Malawi - Cleaned spouse and marriage data

Malawi Epidemiology and Intervention Research Unit

Report generated on: November 19, 2025

Visit our data catalog at: http://192.168.0.96/microdataportal/index.php

Identification

SURVEY ID NUMBER
MWI-MEIRU-SPOUSELINKS-2002-v01

TITLE
Cleaned spouse and marriage data

COUNTRY

Name	Country code
Malawi	MWI

STUDY TYPE
Sample Survey

ABSTRACT
The do-file marital_spouselinks.do combines all data on people's marital statuses and reported spouses to create the following datasets:
1. all_marital_reports - a listing of all the times an individual has reported their current marital status with the id numbers of the reported spouse(s); this listing is as reported so may include discrepancies (i.e. a 'Never married' status following a 'Married' one)
2. all_spouse_pairs_full - a listing of each time each spouse pair has been reported plus summary information on co-residency for each pair
3. all_spouse_pairs_clean_summarised - this summarises the data from all_spouse_pairs_full to give start and end dates of unions
4. marital_status_episodes - this combines data from all the sources to create episodes of marital status, each has a start and end date and a marital status, and if currently married, the spouse ids of the current spouse(s) if reported. There are several variables to indicate where each piece of information is coming from.

The first 2 datasets are made available in case people need the 'raw' data for any reason (i.e. if they only want data from one study) or if they wish to summarise the data in a different way to what is done for the last 2 datasets.

The do-file is quite complicated with many sources of data going through multiple processes to create variables in the datasets so it is not always straightforward to explain where each variable come from on the documentation. The 4 datasets build on each other and the do-file is documented throughout so anyone wanting to understand in great detail may be better off examining that. However, below is a brief description of how the datasets are created:

Marital status data are stored in the tables of the study they were collected in:
AHS Adult Health Study [ahs_ahs1]
CEN Census (initial CRS census) [cen_individ]
CENM In-migration (CRS migration form) [crs_cenm]
GP General form (filled for various reasons) [gp_gpform]
SEI Socio-economic individual (annual survey from 2007 onwards) [css_sei]
TBH TB household (study of household contacts of TB patients) [tb_tbh]
TBO TB controls (matched controls for TB patients) [tb_tbo & tb_tboto2007]
TBX TB cases (TB patients) [tb_tbx & tb_tbxto2007]
In many of the above surveys as well as their current marital status, people were asked to report their current and past spouses along with (sometimes) some information about the marriage (start/end year etc.). These data are stored all together on the table gen_spouse, with variables indicating which study the data came from.
Further evidence of spousal relationships is taken from gen_identity (if a couple appear as co-parents to a CRS member) and from crs_residency_episodes_clean_poly, a combined dataset (if they are living in the same household at the same time). Note that co-parent couples who are not reported in gen_spouse are only retained in the datasets if they have co-resident episodes.

The marital status data are appended together and the spouse id data merged in. Minimal data editing/cleaning is carried out. As the spouse data are in long format, this dataset is reshaped wide to have one line per marital status report (polygamy in the area allows for men to have multiple spouses at one time): this dataset is saved as all_marital_reports.

The list of reported spouses on gen_spouse is appended to a list of co-parents (from gen_identity) and this list is cleaned to

try to identify and remove obvious id errors (incestuous links, same sex [these are not reported in this culture] and large age difference). Data reported by men and women are compared and variables created to show whether one or both of the couple report the union.
Many records have information on start and end year of marriage, and all have the date the union was reported. This listing is compared to data from residency episodes to add dates that couples were living together (not all have start/end dates so this is to try to supplement this), in addition the dates that each member of the couple was last known to be alive or first known to be dead are added (from the residency data as well). This dataset with all the records available for each spouse pair is saved as all_spouse_pairs_full.

The date data from all_spouse_pairs_full are then summarised to get one line per couple with earliest and latest known married date for all, and, if available, marriage and separation date. For each date there are also variables created to indicate the source of the data.
As culture only allows for women having one spouse at a time, records for women with 'overlapping' husbands are cleaned. This dataset is then saved as all_spouse_pairs_clean_summarised.

Both the cleaned spouse pairs and the cleaned marital status datasets are converted into episodes: the spouse listing using the marriage or first known married date as the beginning and the last known married plus a year or separation date as the end, the marital status data records collapsed into periods of the same status being reported (following some cleaning to remove impossible reports) and the start date being the first of these reports, the end date being the last of the reports plus a year. These episodes are appended together and a series of processes run several times to remove overalapping episodes. To be able to assign specific spouse ids to each married episode, some episodes need to be 'split' into more than one (i.e. if a man is married to one woman from 2005 to 2017 and then marries another woman in 2008 and remains married to her till 2017 his intial married episode would be from 2005 to 2017, but this would need to be split into one from 2005 to 2008 which would just have 1 idspouse attached and another from 2008 to 2017, which would have 2 idspouse attached). After this splitting process the spouse ids are merged in.
The final episode dataset is saved as marital_status_episodes.

UNIT OF ANALYSIS
Individual

Version

VERSION DESCRIPTION
v1: Edited data, first version, for internal use only

VERSION DATE
2020-07-23

Producers and sponsors

PRIMARY INVESTIGATORS

Name	Affiliation
Professor Amelia (Mia) Crampin	MEIRU, London School of Hygiene and Tropical Medicine (LSHTM)

Data collection

DATA COLLECTION MODE
Face-to-face [f2f]

DATA COLLECTORS

Name	Abbreviation
Malawi Epidemiology and Intervention Research Unit	MEIRU

Access policy

ACCESS CONDITIONS
This data is made available for licensed access under the following conditions:

1. Data and other material provided by MEIRU will not be redistributed or sold to other individuals, institutions or organisations without MEIRU's written agreement.

2. In the case of multi-centre datasets, data originating from a single contributing member centre of the collaboration may not be analysed or reported on in isolation without the express permission of the member centre concerned.

3. No attempt will be made to re-identify respondents, and there will be no use of the identity of any person or establishment discovered inadvertently. Any such discovery will be reported immediately to MEIRU.

4. No attempt will be made to produce links between datasets provided by MEIRU or between MEIRU data and other datasets that could identify individuals.

5. Any books, articles, conference papers, theses, dissertations, reports or other publications employing data obtained from MEIRU will cite the source, in line with the citation requirement provided with the dataset.

6. An electronic copy of all publications based on the requested data will be sent to MEIRU.

7. MEIRU, MEIRU research collaborators and the relevant funding agencies bear no responsibility for the data's use or interpretation or inferences based upon it.

