	4936	8019	19	1	12975
2006	4699	6993	16	1	11709
2007	4696	6126	11	3	10836
Total	34954	56772	136	12	91874

year	ordmav				Total
	missing	ordered no problem	missings in the middle	none missing but out of sequence	
2000	3468	6758	6	1	10233
2001	3987	7338	13	0	11338
2002	4018	6910	6	0	10934
2003	4770	8008	14	0	12792
2004	4391	6654	12	0	11057
2005	4915	8052	8	0	12975
2006	4688	7013	6	2	11709
2007	4682	6149	4	1	10836
Total	34919	56882	69	4	91874

year	ordendlvv				Total
	missing	ordered no problem	missings in the middle	none missing but out of sequence	
2000	8199	2026	8	0	10233
2001	9145	2185	7	1	11338
2002	8819	2112	3	0	10934
2003	10296	2493	3	0	12792
2004	8897	2155	4	1	11057
2005	10423	2549	2	1	12975
2006	9589	2113	7	0	11709
2007	8987	1841	8	0	10836
Total	74355	17474	42	3	91874

year	orddivv				Total
	missing	ordered no problem	missings in the middle	none missing but out of sequence	
2000	8587	1629	16	1	10233
2001	9575	1724	37	2	11338
2002	9273	1641	20	0	10934
2003	10782	1984	26	0	12792
2004	9307	1729	21	0	11057
2005	10901	2045	29	0	12975
2006	9958	1729	22	0	11709
2007	9306	1511	19	0	10836
Total	77689	13992	190	3	91874

Tables MH14c

year	ordstrt			Total
	missing	ordered no problem	missings in the middle	
2000	3482	6740	11	10233
2001	4013	7309	16	11338
2002	4043	6865	26	10934
2003	4784	7992	16	12792
2004	4301	6737	19	11057
2005	4936	8020	19	12975
2006	4699	6996	14	11709
2007	4696	6130	10	10836
Total	34954	56789	131	91874

year	ordma				Total
	missing	ordered no problem	missings in the middle	none missing but out of sequence	
2000	3468	6758	6	1	10233
2001	3987	7335	13	3	11338
2002	4018	6908	6	2	10934
2003	4770	8008	13	1	12792
2004	4391	6652	12	2	11057
2005	4915	8049	8	3	12975
2006	4688	7013	6	2	11709
2007	4682	6147	5	2	10836
Total	34919	56870	69	16	91874

year	ordendlv				Total
	missing	ordered no problem	missings in the middle	none missing but out of sequence	
2000	8199	2024	10	0	10233
2001	9145	2182	10	1	11338
2002	8819	2110	3	2	10934
2003	10296	2491	4	1	12792
2004	8897	2152	6	2	11057
2005	10423	2546	5	1	12975
2006	9589	2113	7	0	11709
2007	8987	1844	5	0	10836
Total	74355	17462	50	7	91874

year	orddiv				Total
	missing	ordered no problem	missings in the middle	none missing but out of sequence	
2000	8587	1628	17	1	10233
2001	9575	1721	40	2	11338
2002	9273	1640	20	1	10934
2003	10782	1982	27	1	12792
2004	9307	1726	23	1	11057
2005	10901	2042	32	0	12975
2006	9958	1729	22	0	11709
2007	9306	1514	16	0	10836
Total	77689	13982	197	6	91874

Postcheck marriage 15

For 3 marriages:
Presence of the year of marriage, year of cohabitation and year of end, and when not, presence of any other date for this marriage.

Tables MH15

year	varmar1					Total
	missing	year marriage present	no yr marriage but yr premar	no yr marriage but yr death	no yr marriage but yr sep	
2000	3465	6760	0	1	7	10233
2001	3986	7340	1	4	7	11338
2002	4014	6911	2	1	6	10934
2003	4761	8011	4	1	15	12792
2004	4269	6655	116	4	13	11057
2005	4906	8055	3	2	9	12975
2006	4675	7016	7	1	10	11709
2007	4676	6152	4	0	4	10836
Total	34752	56900	137	14	71	91874

year	varmar2					Total
	missing	year marriage present	no yr marriage but yr premar	no yr marriage but yr death	no yr marriage but yr sep	
2000	9204	1025	2	0	2	10233
2001	10194	1139	1	0	4	11338
2002	9821	1111	0	2	0	10934
2003	11581	1201	4	1	5	12792
2004	9943	1106	5	0	3	11057
2005	11716	1249	6	0	4	12975
2006	10628	1077	3	0	1	11709
2007	9905	924	4	1	2	10836
Total	82992	8832	25	4	21	91874

year	varmar3				Total
	missing	year marriage present	no yr marriage but yr death	no yr marriage but yr sep	
2000	10148	85	0	0	10233
2001	11230	107	1	0	11338
2002	10841	93	0	0	10934
2003	12677	115	0	0	12792
2004	10957	99	0	1	11057
2005	12860	115	0	0	12975
2006	11618	91	0	0	11709
2007	10747	89	0	0	10836
Total	91078	794	1	1	91874

year	varendm1				Total
	missing	year end marriage present	no yr end but yr div	death of the spouse	
2000	8186	1891	21	135	10233
2001	9135	2017	16	170	11338
2002	8808	1969	14	143	10934
2003	10288	2346	11	147	12792
2004	8886	2012	13	146	11057
2005	10411	2397	14	153	12975
2006	9586	2001	10	112	11709
2007	8990	1752	5	89	10836
Total	74290	16385	104	1095	91874

year	varendm2				Total
	missing	year end marriage present	no yr end but yr div	death of the spouse	
2000	9986	225	3	19	10233
2001	11047	265	1	25	11338
2002	10664	245	2	23	10934
2003	12470	299	0	23	12792
2004	10783	245	2	27	11057
2005	12646	306	1	22	12975
2006	11435	253	0	21	11709
2007	10590	225	1	20	10836
Total	89621	2063	10	180	91874

year	varendm3				Total
	missing	year end marriage present	no yr end but yr div	death of the spouse	
2000	10207	25	0	1	10233
2001	11309	26	0	3	11338
2002	10907	26	0	1	10934
2003	12759	27	1	5	12792
2004	11033	20	0	4	11057
2005	12951	21	0	3	12975
2006	11687	21	0	1	11709
2007	10811	23	0	2	10836
Total	91664	189	1	20	91874

Postcheck marriage 16

Check the order of the beginning and end (if exists) within each marriage after the corrections. And if they exist, of premarital cohabitation and marriage, and of separation and divorce.

Tables MH16

Table of year by chkin1						
year	chkin1					Total
	missing	no problem	mar date>=end date	pre mar coh date>mar date	end date>div date	
2000	3472	6721	11	3	26	10233
2001	3998	7306	10	0	24	11338
2002	4023	6875	15	0	21	10934
2003	4779	7947	13	0	53	12792
2004	4400	6607	15	0	35	11057
2005	4920	7991	31	0	33	12975
2006	4692	6949	36	0	32	11709
2007	4685	6097	30	0	24	10836
Total	34969	56493	161	3	248	91874

year	chkin2					Total
	missing	no problem	mar date>=end date	pre mar coh date>mar date	end date>div date	
2000	9216	1009	4	0	4	10233
2001	10212	1118	5	0	3	11338
2002	9840	1089	2	0	3	10934
2003	11601	1177	9	0	5	12792
2004	9960	1093	3	0	1	11057
2005	11738	1228	5	0	4	12975
2006	10644	1052	6	1	6	11709
2007	9920	906	6	1	3	10836
Total	83131	8672	40	2	29	91874

year	chkin3				Total
	missing	no problem	mar date>=end date	end date>div date	
2000	10149	83	1	0	10233
2001	11237	100	0	1	11338
2002	10843	89	2	0	10934
2003	12679	112	1	0	12792
2004	10962	94	0	1	11057
2005	12864	111	0	0	12975
2006	11620	89	0	0	11709
2007	10748	88	0	0	10836
Total	91102	766	4	2	91874