ACCESS AUTHORITY

Name
Malawi Epidemiology and Intervention Research Unit

ARCHIVE WHERE STUDY IS ORIGINALLY STORED
MEIRU

Metadata production

DDI DOCUMENT ID
DDI-MWI-MEIRU-SPOUSELINKS-2002-v01

PRODUCERS

Name	Abbreviation	Affiliation	Role
Malawi Epidemiology and Intervention Research Unit	MEIRU		Agency
Estelle McLean	EM	LSHTM and MEIRU	Production of the documentation used to create this DDI document.
Chifundo Kanjala	CK	MEIRU	Data documentalist
Jacky Saul	JS	LSHTM and MEIRU	Production of Data dictionaries
Kieth Branson	KB	LSHTM and MEIRU	Production of Data dictionaries
Joseph Kamanga	JK	MEIRU	Metadata entry
Dominic Nzundah	DN	MEIRU	Metadata entry and editing

DATE OF METADATA PRODUCTION
2020-07-23

DDI DOCUMENT VERSION

version 1 (July, 2020)

Data Dictionary

Data file	Cases	Variables
all_marital_reports This is a listing of all the times an individual has reported their current marital status with the id numbers of the reported spouse(s); this listing is as reported so may include discrepancies (i.e. a 'Never married' status following a 'Married' one) In the variable labels (S) indicates that the variable is unchanged from the original source and (D) indicates that it is a 'derived' variable, often using more than one source variable. For both types information on where the data come from and how they are processed is found in the 'Recoding and Derivation' section	263339	16
all_spouse_pairs_clean_summarised This summarises the data from all_spouse_pairs_full to give start and end dates of unions In the variable labels (S) indicates that the variable is unchanged from the original source and (D) indicates that it is a 'derived' variable, often using more than one source variable. For both types information on where the data come from and how they are processed is found in the 'Recoding and Derivation' section	22800	32
all_spouse_pairs_full This a listing of each time each spouse pair has been reported plus summary information on co-residency for each pair In the variable labels (S) indicates that the variable is unchanged from the original source and (D) indicates that it is a 'derived' variable, often using more than one source variable. For both types information on where the data come from and how they are processed is found in the 'Recoding and Derivation' section	196403	33
marital_status_episodes This combines data from all the sources to create episodes of marital status, each has a start and end date and a marital status, and if currently married, the spouse ids of the current spouse(s) if reported. There are several variables to indicate where each piece of information is coming from. In the variable labels (S) indicates that the variable is unchanged from the original source and (D) indicates that it is a 'derived' variable, often using more than one source variable. For both types information on where the data come from and how they are processed is found in the 'Recoding and Derivation' section	134177	32

Data file: all_marital_reports

This is a listing of all the times an individual has reported their current marital status with the id numbers of the reported spouse(s); this listing is as reported so may include discrepancies (i.e. a 'Never married' status following a 'Married' one)
In the variable labels (S) indicates that the variable is unchanged from the original source and (D) indicates that it is a 'derived' variable, often using more than one source variable. For both types information on where the data come from and how they are processed is found in the 'Recoding and Derivation' section

Cases:	263339
Variables:	16

Variables

ID	Name	Label	Question
V1	ident	Unique identifier	
V2	intdate	Date of interview	
V3	ques	Source of data (D)	
V4	poprep	From population representative survey From population representative survey	
V5	slf	Self or proxy report (D)	
V6	marital	Marital status (D)	
V7	spousenum	Number of current spouses (D)	
V8	idspouse1	Spouse id (1)	
V9	idspouse2	Spouse id (2)	
V10	idspouse3	Spouse id (3)	
V11	idspouse4	Spouse id (4)	
V12	idspouse5	Spouse id (5)	
V13	idspouse6	Spouse id (6)	
V14	idspouse7	Spouse id (7)	
V15	idspouse8	Spouse id (8)	
V16	idspouse9	Spouse id (9)	

Total: 16

Data file: all_spouse_pairs_clean_summarised

This summarises the data from all_spouse_pairs_full to give start and end dates of unions
In the variable labels (S) indicates that the variable is unchanged from the original source and (D) indicates that it is a 'derived' variable, often using more than one source variable. For both types information on where the data come from and how they are processed is found in the 'Recoding and Derivation' section

Cases:	22800
Variables:	32

Variables

ID	Name	Label	Question
V17	wife	Unique identifier of wife (D)	
V18	husband	Unique identifier of husband (D)	
V19	repby	Union ever reported by (D)	
V20	agediff	Age difference between spouses (D)	
V21	newparlink	Partnership only from co-parent link (D)	
V22	heverpolyn	Max number of spouses husband reports while in this union (D)	
V23	everpoly	Union was ever polygamous (D)	
V24	evercur	Union ever reported to be current (D)	
V25	mardate	Marriage date (15-April of reported year) (D)	
V26	datatypeM	mardate datatype (D)	
V27	mardtest	if mardate is estimated (D)	
V28	informantM	mardate informant (D)	
V29	fstknwmardate	First known married date (D)	
V30	datatypeFM	fstknwmardate datatype (D)	
V31	fstknwmardtest	if fstknwmardate is estimated (D)	
V32	informantFM	fstknwmardate informant (D)	
V33	stillmardate	Last date still reported to be married (D)	
V34	datatypeSM	stillmardate datatype (D)	
V35	stillmardtest	if stillmardate is estimated (D)	
V36	informantSM	stillmardate informant (D)	
V37	sepdate	Separation date (15-Aug of reported year) (D)	
V38	datatypeS	sepdate datatype (D)	
V39	sepdtest	if sepdate is estimated (D)	
V40	informantS	sepdate informant (D)	
V41	died	Which partner died (D)	
V42	firstrepdate	First date of any report (D)	
V43	lastrepdate	Last date of any report (D)	
V44	firstrepdatecrs	First date of any report while in the CRS (D)	
V45	lastrepdatecrs	Last date of any report while in the CRS (D)	
V46	evercrs	Either partner ever a CRS member (D)	
V47	fstmarrepdate	First date reported married (D)	
V48	lstmarrepdate	Last date reported married (D)	

Total: 32

Data file: all_spouse_pairs_full

This a listing of each time each spouse pair has been reported plus summary information on co-residency for each pair
In the variable labels (S) indicates that the variable is unchanged from the original source and (D) indicates that it is a 'derived' variable, often using more than one source variable. For both types information on where the data come from and how they are processed is found in the 'Recoding and Derivation' section

Cases:	196403
Variables:	33

Variables

ID	Name	Label	Question
V49	wife	Unique identifier of wife (D)	
V50	husband	Unique identifier of husband (D)	
V51	datatype	Type of data (D)	
V52	sumdate	Summary date for record (D)	
V53	repby	Union ever reported by (D)	
V54	confirmed	Both partners report the union at least once (D)	
V55	agediff	Age difference between spouses (D)	
V56	samesex	Union has same sex error (D)	
V57	everU12	Union has under 12 years error (D)	
V58	incest	Union has incest error (D)	
V59	crs	Partners ever CRS members (D)	
V60	newparlink	Partnership only from co-parent link (D)	
V61	ques	Source of marstatus record (D)	
V62	poprep	Source is population representative (D)	
V63	date	Interview date of marstatus record (D)	
V64	informant	Informant of marstatus record (D)	
V65	match	Whether partner reports union in same round (D)	
V66	heverpolyn	Max number of spouses husband reports while in this union (D)	
V67	everpoly	Union was ever polygamous (D)	
V68	evercur	Union ever reported to be current (D)	
V69	mardate	Marriage date (15-April of reported year) (D)	
V70	stillmardate	Last date still reported to be married (D)	
V71	sepdate	Separation date (15-Aug of reported year) (D)	
V72	who	Partner who moved in/out (D)	
V73	type	Type of residency data (D)	
V74	rmardate	Date known to be married with residency data (D)	
V75	rsepdate	Date known to be separated with residency data (D)	
V76	w_rsn	Wife's reason for res ep start/stop (D)	
V77	h_rsn	Husband's reason for res ep start/stop (D)	
V78	w_lastalive	Last known date wife alive (D)	
V79	w_firstdead	First known date wife dead (D)	
V80	h_lastalive	Last known date husband alive (D)	
V81	h_firstdead	First known date husband dead (D)	

Total: 33

Data file: marital_status_episodes

This combines data from all the sources to create episodes of marital status, each has a start and end date and a marital status, and if currently married, the spouse ids of the current spouse(s) if reported. There are several variables to indicate where each piece of information is coming from.
In the variable labels (S) indicates that the variable is unchanged from the original source and (D) indicates that it is a 'derived' variable, often using more than one source variable. For both types information on where the data come from and how they are processed is found in the 'Recoding and Derivation' section