Tables MH16a Proportion with mar date>=end date, by marriage order (in %)

	Partnership order			
year	1	2	3	4
2000	0.55	1.65	3.85	0.00
2001	0.46	1.77	0.00	0.00
2002	0.71	0.75	7.41	0.00
2003	0.53	2.88	3.13	0.00
2004	0.71	1.14	0.00	0.00
2005	1.22	1.56	0.00	0.00
2006	1.76	2.21	0.00	0.00
2007	1.63	2.52	0.00	0.00

Tables MH16b Proportion with pre mar coh date>mar date, by marriage order (in %)

	Partnership order			
year	1	2	3	4
2000	0.04	0.00	0.00	0.00
2001	0.00	0.00	0.00	0.00
2002	0.00	0.00	0.00	0.00
2003	0.00	0.00	0.00	0.00
2004	0.00	0.00	0.00	0.00
2005	0.00	0.00	0.00	0.00
2006	0.00	0.10	0.00	0.00
2007	0.00	0.11	0.00	0.00

Tables MH16c Proportion with end date>div date, by marriage order (in %)

	Partnership order			
year	1	2	3	4
2000	1.61	2.17	0.00	0.00
2001	1.40	1.46	4.76	0.00
2002	1.29	1.60	0.00	0.00
2003	2.68	2.13	0.00	0.00
2004	2.04	0.54	6.67	0.00
2005	1.62	1.69	0.00	0.00
2006	1.86	3.00	0.00	0.00
2007	1.59	1.74	0.00	0.00

Postcheck marriage 17

If the person reports a marriage of order i>1, do we have the elements about the previous marriage and did they all happen before the beginning of the i^{th}?

Tables MH17

year	chkout1						Total
	missing	marr date for i+1, no problem	marr date for i+1 but no marr date for i	marr date for i+1 but no end date for i	marr date i>marr date i+1	end date i>marr date i+1	
2000	9216	986	3	25	1	2	10233
2001	10212	1091	9	22	2	2	11338
2002	9840	1052	5	33	0	4	10934
2003	11601	1167	8	15	0	1	12792
2004	9960	1069	8	16	1	3	11057
2005	11738	1207	5	20	2	3	12975
2006	10644	1034	5	20	2	4	11709
2007	9920	887	3	21	2	3	10836
Total	83131	8493	46	172	10	22	91874

year	chkout2						Total
	missing	marr date for i+1, no problem	marr date for i+1 but no marr date for i	marr date for i+1 but no end date for i	marr date i>marr date i+1	end date i>marr date i+1	
2000	10149	81	1	2	0	0	10233
2001	11237	96	2	2	1	0	11338
2002	10843	85	1	4	0	1	10934
2003	12679	109	3	1	0	0	12792
2004	10962	90	2	2	1	0	11057
2005	12864	105	3	1	1	1	12975
2006	11620	86	1	1	0	1	11709
2007	10748	83	1	4	0	0	10836
Total	91102	735	14	17	3	3	91874

year	chkout3				Total
	missing	marr date for i+1, no problem	marr date for i+1 but no end date for i	end date i>marr date i+1	
2000	10227	6	0	0	10233
2001	11328	9	0	1	11338
2002	10926	8	0	0	10934
2003	12786	6	0	0	12792
2004	11048	9	0	0	11057
2005	12970	5	0	0	12975
2006	11704	5	0	0	11709
2007	10829	6	1	0	10836
Total	91818	54	1	1	91874

Postcheck marriage 18

These tables show the number of marriages that are overlapping each other after corrections.

Tables MH18

Overlapping marriages 1 and 2			
year	no	yes	% yes
2000	1007	2	0.2%
2001	1110	2	0.2%
2002	1068	4	0.4%
2003	1184	1	0.1%
2004	1090	4	0.4%
2005	1225	3	0.2%
2006	1048	4	0.4%
2007	899	3	0.3%
Total	8631	23	0.3%

Overlapping marriages 2 and 3			
year	no	yes	% yes
2000	82	0	0.0%
2001	99	0	0.0%
2002	87	1	1.1%
2003	112	0	0.0%
2004	93	0	0.0%
2005	109	1	0.9%
2006	87	1	1.1%
2007	85	0	0.0%
Total	754	3	0.4%

Overlapping marriages 3 and 4			
year	no	yes	% yes
2000	6	0	0.0%
2001	9	1	10.0%
2002	8	0	0.0%
2003	6	0	0.0%
2004	10	0	0.0%
2005	5	0	0.0%
2006	5	0	0.0%
2007	6	0	0.0%
Total	55	1	1.8%

Classification of errors 21: MARQU1-n (summary classification[11])

		date in spell n	mar date	end date	order
0	no marriage of that order declared				
1	no problem	yes	yes	yes	yes
2	OK after editing	yes	yes edited	yes edited	yes edited
3	partially OK	yes	yes	no	yes
4	unusable	yes	no		OR no
-9	DNA: not eligible FI, proxy				
-8	Missing				

Explanation:
Code 1 "no problem": when the marriage has at least a beginning date, an end date (if this applies) and that the end date>beginning date.
Code 2: where code 1 applies but that at least one of the dates present has been edited
Code 3: where a beginning date is present but an end date which should be present is missing
Code 4: missing beginning date or beginning and end date out of sequence

Tables MH21

year	marqu1					Total
	no nth marr declared	no problem	ok after editing	partially ok	error	
2000	3456	6647	1	110	19	10233
2001	3966	7211	0	128	33	11338
2002	3993	6829	0	80	32	10934
2003	4738	7909	1	99	45	12792
2004	4217	6621	1	32	186	11057
2005	4881	7971	1	75	47	12975
2006	4662	6947	2	54	44	11709
2007	4659	6088	3	50	36	10836
Total	34572	56223	9	628	442	91874

[11] For the full classification see the User Guide.

year	marqu2					Total
	no nth marr declared	no problem	ok after editing	partially ok	error	
2000	9214	983	0	34	2	10233
2001	10206	1098	0	27	7	11338
2002	9836	1085	0	9	4	10934
2003	11586	1173	0	16	17	12792
2004	9949	1091	0	6	11	11057
2005	11727	1228	0	9	11	12975
2006	10638	1053	0	10	8	11709
2007	9912	906	2	7	9	10836
Total	83068	8617	2	118	69	91874

year	marqu3				Total
	no nth marr declared	no problem	partially ok	error	
2000	10149	80	4	0	10233
2001	11237	99	2	0	11338
2002	10843	91	0	0	10934
2003	12679	110	3	0	12792
2004	10961	95	0	1	11057
2005	12864	109	2	0	12975
2006	11620	87	2	0	11709
2007	10748	86	2	0	10836
Total	91101	757	15	1	91874

year	marqu4				Total
	no nth marr declared	no problem	partially ok	error	
2000	10227	6	0	0	10233
2001	11328	10	0	0	11338
2002	10926	8	0	0	10934
2003	12786	6	0	0	12792
2004	11047	8	1	1	11057
2005	12970	5	0	0	12975
2006	11704	5	0	0	11709
2007	10829	7	0	0	10836
Total	91817	55	1	1	91874

year	marqu5		Total
	no nth marr declared	no problem	
2000	10233	0	10233
2001	11337	1	11338
2002	10934	0	10934
2003	12792	0	12792
2004	11056	1	11057
2005	12975	0	12975
2006	11709	0	11709
2007	10836	0	10836
Total	91872	2	91874

MARQUAL

		number of dates	marr ok	marr edited	marr partially ok	marr unusable	overlap
0	No marriage declared	0					
1	no problem	>=1	yes	no	no	no	no
2	OK after editing	>=1		yes	no	no	OR yes edited
3	Partially OK	>=1			yes	no	no
4	error: inconsistent number	incons with decl					
5	error: other	>=1				yes	OR yes
-9	DNA: not elig FI, proxy						
-8	Missing						

The coding is, for marriages, as for PARTQUAL, explained above.