Cases:	134177
Variables:	32

Variables

ID	Name	Label	Question
V82	ident	Unique identifier	
V83	sex	Sex	
V84	date1	Start date of episode (D)	
V85	dtest1	if date1 is estimated (D)	
V86	datatype1	data type of date1 (D)	
V87	ques1	source if marital status report1 (D)	
V88	informant1	informant of date1 (D)	
V89	rsn1	more detailed description of date1 (D)	
V90	marital	marital status for episode (D)	
V91	spousenum	number of spouses for episode (D)	
V92	date2	End date of episode (D)	
V93	dtest2	if date2 is estimated (D)	
V94	datatype2	data type of date2 (D)	
V95	ques2	source if marital status report2 (D)	
V96	informant2	informant of date2 (D)	
V97	rsn2	more detailed description of date2 (D)	
V98	idspouse1	spouses's id [1] (D)	
V99	agediff1	age difference between individual and their spouse [1] (D)	
V100	repby1	who reported the union [1] (D)	
V101	parlink1	whether union only known about through co-parent link and co-residency [1] (D)	
V102	idspouse2	spouses's id [2] (D)	
V103	agediff2	age difference between individual and their spouse [2] (D)	
V104	repby2	who reported the union [2] (D)	
V105	parlink2	whether union only known about through co-parent link and co-residency [2] (D)	
V106	idspouse3	spouses's id [3] (D)	
V107	agediff3	age difference between individual and their spouse [3] (D)	
V108	repby3	who reported the union [3] (D)	
V109	parlink3	whether union only known about through co-parent link and co-residency [3] (D)	
V110	idspouse4	spouses's id [4] (D)	
V111	agediff4	age difference between individual and their spouse [4] (D)	
V112	repby4	who reported the union [4] (D)	
V113	parlink4	whether union only known about through co-parent link and co-residency [4] (D)	

Total: 32

IDENT: Unique identifier

Data file: all_marital_reports

Overview

Valid: 263339 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Description

DEFINITION
Unique identifier

Imputation and derivation

DERIVATION
variable: ident (table[s]: ahs_ahs1, cen_individ, crs_cenm, gp_gpform, css_sei, tb_tbh, tb_tbo, tb_tboto2007, tb_tbx, tb_tbxto2007)

IDSPOUSE3: Spouse id (3)

Data file: all_marital_reports

Overview

Valid: 1827 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Description

DEFINITION
Spouse id

Imputation and derivation

DERIVATION
variable: idspouse (table[s]: gen_spouse)

This is mostly the spouse id variable as recorded reshaped wide by interview (so it repeats, idspouse1, idspouse2 etc.), but the code 'noid' is added when a spouse is reported but id not recorded.
replace idspouse="noid" if _m~=1 & idspouse==""
replace idspouse="" if idspouse~="noid" | marital~=1
bys ident intdate idspouse: gen n=_n
tostring n, replace
replace idspouse=idspouse+n if idspouse=="noid"
bys ident intdate (idspouse): gen n=_n
reshape wide idspouse, i(ident intdate) j(n)

INTDATE: Date of interview

Data file: all_marital_reports

Overview

Valid: 263339
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Date of interview

Imputation and derivation

DERIVATION
variable: intdate (table[s]: ahs_ahs1, cen_individ, crs_cenm, gp_gpform, css_sei, tb_tbh, tb_tbo, tb_tboto2007, tb_tbx, tb_tbxto2007)

QUES: Source of data (D)

Data file: all_marital_reports

Overview

Valid: 263339 Invalid: 0
Type: Discrete Width: 3 Range: - Format: character

Description

DEFINITION
Source of data

Imputation and derivation

DERIVATION
Created when each data source is appended together, e.g.:
use ident intdate marital using ${newstata}\ahs_ahs1, clear
gen ques="AHS"
gen slf=1
append using ${newstata}\cen_individ, keep(ident intdate marital spoucur spoupast self informtid)
replace ques="CEN" if ques==""
etc.

POPREP: From population representative survey From population representative survey

Data file: all_marital_reports

Overview

Valid: 263339 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Description

DEFINITION
From population representative survey

Imputation and derivation

DERIVATION
variable: ques (table[s]: this dataset)

gen poprep=ques=="AHS" | ques=="CEN" | ques=="CNM" | ques=="SEI"

SLF: Self or proxy report (D)

Data file: all_marital_reports

Overview

Valid: 263195 Invalid: 144
Type: Discrete Decimal: 0 Width: 2 Range: -1 - 1 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
-1	Unknown	1165	0.4%
0	Proxy-report	210895	80.1%
1	Self-report	51135	19.4%
Sysmiss		144	

Description

DEFINITION
Self or proxy report

Imputation and derivation

DERIVATION
variable: self (table[s]: cen_individ, crs_cenm, gp_gpform, tb_tbh, tb_tbo, tb_tboto2007, tb_tbx, tb_tbxto2007)
variable: infmtype (table[s]: crs_cenm)
variable: infmtype2 (table[s]: crs_cenm)
variable: relinform (table[s]: tb_tbx)
variable: informtid (table[s]: gp_gpform)

<for AHS (no proxies):>
gen slf=1
<for CEN:>
recode slf .=-1
replace slf=0 if self==0 | informtid~=""

replace slf=1 if self==1
<for CENM:>
recode slf .=-1
replace slf=0 if self==0 | (infmtype~=. & infmtype~=6) //6=self
replace slf=1 if self==1 | infmtype==6
<for GP:>
recode slf .=-1
replace slf=0 if self==0 | informtid~=""
replace slf=1 if self==1
<for SEI:>
recode slf .=-1
replace slf=0 if infmtype2~=1 & infmtype2~=99 & infmtype~=6
replace slf=1 if infmtype2==1 | infmtype==6
<for TBH:>
recode slf .=-1
replace slf=0 if self==0
replace slf=1 if self==1
<for TBO:>
recode slf .=-1
replace slf=0 if self==0
replace slf=1 if self==1
<for TBX:>
recode slf .=-1
replace slf=0 if (relinform~=8 & relinform~=.) | informtid~=""
replace slf=1 if relinform==8

MARITAL: Marital status (D)

Data file: all_marital_reports

Overview

Valid: 259912 Invalid: 3427
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 9 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	Never married	84882	32.7%
1	Married	141125	54.3%
2	Divorced/separated	18136	7%
3	Widowed	15769	6.1%
4	N/A (dead)	0	0%
8	Unknown	0	0%
9	Missing	3414	
Sysmiss		13	

Description

DEFINITION
Marital status

Imputation and derivation

DERIVATION
variable: marital (table[s]: ahs_ahs1, cen_individ, crs_cenm, gp_gpform, css_sei, tb_tbh, tb_tbo, tb_tboto2007, tb_tbx, tb_tbxto2007)
variable: spoucur (table[s]: cen_individ, crs_cenm)
variable: spoupast (table[s]: cen_individ, crs_cenm)
variable: maridill (table[s]: tb_tbxto2007)
variable: idspouse (table[s]: gen_spouse)
variable: marendyr (table[s]: gen_spouse)

Mostly the variables is used as reported but for a few sources it is updated using other data:
la def marital 4 "N/A (dead)", modify
replace marital=0 if spoucur==0 & spoupast<=0 & marital==.
replace marital=1 if spoucur>=1 & spoucur~=. & marital==.
replace marital=2 if spoucur==0 & spoupast>=1 & spoupast~=. & marital==.
replace marital=0 if maridill==0 /*just assumed that not married is never married for these ones for now*/
replace marital=1 if maridill==1
<and after merging in reported spouse ids from gen_spouse:>
replace marital=1 if idspouse~="" & marendyr==.