Tables MH21q

year	marqual							Total
	missing	no marr declared	no problem	ok after editing	partially ok	error: inconsistent number	error: other	
2000	158	3286	6603	1	144	12	29	10233
2001	324	3633	7174	0	153	19	35	11338
2002	428	3552	6814	0	82	13	45	10934
2003	530	4174	7879	0	112	24	73	12792
2004	530	3667	6601	0	35	24	200	11057
2005	570	4233	7944	1	81	22	124	12975
2006	817	3815	6801	2	63	123	88	11709
2007	899	3552	5977	2	56	92	258	10836
Total	4256	29912	55793	6	726	329	852	91874

ANNEX D DERIVATION OF THE QUALITY FLAGS FOR THE MARRIAGE AND PARTNERSHIP HISTORIES

See also annex C for fuller details.

Construction of the **partq1-11** variables

		date in spell n	beg date	end date	order
0	no date present in the spell	no			
1	no problem	yes	yes	yes	yes
2	OK after editing	yes	yes edited	yes edited	yes edited
3	partially OK	yes	yes	no	yes
4	unusable	yes	no		OR no
-9	DNA: not eligible FI, proxy				
-8	Missing				

Construction of the **partqual** variable

		number of dates	partn ok	partn edited	partn partially ok	partn unusable	overlap
0	No partnership declared	0					
1	no problem	>=1	yes	no	no	no	no
2	OK after editing	>=1		yes	no	no	OR yes edited
3	Partially OK	>=1			yes	no	no
4	error: inconsistent number	incons with decl					
5	error: other	>=1				yes	OR yes
-9	DNA: not elig FI, proxy						
-8	Missing						

Construction of the **marq1-11** variables

		date in spell n	mar date	end date	order
0	no marriage of that order declared				
1	no problem	yes	yes	yes	yes
2	OK after editing	yes	yes edited	yes edited	yes edited
3	partially OK	yes	yes	no	yes
4	unusable	yes	no		OR no
-9	DNA: not eligible FI, proxy				
-8	Missing				

C

		number of dates	marr ok	marr edited	marr partially ok	marr unusable	overlap
0	No marriage declared	0					
1	no problem	>=1	yes	no	no	no	no
2	OK after editing	>=1		yes	no	no	OR yes edited
3	Partially OK	>=1			yes	no	no
4	error: inconsistent number	incons with decl					
5	error: other	>=1				yes	OR yes
-9	DNA: not elig FI, proxy						
-8	Missing						

ANNEX E SPSS CODING OF DEFACTO (DE FACTO PARTNERSHIP STATUS)

SPSS Syntax consistent for years 1979-1996/7 using the year 1979 as an example. In later years the exact variable levels and variable names differ. Some categories in **defacto** in this set of years cannot be generated as the category does not apply (e.g. civil partnership).

```
compute defacto=-9.
execute.
*default to NA code.
do if (marital=1).
recode
defacto
(-9=11).
end if.
execute.
*marital=1 indicates respondent is married.
do if ((marital=2)|(marital=3)|(marital=4)|(marital=5)).
recode
defacto
(-9=13).
end if.
execute.
*2=single, 3,4, 5 indicates widowed, separated and divorced.
variable labels defacto 'defacto indicator'.
value labels defacto
1 'Living with spouse- FI' 2 'Living with partner- FI' 3 'Not living with anyone-FI'
11 'Living with spouse- non FI' 12 'Living with partner- non FI' 13 'Not living with anyone- non
FI' 14 ' Same sex civil partnership- non FI' 15 'Living with same sex partner- non FI'.
*Set value labels.

do if (age<16).
compute defacto=-9.
end if.
execute.
*exclude the under 16s from the coding process.

missing values age().
do if (age=-9).
compute defacto=-9.
end if.
execute.
*exclude those with missing age from the coding process.

do if (fiself<4).
recode
defacto
(11=1) (12=2) (13=3).
end if.
execute.
do if (fiself<0).
recode
defacto
(1=11) (2=12) (3=13).
end if.
execute.
*recode those eligible for the Family information to the relevant code.
```

```
do if (wherewed=5).
recode
defacto
(1=2).
end if.
execute.
*Those not actually married are recoded from married to cohabiting.

do if (wherewed=4).
recode
defacto
(1=2) (3=2).
end if.
execute.
*Those WDS not married are recoded from married to cohabiting.

missing values wherewed husaway xslivtog ().
do if (wherewed=8).
recode
defacto
(1=11) (2=12) (3=13).
end if.
execute.
*Exclude those with missing wherewed values from the FI codes.

do if (husaway=3).
recode
defacto
(1=3) .
end if.
execute.
*Identify the respondents whose spouse is not in the household due to
*marital dissolution.

do if (husaway=8).
recode
defacto
(1=11) (2=12) (3=13).
end if.
execute.
*Exclude the respondents with a missing husbaway code.

do if (xslivtog=1).
recode
defacto
(1=2) (3=2).
end if.
execute.
*Identify the single respondents and recode to cohabiting if living with a partner.

do if (xslivtog=8).
recode
defacto
(1=11) (2=12) (3=13) .
end if.
execute.
*Exclude the respondents with a missing xslivtog code.

*Syntax consistent for years 1998/99-2007 using the year 2000 as an example*.
* In years after 2000 the exact variable levels and variable names may differ*.
```

```
compute defacto=-9.
execute.
*Start from the position of NA.

variable labels defacto 'Partnership indicator'.
value labels defacto
1 'Living with spouse- FI' 2 'Living with partner- FI' 3 'Not living with anyone-FI'
11 'Living with spouse- non FI' 12 'Living with partner- non FI' 13 'Not living with anyone- non
FI' 14 ' Same sex civil partnership- non FI' 15 'Living with same sex partner- non FI'.
*Set up the variable and value labels.

do if (marstat=2).
recode
defacto
(-9=11).
end if.
execute.
*Those married defined as married.

do if ((marstat=1)|(marstat=3)|(marstat=4)|(marstat=5)).
recode
defacto
(-9=13).
end if.
execute.
*Those who are single, widowed, separated and divorced as defined as not living with anyone.


do if (livewith=1).
recode
defacto
(-9=12) (11=12) (13=12).
end if.
execute.
*Those who reported living with a partner are defined as cohabiting.

do if (livewith=3).
recode
defacto
(-9=15) (13=15).
end if.
execute.
*Same sex cohabitation.

do if (age<16).
compute defacto=-9.
end if.
execute.
*Exclude the under-16s from the coding process.

do if (selfcom3=1).
recode
defacto
(11=1) (12=2) (13=3).
end if.
execute.
do if (selfcom3=2).
recode
defacto
```

```
(11=1) (12=2) (13=3).
end if.
execute.
do if (selfcom3=3).
recode
defacto
(11=1) (12=2) (13=3).
end if.
execute.
*Identify the respondents who repond to the Family information.

do if (wherewed=4).
recode
defacto
(1=2).
end if.
execute.
*Those who are not actually married are defined as cohabiting.

missing values wherewed husbaway tgthr2 ().
do if (wherewed=-8).
recode
defacto
(1=11) (2=12) (3=13).
end if.
EXECUTE .
*Respondents with a missing wherewed code are defined as non-respondents to the FI.

do if (husbaway=2).
recode
defacto
(1=3).
end if.
execute.
*Defining those with a spouse not in the household due to marital
*breakdown are defined as not living with anyone.

do if (husbaway=-8).
recode
defacto
(1=11) (2=12) (3=13).
end if.
EXECUTE .
*Non respondents to husbaway are defined as not responding to FI.

do if (tgthr2=1).
recode
defacto
(1=2) (3=2).
end if.
execute.
*Those single and living with a non-related partern are defined as cohabiting.

do if (tgthr2=-8).
recode
defacto
(1=11) (2=12) (3=13).
end if.
EXECUTE .
*Non respondents to tgthr are defined as not responding to FI.
```

ANNEX F ELIGIBILITY FOR THE CONTRACEPTION/FERTILITY INTENTIONS QUESTIONS.

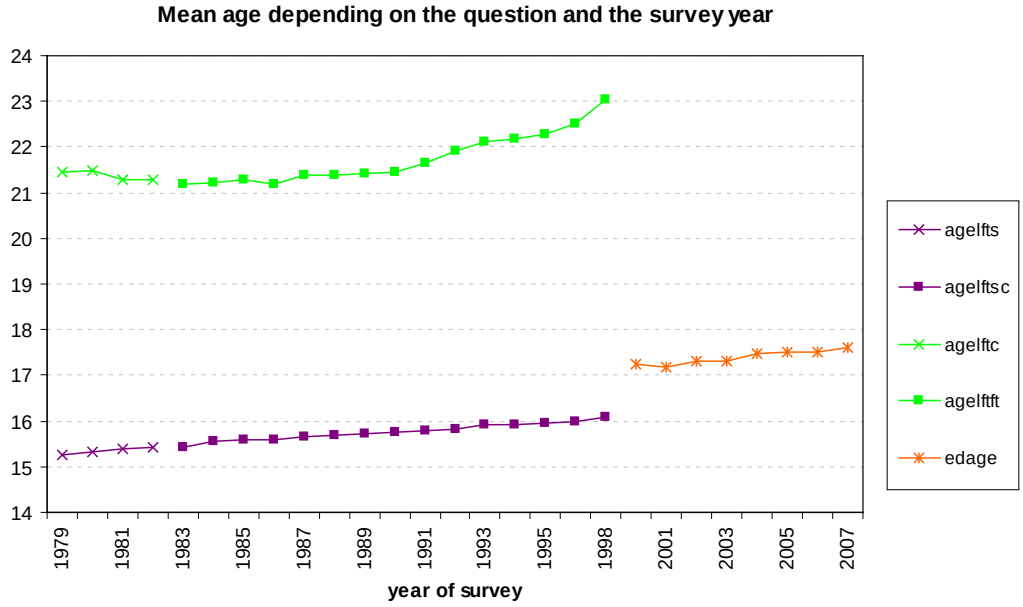
See accompanying Excel file on eligibility for several Family Information section questions.

ANNEX G QUALITY CHECKS ON AGE AT LEAVING FULL TIME EDUCATION 1979-2007

Aim:
To construct a variable 'age at the end of the full time education' that is consistent over time.

Construction:
Variables giving the age at end of school and/or at end of full time education:
Until 1998, the age at leaving school is asked first, and after a few other questions and filters, the age at end of full time education. From 2000, only one age is asked of everybody, which is the age at end of the continuous full time education. For check, the means are produced in this Figure G1.

Figure G1

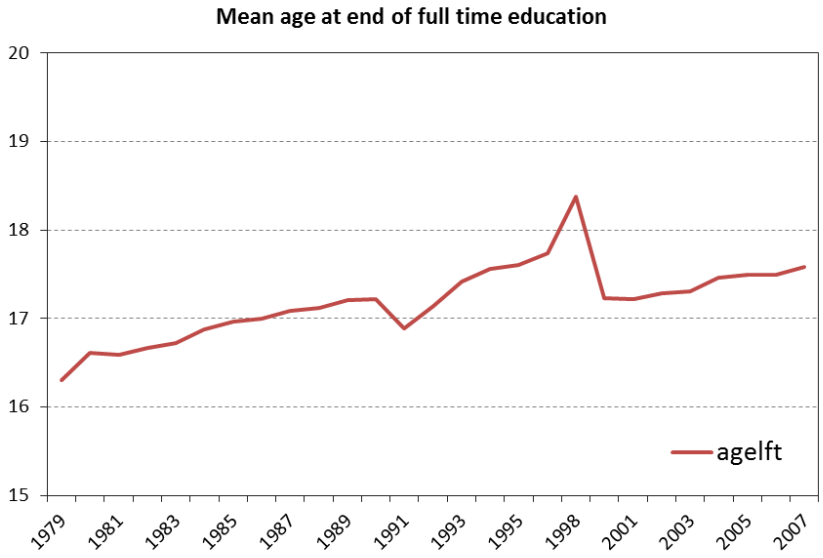


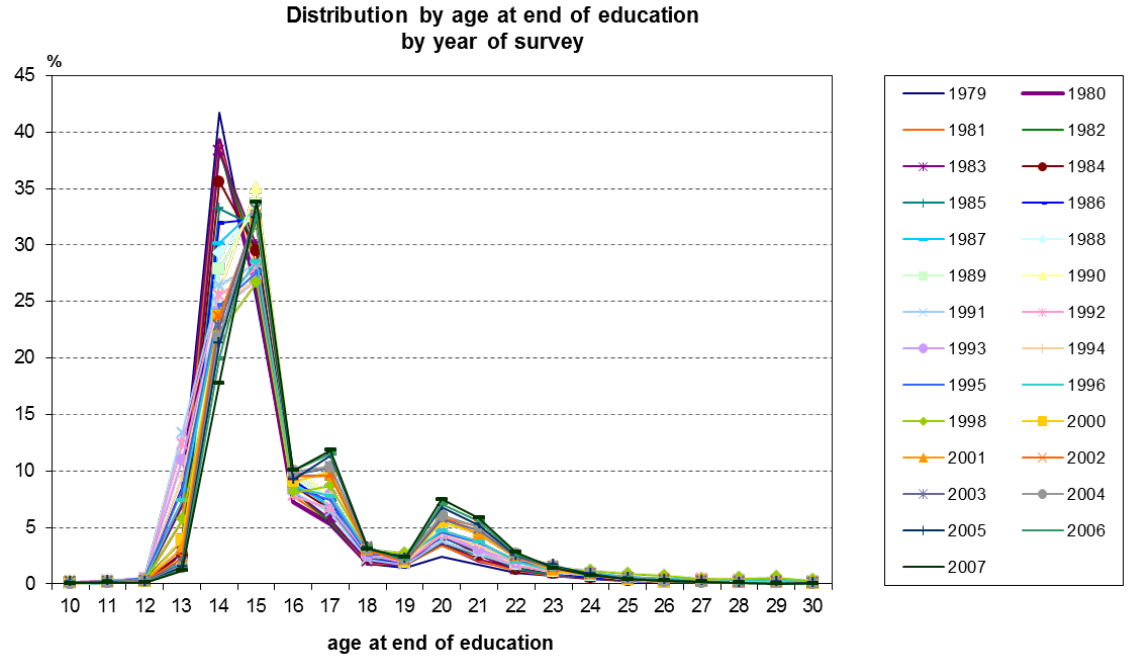
Source: GHS 1979-2007
Sample: Persons aged 16-69 at survey

We combined the variables of age at end of school and end of full time education (for a selected group) into one variable (named **agelft**), in an attempt to produce an age at end of full time education for everyone, and intended to have the same definition as the variable EDAGE that is present from 2000 onwards.
Using the variable FURTHED which indicates whether the person attended further education after school (only for people aged 16-49 before 1991) and the variable LASTSCH (EDTYPE before 1983) which is a further filter giving details on the type

of courses attended after school[12] for everybody, we assigned to the new variable **agelft** the value of AGELFTS(C) if the person did not go on to further study after leaving school, and AGELFTC(FT) if the person continued studying after school.
The mean and age distribution of the new variable are shown in Figures G2a and G2b, for men and women aged 16-69.