SPOUSENUM: Number of current spouses (D)

Data file: all_marital_reports

Overview

Valid: 112388 Invalid: 150951
Type: Discrete Decimal: 0 Width: 2 Range: 0 - 99 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category
97	Not applicable
98	Unknown
99	Missing

Description

DEFINITION
Number of current spouses

Imputation and derivation

DERIVATION
variable: spousenum (table[s]: css_sei)
variable: spoucur (table[s]: cen_individ, crs_cenm)

replace spousenum=spoucur if spousenum==.

IDSPOUSE1: Spouse id (1)

Data file: all_marital_reports

Overview

Valid: 138401 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Description

DEFINITION
Spouse id

Imputation and derivation

DERIVATION
variable: idspouse (table[s]: gen_spouse)

This is mostly the spouse id variable as recorded reshaped wide by interview (so it repeats, idspouse1, idspouse2 etc.), but the code 'noid' is added when a spouse is reported but id not recorded.
replace idspouse="noid" if _m~=1 & idspouse==""
replace idspouse="" if idspouse~="noid" | marital~=1
bys ident intdate idspouse: gen n=_n
tostring n, replace
replace idspouse=idspouse+n if idspouse=="noid"
bys ident intdate (idspouse): gen n=_n
reshape wide idspouse, i(ident intdate) j(n)

IDSPOUSE2: Spouse id (2)

Data file: all_marital_reports

Overview

Valid: 11388 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Description

DEFINITION
Spouse id

Imputation and derivation

DERIVATION
variable: idspouse (table[s]: gen_spouse)

This is mostly the spouse id variable as recorded reshaped wide by interview (so it repeats, idspouse1, idspouse2 etc.), but the code 'noid' is added when a spouse is reported but id not recorded.
replace idspouse="noid" if _m~=1 & idspouse==""
replace idspouse="" if idspouse~="noid" | marital~=1

bys ident intdate idspouse: gen n=_n
tostring n, replace
replace idspouse=idspouse+n if idspouse=="noid"
bys ident intdate (idspouse): gen n=_n
reshape wide idspouse, i(ident intdate) j(n)

IDSPOUSE4: Spouse id (4)

Data file: all_marital_reports

Overview

Valid: 382 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Description

DEFINITION
Spouse id

Imputation and derivation

DERIVATION
variable: idspouse (table[s]: gen_spouse)

This is mostly the spouse id variable as recorded reshaped wide by interview (so it repeats, idspouse1, idspouse2 etc.), but the code 'noid' is added when a spouse is reported but id not recorded.
replace idspouse="noid" if _m~=1 & idspouse==""
replace idspouse="" if idspouse~="noid" | marital~=1
bys ident intdate idspouse: gen n=_n
tostring n, replace
replace idspouse=idspouse+n if idspouse=="noid"
bys ident intdate (idspouse): gen n=_n
reshape wide idspouse, i(ident intdate) j(n)

IDSPOUSE5: Spouse id (5)

Data file: all_marital_reports

Overview

Valid: 26 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Description

DEFINITION
Spouse id

Imputation and derivation

DERIVATION

variable: idspouse (table[s]: gen_spouse)

This is mostly the spouse id variable as recorded reshaped wide by interview (so it repeats, idspouse1, idspouse2 etc.), but the code 'noid' is added when a spouse is reported but id not recorded.
replace idspouse="noid" if _m~=1 & idspouse==""
replace idspouse="" if idspouse~="noid" | marital~=1
bys ident intdate idspouse: gen n=_n
tostring n, replace
replace idspouse=idspouse+n if idspouse=="noid"
bys ident intdate (idspouse): gen n=_n
reshape wide idspouse, i(ident intdate) j(n)

IDSPOUSE6: Spouse id (6)

Data file: all_marital_reports

Overview

Valid: 8 Invalid: 0
Type: Discrete Width: 5 Range: - Format: character

Description

DEFINITION
Spouse id

Imputation and derivation

DERIVATION
variable: idspouse (table[s]: gen_spouse)

This is mostly the spouse id variable as recorded reshaped wide by interview (so it repeats, idspouse1, idspouse2 etc.), but the code 'noid' is added when a spouse is reported but id not recorded.
replace idspouse="noid" if _m~=1 & idspouse==""
replace idspouse="" if idspouse~="noid" | marital~=1
bys ident intdate idspouse: gen n=_n
tostring n, replace
replace idspouse=idspouse+n if idspouse=="noid"
bys ident intdate (idspouse): gen n=_n
reshape wide idspouse, i(ident intdate) j(n)

IDSPOUSE7: Spouse id (7)

Data file: all_marital_reports

Overview

Valid: 6 Invalid: 0
Type: Discrete Width: 5 Range: - Format: character

Description

DEFINITION
Spouse id

Imputation and derivation

DERIVATION
variable: idspouse (table[s]: gen_spouse)

This is mostly the spouse id variable as recorded reshaped wide by interview (so it repeats, idspouse1, idspouse2 etc.), but the code 'noid' is added when a spouse is reported but id not recorded.
replace idspouse="noid" if _m~=1 & idspouse==""
replace idspouse="" if idspouse~="noid" | marital~=1
bys ident intdate idspouse: gen n=_n
tostring n, replace
replace idspouse=idspouse+n if idspouse=="noid"
bys ident intdate (idspouse): gen n=_n
reshape wide idspouse, i(ident intdate) j(n)

IDSPOUSE8: Spouse id (8)

Data file: all_marital_reports

Overview

Valid: 2 Invalid: 0
Type: Discrete Width: 5 Range: - Format: character

Description

DEFINITION
Spouse id

Imputation and derivation

DERIVATION
variable: idspouse (table[s]: gen_spouse)

This is mostly the spouse id variable as recorded reshaped wide by interview (so it repeats, idspouse1, idspouse2 etc.), but the code 'noid' is added when a spouse is reported but id not recorded.
replace idspouse="noid" if _m~=1 & idspouse==""
replace idspouse="" if idspouse~="noid" | marital~=1
bys ident intdate idspouse: gen n=_n
tostring n, replace
replace idspouse=idspouse+n if idspouse=="noid"
bys ident intdate (idspouse): gen n=_n
reshape wide idspouse, i(ident intdate) j(n)

IDSPOUSE9: Spouse id (9)

Data file: all_marital_reports

Overview

Valid: 1 Invalid: 0
Type: Discrete Width: 5 Range: - Format: character

Description

DEFINITION
Spouse id

Imputation and derivation

DERIVATION
variable: idspouse (table[s]: gen_spouse)

This is mostly the spouse id variable as recorded reshaped wide by interview (so it repeats, idspouse1, idspouse2 etc.), but the code 'noid' is added when a spouse is reported but id not recorded.
replace idspouse="noid" if _m~=1 & idspouse==""
replace idspouse="" if idspouse~="noid" | marital~=1
bys ident intdate idspouse: gen n=_n
tostring n, replace
replace idspouse=idspouse+n if idspouse=="noid"
bys ident intdate (idspouse): gen n=_n
reshape wide idspouse, i(ident intdate) j(n)

WIFE: Unique identifier of wife (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 22800 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Description

DEFINITION
Unique identifier of wife

Imputation and derivation

DERIVATION
variable: wife (table[s]: all_spouse_pairs_full [derived table])

HUSBAND: Unique identifier of husband (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 22800 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Description

DEFINITION
Unique identifier of husband

Imputation and derivation

DERIVATION
variable: husband (table[s]: all_spouse_pairs_full [derived table])

REPBY: Union ever reported by (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 21253 Invalid: 1547
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 4 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	3992	18.8%
2	husband	1686	7.9%
3	both	15575	73.3%
4	Residency data	0	0%
Sysmiss		1547	