Figures G2a and G2b



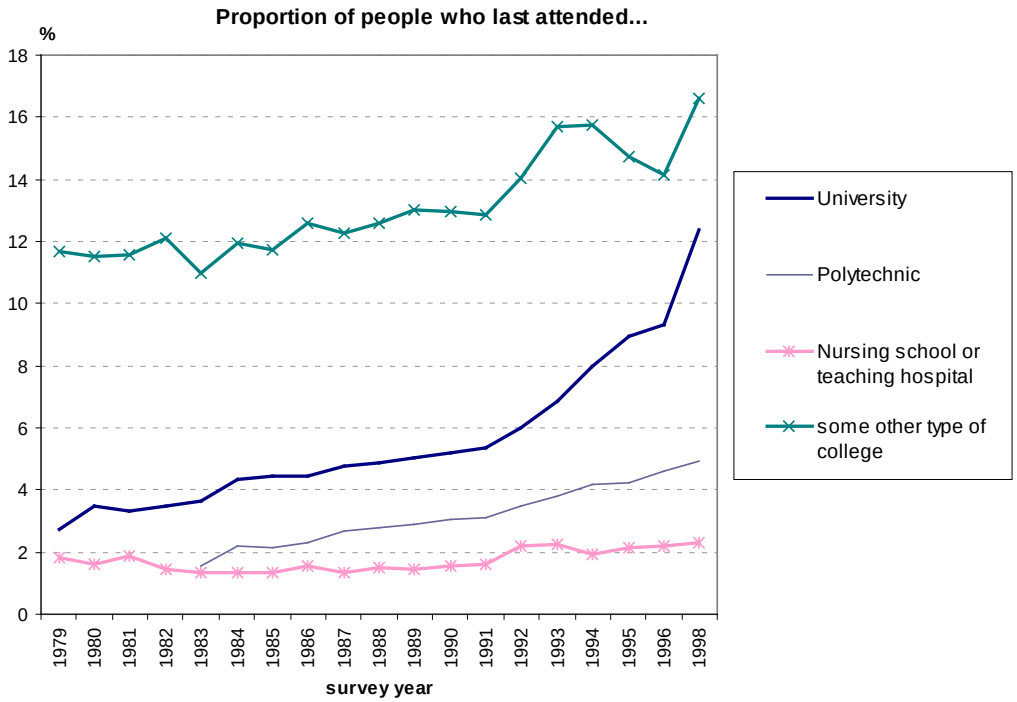


Source: GHS 1979-2007
Sample: persons aged 16-69 at survey

[12] Now thinking of your full-time education, what type of school or college did you last attend full-time? Was it: "elementary or secondary school", "university", "polytechnic", nursing school or teaching hospital", "or some other type of college?"

There is an anomaly in the mean age at the end of full time education between 1991 and 1998.
To understand where this hump comes from, we first looked at the distribution of respondents in the different types of education (Figure G3: we can't continue the curve after 1998 because there is no equivalent to that question).

Figure G3

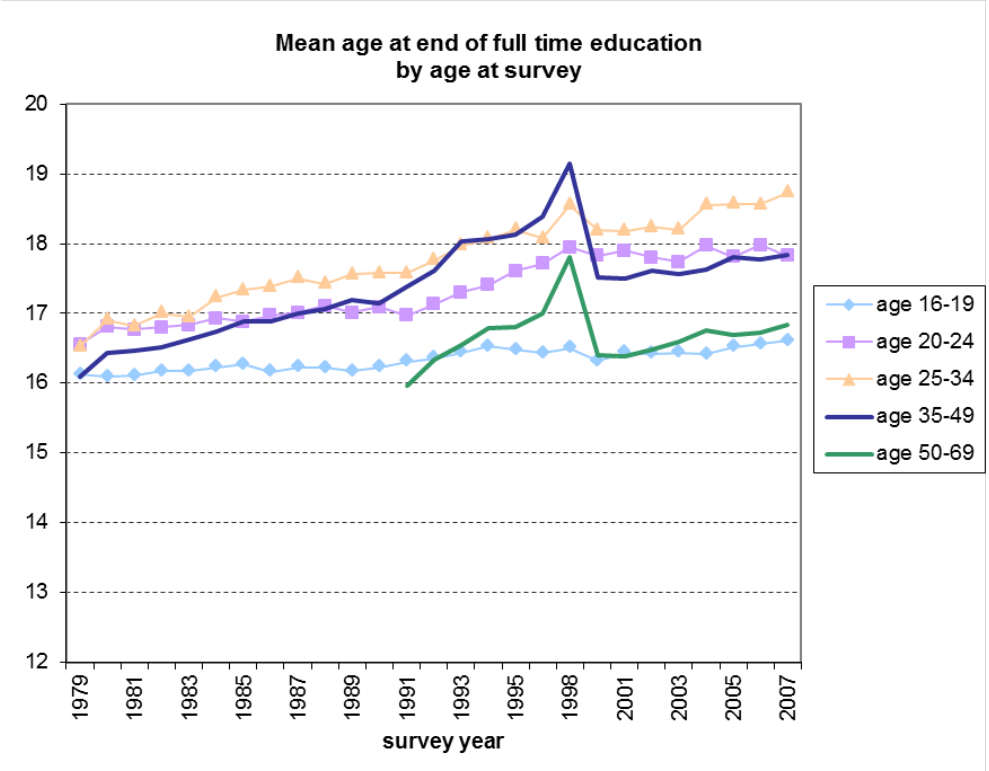


Source: GHS 1979-2007
Sample: Persons aged 16-69 at survey and who have had education

It appears that the proportion of people who have attended university rose very rapidly after 1991, which is believable, but 1998 looks too high compared with the previous years. There is also a discontinuity for the "other type of college" answer in 1998.

One particularly important change in the questionnaire occurs between 1990 and 1991: the age filter on the question FURTHRED has changed, from 16-49 to 16-69. This suggests looking at the mean age at end of full time education by age group at survey (Figure G4).

Figure G4

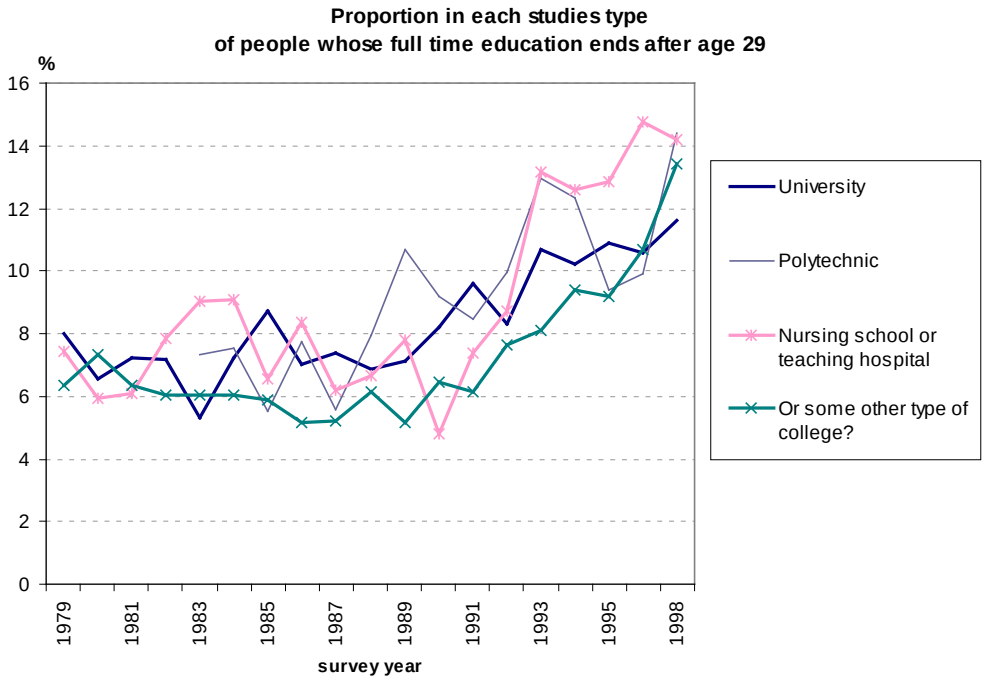


Source: GHS 1979-2007
Sample: Persons aged 16-69 at survey and who have had education

The problem in 1991-1998 appears limited to people age 35+ at survey. In 1992 there is a change in the order of the questions, with a stronger separation of people currently studying vs. others. Even if this does not seem to have a direct impact, and if the rest of the wording of the questions has not changed, it is possible that it changes the way of understanding the question on age at end of the full time education, and people would more often give other courses they have had in between, possibly including part-time courses.

For example among persons who have studied in nursing school or at teaching hospital, we see a strong increase in the proportion of persons who declare having finished nursing studies after age 30 (Figure G5).

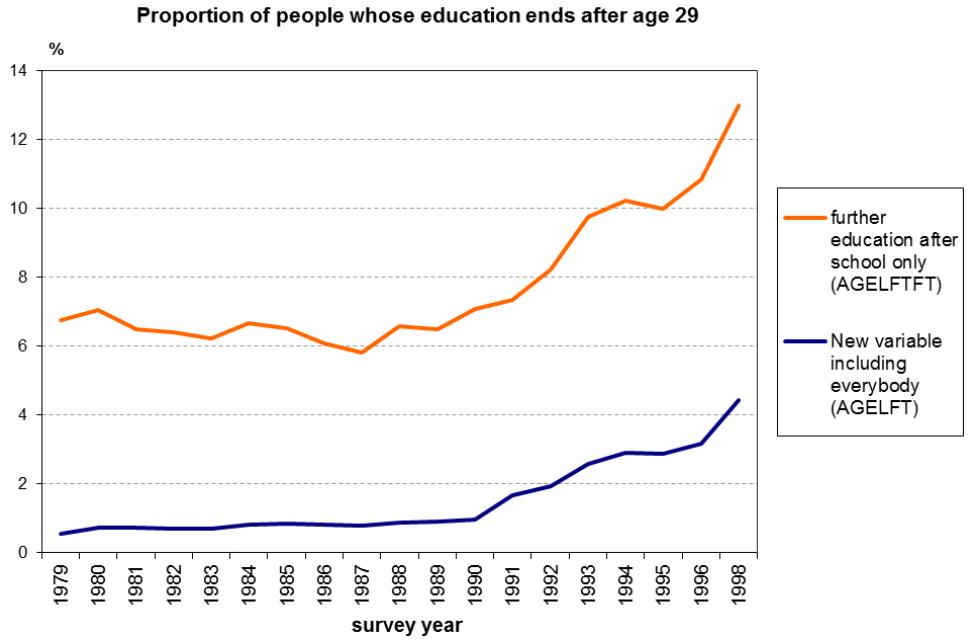
Figure G5



Source: GHS 1979-2007
Sample: Persons aged 16-69 at survey and who have taken further education

We notice more generally that there appears to be an especially strong proportion of persons who finish their studies after age 30 in the problematic years (Figure G6).

Figure G6

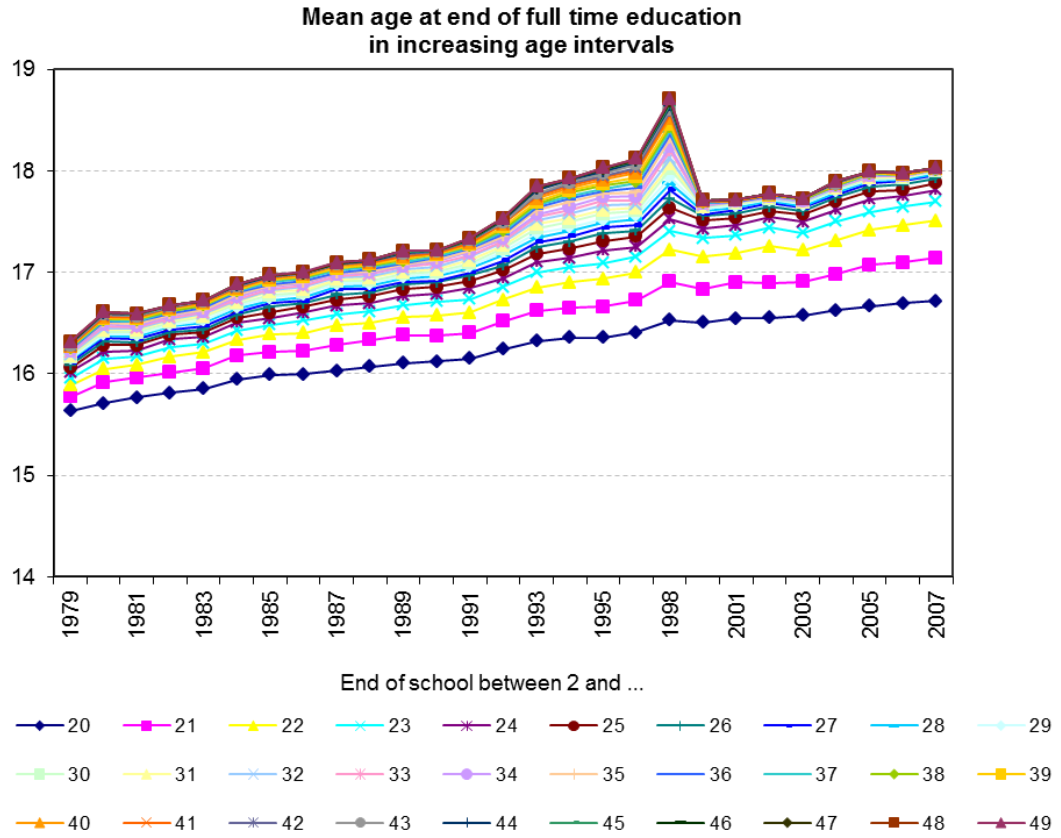


Source: GHS 1979-2007
Sample: All persons aged 16-69 at survey; and those with further education

To deal with this implausible rise in age at end of full time education between 1991 and 1998, it is possible to compute the mean age at end of full time education only for people who finished their studies before a certain age.

Figure G7 shows this calculation for various age limits (for instance people aged 2-19 at the end of their studies, then people aged 2-20, etc.).