Description

DEFINITION
Union ever reported by

Imputation and derivation

DERIVATION
variable: repby (table[s]: all_spouse_pairs_full [derived table])

AGEDIFF: Age difference between spouses (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 22800 Invalid: 0
Type: Continuous Decimal: 0 Width: 2 Range: 0 - 44 Format: Numeric

Description

DEFINITION
Age difference between spouses

Imputation and derivation

DERIVATION
variable: agediff (table[s]: all_spouse_pairs_full [derived table])

NEWPARLINK: Partnership only from co-parent link (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 22800 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Description

DEFINITION
Partnership only from co-parent link

Imputation and derivation

DERIVATION
variable: newparlink (table[s]: all_spouse_pairs_full [derived table])

HEVERPOLYN: Max number of spouses husband reports while in this union (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 13219 Invalid: 9581
Type: Continuous Decimal: 0 Width: 2 Range: 0 - 99 Format: Numeric

Description

DEFINITION
Max number of spouses husband reports while in this union

Imputation and derivation

DERIVATION
variable: heverpolyn (table[s]: all_spouse_pairs_full [derived table])

EVERPOLY: Union was ever polygamous (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 13219 Invalid: 9581
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Description

DEFINITION
Union was ever polygamous (D)

Imputation and derivation

DERIVATION
variable: everpoly (table[s]: all_spouse_pairs_full [derived table])

EVERCUR: Union ever reported to be current (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 21253 Invalid: 1547
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Description

DEFINITION
Union ever reported to be current

Imputation and derivation

DERIVATION
variable: evercur (table[s]: all_spouse_pairs_full [derived table])

MARDATE: Marriage date (15-April of reported year) (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 14362
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Marriage date (15-April of reported year)

Imputation and derivation

DERIVATION
variable: mardate (table[s]: all_spouse_pairs_full [derived table])
variable: rmardate (table[s]: all_spouse_pairs_full [derived table])
variable: rsepdate (table[s]: all_spouse_pairs_full [derived table])

the mode of the reported marriage dates is taken first (this is the date estimated from the marriage start year reported by either partner). It is updated to the earliest co-residency date if this is earlier, or in the same year but different date (as the reported values are all estimated as 15 April).

DATATYPEM: mardate datatype (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 14371 Invalid: 8429
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 6 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	Report	9118	63.4%
2	Residency	5253	36.6%
3	Both	0	0%
4	Death	0	0%
5	ep censor	0	0%
6	ep split	0	0%
Sysmiss		8429	

Description

DEFINITION
mardate datatype (D)

Imputation and derivation

DERIVATION
The datatype variable is updated depending on which final mardate is used.

MARDTEST: if mardate is estimated (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 14371 Invalid: 8429
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 5 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	No	5253	36.6%
1	Day est	0	0%
2	Day&month	8949	62.3%
3	est due to overlap	169	1.2%
4	ep censor	0	0%

5	ep split	0	0%
Sysmiss		8429	

Description

DEFINITION
if mardate is estimated (D)

Imputation and derivation

DERIVATION
The dtest variable is updated depending on which final mardate is used.

INFORMANTM: mardate informant (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 14371 Invalid: 8429
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 4 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	2674	18.6%
2	husband	480	3.3%
3	both	5964	41.5%
4	Residency data	5253	36.6%
Sysmiss		8429	

Description

DEFINITION
mardate informant (D)

Imputation and derivation

DERIVATION
The informant variable is updated depending on which final mardate is used.

FSTKNWMARDATE: First known married date (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 8304
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
First known married date (D)

Imputation and derivation

DERIVATION
variable: stillmardate (table[s]: all_spouse_pairs_full [derived table])
variable: rmardate (table[s]: all_spouse_pairs_full [derived table])
variable: rsepdate (table[s]: all_spouse_pairs_full [derived table])

The earliest date either couple report being married to this partner is used first (i.e. the interview date when the partnership first reported when they are married to them). It is updated to the earliest co-residency date if this is earlier. After all the dates have been finalised this date is removed if mardate is available.

DATATYPEFM: fstknwmardate datatype (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 8304 Invalid: 14496
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 6 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	Report	5916	71.2%
2	Residency	2311	27.8%
3	Both	0	0%
4	Death	77	0.9%
5	ep censor	0	0%
6	ep split	0	0%
Sysmiss		14496	

Description

DEFINITION
fstknwmardate datatype (D)

Imputation and derivation

DERIVATION
The datatype variable is updated depending on which final fstknwmardate is used.

FSTKNWMARDTEST: if fstknwmardate is estimated (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 8268 Invalid: 14532
Type: Discrete Decimal: 0 Width: 5 Range: 0 - 17800 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	No	8041	97.3%
1	Day est	0	0%
2	Day&month	221	2.7%
3	est due to overlap	4	0%
4	ep censor	0	0%
5	ep split	0	0%
15776		1	0%
17800		1	0%
Sysmiss		14532	

Description

DEFINITION
if fstknwmardate is estimated (D)

Imputation and derivation

DERIVATION
The dtest variable is updated depending on which final fstknwmardate is used.

INFORMANTFM: fstknwmardate informant (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 8304 Invalid: 14496
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 4 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	1547	18.6%
2	husband	1445	17.4%
3	both	2924	35.2%
4	Residency data	2388	28.8%
Sysmiss		14496	

Description

DEFINITION
fstknwmardate informant (D)

Imputation and derivation

DERIVATION
The informant variable is updated depending on which final fstknwmardate is used.

STILLMARDATE: Last date still reported to be married (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 15209
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Last date still reported to be married

Imputation and derivation

DERIVATION
variable: stillmardate (table[s]: all_spouse_pairs_full [derived table])
variable: rmardate (table[s]: all_spouse_pairs_full [derived table])
variable: rsepdate (table[s]: all_spouse_pairs_full [derived table])

The latest date either couple report being married to this partner is used first (i.e. the interview date when the partnership first reported when they are married to them). It is updated to the latest co-residency date if this is later. After all the dates have been finalised this date is removed if sepdate is available.

DATATYPESM: stillmardate datatype (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 15209 Invalid: 7591
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 6 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	Report	6146	40.4%
2	Residency	9063	59.6%
3	Both	0	0%
4	Death	0	0%
5	ep censor	0	0%
6	ep split	0	0%
Sysmiss		7591	

Description

DEFINITION
stillmardate datatype (D)

Imputation and derivation

DERIVATION
The datatype variable is updated depending on which final stillmardate is used.

STILLMARDTEST: if stillmardate is estimated (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 15178 Invalid: 7622
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 5 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	No	15113	99.6%
1	Day est	0	0%
2	Day&month	40	0.3%

3	est due to overlap	25	0.2%
4	ep censor	0	0%
5	ep split	0	0%
Sysmiss		7622	

Description

DEFINITION
if stillmardate is estimated (D)

Imputation and derivation

DERIVATION
The dtest variable is updated depending on which final stillmardate is used.

INFORMANTSM: stillmardate informant (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 15209 Invalid: 7591
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 4 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	1970	13%
2	husband	1869	12.3%
3	both	2307	15.2%
4	Residency data	9063	59.6%
Sysmiss		7591	

Description

DEFINITION
stillmardate informant (D)

Imputation and derivation

DERIVATION
The informant variable is updated depending on which final stillmardate is used.