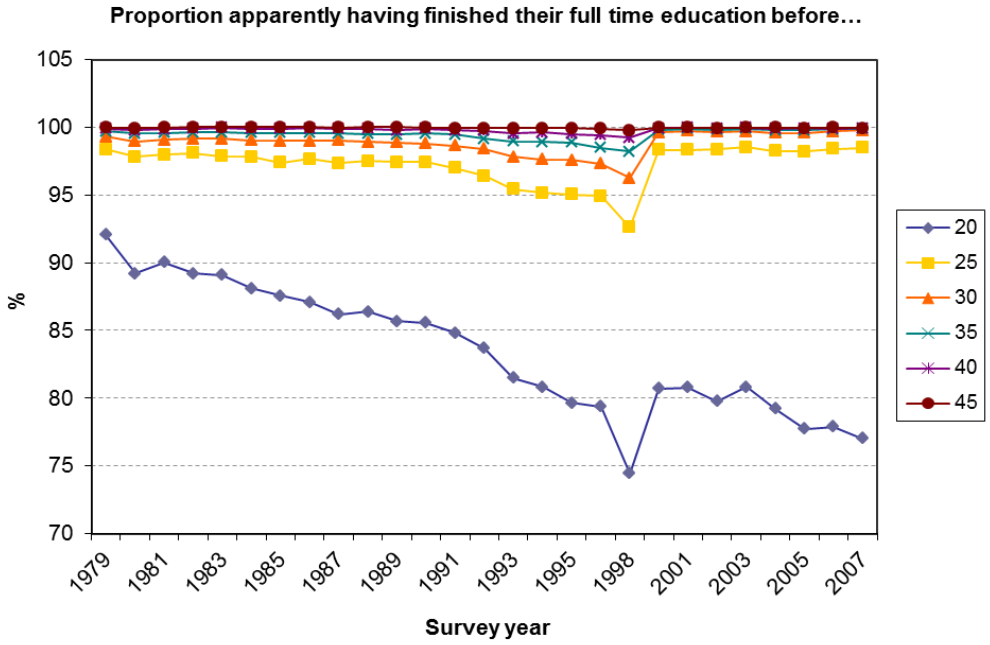
Figure G7



Source: GHS 1979-2007
Sample: Persons aged 16-49 at survey and who have had education

Considering only people who declared having finished before 25, there is some continuity over time. But the essential problem remains and there is no obvious solution. If taking the declared ages up to 24 and then top-coding the variable at 25+, there are two potential difficulties: (a) an especially high proportion of persons will still have finished at age 25+ using this solution (see Figure G8) and (b) most of those apparently completing at very late ages will almost certainly have left continuous education at a much earlier age, and so this potentially introduces a bias. As noted earlier, it is very likely that these persons are in fact persons who declared the age at end of part-time study or other forms of training.
The large increase in 1998 is probably linked with the especially large proportion of persons interviewed this year who declared having attended the university or some other type of college.

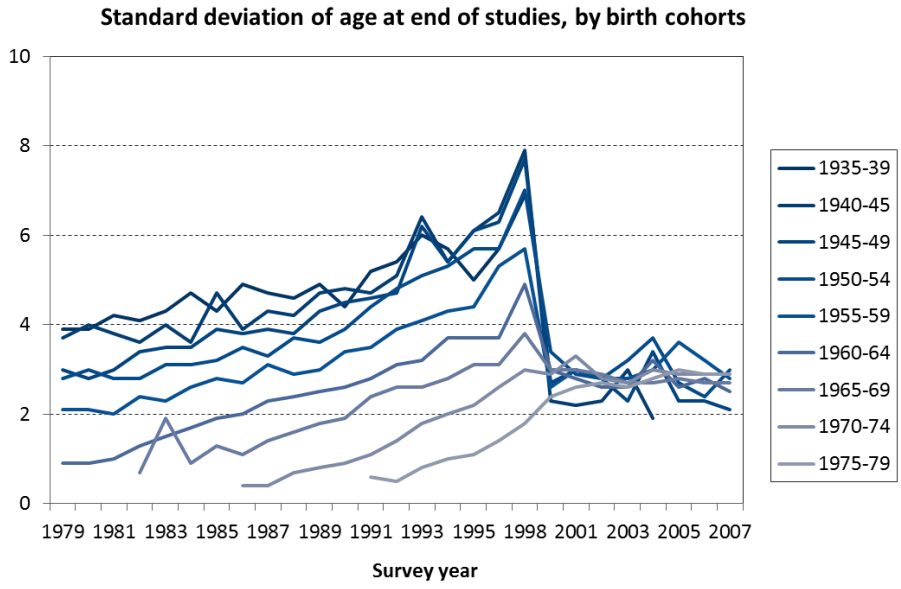
Figure G8



Source: GHS 1979-2007
Sample: Persons aged 16-49 at survey and who are not currently enrolled in education

Figure G9 plots the standard deviation of the age at leaving full time education (**agelft** in the CPC file) by both survey year and cohort (c1935=1935-39, 1940=c1940-44). We see that up to survey year 1998 the within cohort standard deviation increases systematically and substantially, well after the age at which one would have expected the within-cohort dispersion to stabilise. After 2000, however, when the question was revised, the standard deviation declines to a level that is both more stable and within a relatively narrow range. We take this as evidence of the greater validity of the variable **agelft** (in the CPC file; "edage" in the GHS file) from GHS round 2000 onwards.

Figure G9



ANNEX H SAMPLE SIZES IN THE CPC TIME-SERIES DATASET WITH NOTES OF SOME ERRORS IN THE ORIGINAL GHS DATA FILES

Samples sizes by survey year

The datasets provided by ESDS have been merged by household serial number or by household serial number and person number when necessary.

Table 1: Total number of persons in each database

	Main sample	Scotland	Total
1979	30705	-	30705
1980	31443	-	31443
1981	32410	-	32410
1982	27160	-	27160
1983	26587	3686	30273
1984	25354	2640	27994
1985	25555	3727	29282
1986	26073	3520	29593
1987	26419	3411	29830
1988-89	25350	3334	28684
1989-90	25269	3362	28631
1990-91	23663	3218	26881
1991-92	24657	3287	27944
1992-93	24535	3418	27953
1993-94	24079	3187	27266
1994-95	23622	-	23622
1995-96	23385	-	23385
1996-97	22274	-	22274
1998-99	20396	-	20396
2000-01	19266	-	19266
2001-02	21195	-	21195
2002-03	20149	-	20149
2003-04	24489	-	24489
2004-05	20421	-	20421
2005	30069	-	30069
2006	22924	-	22924
2007	21472	-	21472

We have retained only the Main sample, without the Scottish Supplement.
In 1982, 7 entries were duplicated. We have kept only one of each, thus 7 observations are dropped.
The 1972-2004 time-series dataset had 4 more individuals than in the original GHS and they have not been retained in the CPC dataset.

Year of birth

Date of birth is not available for respondents in GHS rounds 1979 to 1981. In 1982 to 1985 date of birth is available only for those answering the FI section. Date of birth is available for all respondents from GHS round 1986. In all years, age was collected for all respondents.

Table 2: Number of year of birth missing by year

year	brthyr(INFORMANT`S YEAR OF BIRTH)				Total
	missing	not missing	2202	2995	
1979	30705	0	0	0	30705
1980	31443	0	0	0	31443
1981	32410	0	0	0	32410
1982	21847	5313	0	0	27160
1983	21217	5370	0	0	26587
1984	20063	5291	0	0	25354
1985	20149	5406	0	0	25555
1986	51	26022	0	0	26073
1987	34	26385	0	0	26419
1988	37	25313	0	0	25350
1989	23	25246	0	0	25269
1990	11	23652	0	0	23663
1991	36	24621	0	0	24657
1992	34	24501	0	0	24535
1993	17	24062	0	0	24079
1994	24	23598	0	0	23622
1995	0	23385	0	0	23385
1996	0	22274	0	0	22274
1998	53	20343	0	0	20396
2000	32	19234	0	0	19266
2001	46	21134	0	0	21180
2002	71	20077	1	0	20149
2003	95	24394	0	0	24489
2004	39	20382	0	0	20421
2005	7367	27291	0	1	34659
2006	53	22871	0	0	22924
2007	45	21427	0	0	21472
Total	185902	487592	1	1	673496

Sample quarter

Table 3b: Number of individual by quarter and year, revised quarter variable

Year	**sampq** (Sample quarter)				Total
	1	2	3	4	
1998	5,208	5,242	4,713	5,233	20,396
2000	5,100	4,848	4,707	4,611	19,266
2001	3,934	6,092	5,693	5,461	21,180
2002	5,076	5,143	4,970	4,960	20,149
2003	5,210	5,374	6,849	7,056	24,489
2004	5,118	5,191	5,136	4,976	20,421
2005	0	8,470	8,430	8,193	25,093
2006	5,960	5,838	5,690	5,436	22,924
2007	5,247	5,629	5,429	5,167	21,472
2008	4,972	5,299	5,087	5,145	20,503
2009	4,767	4,887	4,788	4,546	18,988

The 1998 GHS data file from ESDS was found to have an error in codes 3 and 4 of SAMPQTR and an attempt has been made to correct these using the month of interview. All interviews either coded 3 originally or coded 4 and taking place in November and December 1998 are now coded 3. The rest of those coded 4 originally retain the code 4. This is not exactly accurate as interviews scheduled for a particular quarter sometimes take place in a later quarter. So a few of those retaining the code 4 in the 1998 survey may in fact belong to quarter 3 although the interview took place in January to March 1999.

Errors in the coding of sex

In 2006, some men appear to have answered the fertility history of the Family Information section, but this is due to an error in the coding of the variable SEX. These are, in fact, women who were correctly coded as female in the 2005 round, but mistakenly recoded as male in 2006 (in the second wave as the survey has become longitudinal in that year). They are not present in the CPC database for 2006-2009, only respondents in the survey for the first time (wave=1) were retained.

Ineligible women answering the FI section

Also in 2006, some women aged 60+ appear to have answered the FI section but these are women who were eligible in 2005 and who in 2006 on the repeat interview had reached their 60th birthday. We have not established why these respondents have answers recorded in the FI section in 2006, despite their ineligibility on age grounds. Possibly ONS recovered the history from the 2005 dataset or continued in 2006 using the same filter as the previous year. Again, they are not present in the CPC database as they were in the second or later wave in 2006-2009.

Year of interview

There are four years where a few dates of interview are missing (in all, 13 cases between 1998 and 2002). The year of survey has then been attributed as the year of interview and the month as 6.5.

The year of interview sometimes does not correspond to the year of the survey. When the interview year is after the survey year it is most often due to the normal process of interviews that can take place until six months after the planned date. However we cannot explain what happens when the interview year is earlier than the actual year of the survey (see Table 4). Some years that had been obviously miscoded (e.g. 1989 instead of 1998 for the 1998 survey year) have been recoded manually.

Table 4: Cases in which the recorded year of interview is before the year of the survey

year	problint		
	no problem	interview before year of survey	Total
1979	30705	0	30705
1980	31443	0	31443
1981	32410	0	32410
1982	27160	0	27160
1983	26587	0	26587
1984	25354	0	25354
1985	25555	0	25555
1986	26073	0	26073
1987	26419	0	26419
1988	25350	0	25350
1989	25269	0	25269
1990	23663	0	23663
1991	24657	0	24657
1992	24535	0	24535
1993	24079	0	24079
1994	23622	0	23622
1995	23385	0	23385
1996	22228	46	22274
1998	20393	3	20396
2000	19266	0	19266
2001	21177	3	21180
2002	20146	3	20149
2003	24485	4	24489
2004	20421	0	20421
2005	25086	7	25093
2006	22912	12	22924
2007	21466	6	21472
Total	663846	84	663930

Age and year of birth

The age calculated using the date of birth together with the date interview is not always consistent with the age reported in the survey. For some the difference is just one year, which may be mainly due to the fact that the day of birth and of interview are not available.

However, Table 5 shows that the difference between the reported and calculated age can be greater than a year. This is especially the case in 2001, due to the top-coding of age at 85 (ages 85 and higher being coded 85+). For some other survey years it appears as a miscoding, sometime obvious when the calculated age is 10 or 20 years different from the reported age.

These inconsistencies have not been corrected in the CPC dataset. Users may choose how to handle the issue.

Table 5: Total number of persons in each dataset

year	problages		
	no problem	age calculated strictly different from age+/-1	Total
1979	30705	0	30705
1980	31443	0	31443
1981	32410	0	32410
1982	27131	29	27160
1983	26549	38	26587
1984	25319	35	25354
1985	25515	40	25555
1986	26061	12	26073
1987	26387	32	26419
1988	25326	24	25350
1989	25252	17	25269
1990	23648	15	23663
1991	24636	21	24657
1992	24515	20	24535
1993	24062	17	24079
1994	23620	2	23622
1995	23382	3	23385
1996	22270	4	22274
1998	20394	2	20396
2000	19266	0	19266
2001	21000	180	21180
2002	20149	0	20149
2003	24488	1	24489
2004	20421	0	20421
2005	25091	2	25093
2006	22902	22	22924
2007	21450	22	21472
Total	663392	538	663930

REFERENCES

Beaujouan, É., J. Brown, J., & Ní Bhrolchain, M. (2011) Reweighting the General Household Survey 1979-2007. *Population Trends* 145: 119-145. http://www.ons.gov.uk/ons/rel/population-trends-rd/population-trends/no--145--autumn-2011/index.html

Beaujouan, É. & Ní Bhrolcháin, M. (2011) Cohabitation and marriage in Britain since the 1970's. *Population Trends* 145:35-59. http://www.ons.gov.uk/ons/rel/population-trends-rd/population-trends/no--145--autumn-2011/index.html

Berrington, A., Beaujouan, É., Lyons-Amos, M. & Ní Bhrolcháin, M. (2011) Evaluation of the Partnership Histories in the Centre for Population Change GHS Time Series Dataset. CPC Working Paper 12, ESRC Centre for Population Change. http://www.cpc.ac.uk/publications/2011_WP12_Evaluation_of_the_Partnership_Histories_Berrington_et_al.pdf

Ní Bhrolcháin, M., Beaujouan, É. & Berrington, A. (2010) Stability and change in fertility intentions in Britain, 1991-2007. *Population Trends* 141:1-23. http://eprints.soton.ac.uk/187557/1/05-poptrends141-bhrolchain.pdf

Ní Bhrolcháin, M. & Beaujouan, É. (2011) Uncertainty in fertility intentions in Britain 1979-2007. *Vienna Yearbook of Population Research* Vol. 9, 101-134. http://epub.oeaw.ac.at/0xc1aa500d_0x002a70f6.pdf

Ní Bhrolcháin, M. & Beaujouan, É. (2012) Fertility postponement is largely due to rising educational enrolment. *Population Studies* 66: 3: 311-327. http://www.tandfonline.com/doi/abs/10.1080/00324728.2012.697569

Ní Bhrolcháin, M., Beaujouan, É. & Murphy, M. (2011) Sources of error in reported childlessness in a continuous British household survey. *Population Studies* 65:3, 305-318. http://www.tandfonline.com/doi/full/10.1080/00324728.2011.607901

ISSN 2042-4116

ESRC Centre for Population Change
Building 58, Room 2001
Faculty of Social and Human Sciences
University of Southampton
SO17 1BJ

T: +44 (0)2380 592579
E: cpc@soton.ac.uk
www.cpc.ac.uk

To subscribe to the CPC newsletter and keep up-to-date with research activity, news and events, please register online: www.cpc.ac.uk/newsletter

For our latest research updates you can also follow CPC on Twitter, Facebook and Mendeley:

www.facebook.com/CPCpopulation

www.twitter.com/CPCpopulation

www.mendeley.com/groups/3241781/centre-for-population-change

The ESRC Centre for Population Change (CPC) is a joint initiative between the University of Southampton and a consortium of Scottish universities including St Andrews, Edinburgh, Stirling and Strathclyde, in partnership with the Office for National Statistics and National Records of Scotland.