SEPDATE: Separation date (15-Aug of reported year) (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 7457
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Separation date (15-Aug of reported year)

Imputation and derivation

DERIVATION
variable: sepdate (table[s]: all_spouse_pairs_full [derived table])
variable: rmardate (table[s]: all_spouse_pairs_full [derived table])
variable: rsepdate (table[s]: all_spouse_pairs_full [derived table])
variable: w_firstdead (table[s]: all_spouse_pairs_full [derived table])
variable: h_firstdead (table[s]: all_spouse_pairs_full [derived table])

the mode of the reported separation dates is taken first (this is the date estimated from the marriage end year reported by either partner). It is updated to the latest co-residency date [if the reason for the end of the co-res episode was divorce] if this is later or in the same year but different date. If there is a residency date where they are still thought to be married after the separation date the sepdate is deleted. If the sepdate or stillmardate is on or after the death date for either partner, the sepdate is updated to the death date.

DATATYPES: sepdate datatype (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 7457 Invalid: 15343
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 6 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	Report	2446	32.8%
2	Residency	3486	46.7%
3	Both	0	0%
4	Death	1525	20.5%
5	ep censor	0	0%
6	ep split	0	0%
Sysmiss		15343	

Description

DEFINITION
sepdate datatype (D)

Imputation and derivation

DERIVATION
The datatype variable is updated depending on which final sepdate is used.

SEPDTEST: if sepdate is estimated (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 7330 Invalid: 15470
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 5 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	No	4509	61.5%
1	Day est	0	0%
2	Day&month	2731	37.3%
3	est due to overlap	90	1.2%
4	ep censor	0	0%
5	ep split	0	0%
Sysmiss		15470	

Description

DEFINITION
if sepdate is estimated (D)

Imputation and derivation

DERIVATION
The dtest variable is updated depending on which final sepdate is used.

INFORMANTS: sepdate informant (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 7457 Invalid: 15343
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 4 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	2171	29.1%
2	husband	256	3.4%
3	both	19	0.3%
4	Residency data	5011	67.2%
Sysmiss		15343	

Description

DEFINITION
sepdate informant (D)

Imputation and derivation

DERIVATION
The informant variable is updated depending on which final sepdate is used.

DIED: Which partner died (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 1525 Invalid: 21275
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 3 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	348	22.8%
2	husband	1177	77.2%
3	both	0	0%
Sysmiss		21275	

Description

DEFINITION
Which partner died (D)

Imputation and derivation

DERIVATION
Updated to wife or husband if their death date is taken as sepdate

FIRSTREPDATE: First date of any report (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 21253
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
First date reported married (D)

Imputation and derivation

DERIVATION
The first date either partner report this union (regardless of whether they reported still being married)

LASTREPDATE: Last date of any report (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 21253
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Last date reported married (D)

Imputation and derivation

DERIVATION
The last date either partner report this union (regardless of whether they reported still being married)

FIRSTREPDATECRS: First date of any report while in the CRS (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 16581
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
First date of any report (D)

Imputation and derivation

DERIVATION
The first date either partner report this union while they were a CRS member (regardless of whether they reported still being married)

LASTREPDATECRS: Last date of any report while in the CRS (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 16581
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Last date of any report (D)

Imputation and derivation

DERIVATION
The last date either partner report this union while they were a CRS member (regardless of whether they reported still being married)

EVERCRS: Either partner ever a CRS member (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 16581 Invalid: 6219
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 1 Format: Numeric

Description

DEFINITION
First date of any report while in the CRS (D)

Imputation and derivation

DERIVATION
gen evercrs=1 if firstrepdatecrs~=.

FSTMARREPDATE: First date reported married (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 18870
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Last date of any report while in the CRS (D)

Imputation and derivation

DERIVATION
variable: stillmardate (table[s]: all_spouse_pairs_full [derived table])

<after all the summary dates are finalised the earliest date either of the partners reported being married to the other partner that is on or after mardate or fstknwmardate is taken:>
egen start=rowmin(mardate fstknwmardate)
bys wife husband: egen temp1=min(stillmardateO) if stillmardateO~=. & stillmardateO>=start
bys wife husband: egen fstmarrepdate=max(temp1)

LSTMARREPDATE: Last date reported married (D)

Data file: all_spouse_pairs_clean_summarised

Overview

Valid: 18673
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Either partner ever a CRS member (D)

Imputation and derivation

DERIVATION
variable: stillmardate (table[s]: all_spouse_pairs_full [derived table])

<after all the summary dates are finalised the latest date either of the partners reported being married to the other partner that is on or before sepdate or stillmardate is taken:>
egen end=rowmax(stillmardate sepdate)
bys wife husband: egen temp2=max(stillmardateO) if stillmardateO~=. & stillmardateO<=end
bys wife husband: egen lstmarrepdate=max(temp2)

WIFE: Unique identifier of wife (D)

Data file: all_spouse_pairs_full

Overview

Valid: 196403 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Description

DEFINITION
Unique identifier of wife

Imputation and derivation

DERIVATION
variable: ident (table[s]: gen_spouse, crs_residency_episodes_clean_poly [derived table], all_marital_reports [derived table])
variable: idspouse (table[s]: gen_spouse)
variable: idmoth (table[s]: gen_identity)

HUSBAND: Unique identifier of husband (D)

Data file: all_spouse_pairs_full

Overview

Valid: 196403 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Description

DEFINITION
Unique identifier of husband

Imputation and derivation

DERIVATION
variable: ident (table[s]: gen_spouse, crs_residency_episodes_clean_poly [derived table], all_marital_reports [derived table])
variable: idspouse (table[s]: gen_spouse)
variable: idfath (table[s]: gen_identity)

DATATYPE: Type of data (D)

Data file: all_spouse_pairs_full

Overview

Valid: 196403 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 2 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	marstatus	158094	80.5%
2	residency	38309	19.5%

Description

DEFINITION
Type of data

Imputation and derivation

DERIVATION
Created when making the separate datasets and appending them together.

SUMDATE: Summary date for record (D)

Data file: all_spouse_pairs_full

Overview

Valid: 196403
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Summary date for record to make it easier to sort/understand data

Imputation and derivation

DERIVATION
variable: date (table[s]: this dataset)
variable: rmardate (table[s]: this dataset)
variable: rsepdate (table[s]: this dataset)

egen sumdate=rowmin(date rmardate rsepdate)

REPBY: Union ever reported by (D)

Data file: all_spouse_pairs_full

Overview

Valid: 193050 Invalid: 3353
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 3 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	12540	6.5%
2	husband	2958	1.5%
3	both	177552	92%
Sysmiss		3353	

Description

DEFINITION
Union ever reported by

Imputation and derivation

DERIVATION
variable: ident (table[s]: gen_spouse)
variable: idspouse (table[s]: gen_spouse)

The data from gen_spouse is split into male and female reports which are merged into each other, the resulting merge variable is used to create repby. NB. the spouse_match_ datasets referred to below are the separated datasets:

```
use ident idspouse using ${tempfiles}\spouse_match_1, clear
rename (ident idspouse) (wife husband)
duplicates drop
save ${tempfiles}\couples_Freport, replace
use ident idspouse using ${tempfiles}\spouse_match_0, clear
rename (ident idspouse) (husband wife)
duplicates drop
merge 1:1 wife husband using ${tempfiles}\couples_Freport
gen repby=3 if _m==3 /*if they match they have both reported each other at least once*/
replace repby=1 if _m==2 /*if the union is only in the using data then only the wife has ever reported it*/
replace repby=2 if _m==1 /*if the union is only in the master data then only the husband has ever reported it*/
la val repby repby
drop _m
save ${tempfiles}\spouses_ever_reported, replace
```

CONFIRMED: Both partners report the union at least once (D)

Data file: all_spouse_pairs_full

Overview

Valid: 193050 Invalid: 3353
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Description

DEFINITION
Both partners report the union at least once

Imputation and derivation

DERIVATION
variable: repby (table[s]: this dataset)

gen confirmed=0
replace confirmed=1 if repby==3

AGEDIFF: Age difference between spouses (D)

Data file: all_spouse_pairs_full

Overview

Valid: 196385 Invalid: 18
Type: Continuous Decimal: 0 Width: 2 Range: 0 - 69 Format: Numeric

Description

DEFINITION
Age difference between spouses

Imputation and derivation

DERIVATION
variable: ident (table[s]: gen_spouse)
variable: idspouse (table[s]: gen_spouse)
variable: birth_date (table[s]: gen_identity_clean [derived table])

rename (ident idspouse) (ident1 ident)
merge m:1 ident using ${newstata}\gen_identity_clean.dta, keepusing(sex birth_date idmoth idfath) keep(master match) nogen
gen idspouse_age=int((intdate-birth_date)/365.25) if birth_date~=.
gen idspouse_yob=year(birth_date) if birth_date~=.
drop birth_date
rename (ident1 ident sex idmoth idfath) (ident idspouse idspouse_sex idspouse_idmoth idspouse_idfath)
gen agediff=abs(ident_age-idspouse_age)

SAMESEX: Union has same sex error (D)

Data file: all_spouse_pairs_full

Overview

Valid: 196403 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Description

DEFINITION
Union has same sex error

Imputation and derivation

DERIVATION
variable: sex (table[s]: gen_identity)

gen samesex=0
replace samesex=1 if ident_sex==idspouse_sex

EVERU12: Union has under 12 years error (D)

Data file: all_spouse_pairs_full

Overview

Valid: 196403 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Description

DEFINITION
Union has under 12 years error

Imputation and derivation

DERIVATION
bys ident idspouse: egen youngestident=min(ident_age)
bys ident idspouse: egen youngestidspouse=min(idspouse_age)
gen everU12=0
replace everU12=1 if (youngestident<12 & youngestident~=.) | (youngestidspouse<12 & youngestidspouse~=.)

INCEST: Union has incest error (D)

Data file: all_spouse_pairs_full

Overview

Valid: 196403 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Description

DEFINITION
Union has incest error

Imputation and derivation

DERIVATION
variable: ident (table[s]: gen_spouse)
variable: idspouse (table[s]: gen_spouse)
variable: idmoth (table[s]: gen_identity)

variable: idfath (table[s]: gen_identity)

gen incest=0
replace incest=1 if (ident_idmoth==idspouse & ident_idmoth~="" & idspouse~="") | (ident_idfath==idspouse & ident_idfath~="" & idspouse~="") | (idspouse_idmoth==ident & idspouse_idmoth~="" & ident~="") | (idspouse_idfath==ident & idspouse_idfath~="" & ident~="")

CRS: Partners ever CRS members (D)

Data file: all_spouse_pairs_full

Overview

Valid: 188811 Invalid: 7592
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 3 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	7473	4%
2	husband	774	0.4%
3	both	180564	95.6%
Sysmiss		7592	

Description

DEFINITION
Partners ever CRS members

Imputation and derivation

DERIVATION
<A listing of all idents that were ever CRS members is created:>
use ${gendata}\crs_residency_episodes_clean_nopoly, clear
bys ident: egen lastincrs=max(stopdate)
bys ident (stopdate): gen laststop=stoptype[_N]
format %td lastincrs
keep ident lastincrs laststop
duplicates drop
gen crs=1
save ${tempfiles}\allcrsmembers, replace
<then this is merged into the main dataset for each partner:>
merge m:1 ident using ${tempfiles}\allcrsmembers, keep(master match) keepusing(crs) nogen
merge m:1 ident using ${tempfiles}\indices_lafd, keep(master match) nogen
rename (ident husband lastalive firstdead crs) (wife ident w_lastalive w_firstdead w_crs)
merge m:1 ident using ${tempfiles}\allcrsmembers, keep(master match) keepusing(crs) nogen
merge m:1 ident using ${tempfiles}\indices_lafd, keep(master match) nogen
rename (ident lastalive firstdead crs) (husband h_lastalive h_firstdead h_crs)
gen crs=1 if w_crs==1 & h_crs~=1
replace crs=2 if w_crs~=1 & h_crs==1
replace crs=3 if w_crs==1 & h_crs==1

la val crs repby

NEWPARLINK: Partnership only from co-parent link (D)

Data file: all_spouse_pairs_full

Overview

Valid: 196403 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Description

DEFINITION
Partnership only from co-parent link

Imputation and derivation

DERIVATION
<The variable parlink is created when co-parent data is processed before being appended to other spouse data:>
use ${newstata}\gen_identity_clean, clear
replace idmoth="" if idmoth=="UUUUUUU"
replace idfath="" if idfath=="UUUUUUU"
keep if idmoth~="" & idfath~=""
rename (idmoth idfath) (ident idspouse)
gen parlink=1
<After all data are combined parlink is used to id couples who remain in the dataset as they have lived together but have never reported each other as spouses:>
recode parlink .=0
bys wife husband: egen newparlink=min(parlink)

QUES: Source of marstatus record (D)

Data file: all_spouse_pairs_full

Overview

Valid: 158094 Invalid: 0
Type: Discrete Width: 3 Range: - Format: character

Description

DEFINITION
Source of marstatus record

Imputation and derivation

DERIVATION
variable: ques (table[s]: all_marital_reports [derived table])

POPREP: Source is population representative (D)

Data file: all_spouse_pairs_full

Overview

Valid: 193050 Invalid: 3353
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Description

DEFINITION
Source is population representative

Imputation and derivation

DERIVATION
variable: poprep (table[s]: all_marital_reports [derived table])

DATE: Interview date of marstatus record (D)

Data file: all_spouse_pairs_full

Overview

Valid: 158094
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Interview date of marstatus record

Imputation and derivation

DERIVATION
variable: date (table[s]: all_marital_reports [derived table])

INFORMANT: Informant of marstatus record (D)

Data file: all_spouse_pairs_full

Overview

Valid: 158094 Invalid: 38309
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 3 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	85559	54.1%
2	husband	72535	45.9%
3	both	0	0%
Sysmiss		38309	

Description

DEFINITION
Informant of marstatus record

Imputation and derivation

DERIVATION
variable: informant (table[s]: all_marital_reports [derived table])

MATCH: Whether partner reports union in same round (D)

Data file: all_spouse_pairs_full

Overview

Valid: 158094 Invalid: 38309
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 2 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	Matched	128948	81.6%
1	Missing data	25335	16%
2	Mismatch	3811	2.4%
Sysmiss		38309	

Description

DEFINITION
Whether partner reports union in same round

Imputation and derivation

DERIVATION
variable: ident (table[s]: gen_spouse)
variable: idspouse (table[s]: gen_spouse)
variable: marital (table[s]: ahs_ahs1, cen_individ, crs_cenm, gp_gpform, css_sei, tb_tbh, tb_tbo, tb_tboto2007, tb_tbx, tb_tbxto2007)

variable: spoucur (table[s]: cen_individ, crs_cenm)
variable: spoupast (table[s]: cen_individ, crs_cenm)
variable: maridill (table[s]: tb_tbxto2007)
variable: idspouse (table[s]: gen_spouse)
variable: marendyr (table[s]: gen_spouse)

<Once the marital status report and spouse listing datasets are merged together they are split into what men and women are reporting per round/survey and merged together to assess whether each report is reported by the otehr partner:>

```
foreach spouse in wife husband {
preserve
if "`spouse'"=="wife" {
local sexi=1
local sexs=0
local S="H"
}
if "`spouse'"=="husband" {
local sexi=0
local sexs=1
local S="W"
}
*create dataset of what the spouse 2 (idspouse) reports
drop if ident_sex==`sexi'
rename (idspouse ident marital) (x`spouse' idspouse xmarital)
keep idspouse x`spouse' rd xmarital
bys idspouse rd: gen n=_n
sum n
local max=r(max)
reshape wide xmarital x`spouse' ,i(idspouse rd) j(n)
save ${tempfiles}\gen_spouse_`sexs', replace
restore
preserve
*create dataset of what spouse 1 (ident) reports and merge with the spouse 2 dataset
drop if ident_sex==`sexs'
drop if idspouse=="" | idspouse=="noid"
merge m:1 idspouse rd using ${tempfiles}\gen_spouse_`sexs', keep(master match) nogen
*create match var
gen match=.
forvalues x=1/`max' {
replace match=0 if ident==x`spouse'`x'
}
replace match=1 if match==. &(xmarital1==. | (xmarital1~=0 & x`spouse'1=="))
replace match=2 if match==. & x`spouse'1~="" | xmarital1==0
lab def match 0 "Matched" 1 "Missing data" 2 "Mismatch"
la val match match
drop x*
save ${tempfiles}\spouse_match_`sexi', replace
restore
}
```

HEVERPOLYN: Max number of spouses husband reports while in this union (D)

Data file: all_spouse_pairs_full

Overview

Valid: 164160 Invalid: 32243
Type: Continuous Decimal: 0 Width: 2 Range: 0 - 99 Format: Numeric

Description

DEFINITION
Max number of spouses husband reports while in this union

Imputation and derivation

DERIVATION
variable: informant (table[s]: all_marital_reports [derived table])
variable: spousenum (table[s]: all_marital_reports [derived table])

replace spousenum=. if informant==1 //only want husband number of spouses to know if union is polygamous
bys wife husband: egen heverpolyn=max(spousenum)

EVERPOLY: Union was ever polygamous (D)

Data file: all_spouse_pairs_full

Overview

Valid: 164160 Invalid: 32243
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Description

DEFINITION
Union was ever polygamous (D)

Imputation and derivation

DERIVATION
variable: heverpolyn (table[s]: this dataset)

recode heverpolyn 1=0 2/max=1, gen(everpoly)

EVERCUR: Union ever reported to be current (D)

Data file: all_spouse_pairs_full

Overview

Valid: 193050 Invalid: 3353
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Description

DEFINITION
Union ever reported to be current

Imputation and derivation

DERIVATION
gen evercur=marital==1

MARDATE: Marriage date (15-April of reported year) (D)

Data file: all_spouse_pairs_full

Overview

Valid: 118254
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Marriage date (15-April of reported year)

Imputation and derivation

DERIVATION
gen mardate=mdy(4,15,marstyr) if marstyr~=. & marstyr<9998

STILLMARDATE: Last date still reported to be married (D)

Data file: all_spouse_pairs_full

Overview

Valid: 139996
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Last date still reported to be married

Imputation and derivation

DERIVATION
gen stillmardate=intdate if wife~="" & husband~="" & marendyr==.

SEPDATE: Separation date (15-Aug of reported year) (D)

Data file: all_spouse_pairs_full

Overview

Valid: 15846
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Separation date (15-Aug of reported year)

Imputation and derivation

DERIVATION
gen sepdate=mdy(8,15,marendyr) if marendyr~=. & marendyr<9998

WHO: Partner who moved in/out (D)

Data file: all_spouse_pairs_full

Overview

Valid: 38309 Invalid: 158094
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 3 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	6535	17.1%
2	husband	2127	5.6%
3	both	29647	77.4%
Sysmiss		158094	

Description

DEFINITION
Who was 'responsible' for the residency start/stop on this date: i.e. who moved in last or out first, or did they move in/out together?

Imputation and derivation

DERIVATION
<Spouse listing (from co-parent links and gen_spouse) are merged into residency episodes, those where there is overlap for the wife and husband in the same household are kept. The start & end dates of the overlapping period are compared to the start and end dates of the wife's/husband's episode to see who moved in/out:>
use ${tempfiles}\parlinks_checked, clear
*add in spouse links
append using ${tempfiles}\spouse_pairs_all_data, keep(wife husband)
duplicates drop

```
drop if wife=="" | husband==""
rename wife ident
*merge to residency data (wife)
joinby ident using ${gendata}\crs_residency_episodes_clean_poly, unmatched(both) _merge(wifem)
rename (ident startdate stopdate starttype stoptype rsnmovecenm rsnmovedep crshse) (wife w_startdate w_stopdate
w_starttype w_stoptype w_rsnstart w_rsnstop w_crshse)
drop sex birth_date
rename husband ident
joinby ident using ${gendata}\crs_residency_episodes_clean_poly, unmatched(both) _merge(husbandm)
rename (ident startdate stopdate starttype stoptype rsnmovecenm rsnmovedep crshse) (husband h_startdate h_stopdate
h_starttype h_stoptype h_rsnstart h_rsnstop h_crshse)
drop sex birth_date
keep if w_crshse==h_crshse & h_startdate~=. & h_stopdate~=. & w_startdate~=. & w_stopdate~=. &
((h_startdate>=w_startdate & h_startdate<=w_stopdate) | (h_stopdate>=w_startdate & h_stopdate<=w_stopdate) |
(w_startdate>=h_startdate & w_startdate<=h_stopdate) | (w_stopdate>=h_startdate & w_stopdate<=h_stopdate))
egen resstartdt=rowmax(w_startdate h_startdate)
egen resstopdt=rowmin(w_stopdate h_stopdate)
format %td resstartdt resstopdt
la def whost 1 "wife" 2 "husband" 3 "both"
foreach st in start stop {
gen who`st'=1 if res`st'dt==w_`st'date
replace who`st'=2 if res`st'dt==h_`st'date
replace who`st'=3 if res`st'dt==w_`st'date & res`st'dt==h_`st'date
la val who`st' whost
replace h_`st'rsn=. if who`st'==1
replace w_`st'rsn=. if who`st'==2
}
```

TYPE: Type of residency data (D)

Data file: all_spouse_pairs_full

Overview

Valid: 38309 Invalid: 158094
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 2 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	ep end	16753	43.7%
2	ep start	21556	56.3%
Sysmiss		158094	

Description

DEFINITION
Type of residency data

RMARDATE: Date known to be married with residency data (D)

Data file: all_spouse_pairs_full

Overview

Valid: 33259
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Date known to be married with residency data

Imputation and derivation

DERIVATION
<couple are assumed to be married according to residency data if date is the start of a co-resident episode, or if the reason for the end of episode is not birth/death nor divorce:>
gen rmardate=rdate if type==2 | (type==1 & w_rsn~=2 & w_rsn~=4 & h_rsn~=2 & h_rsn~=4)

RSEPDATE: Date known to be separated with residency data (D)

Data file: all_spouse_pairs_full

Overview

Valid: 5050
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Date known to be separated with residency data

Imputation and derivation

DERIVATION
<couple is assumed to have separated if reason for end of co-residency episode is birth/death or divorce:>
gen rsepdate=rdate if type==1 & (w_rsn==2 | w_rsn==4 | h_rsn==2 | h_rsn==4)

W_RSN: Wife's reason for res ep start/stop (D)

Data file: all_spouse_pairs_full

Overview

Valid: 36166 Invalid: 160237
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 7 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	Baseline/present	11079	30.6%
2	Birth/Death	343	0.9%
3	Migr - marriage	9618	26.6%
4	Migr - Div/Sep	3888	10.8%
5	Migr - educ	97	0.3%
6	Migr - work	6091	16.8%
7	Migr - other	5050	14%
Sysmiss		160237	

Description

DEFINITION
Wife's reason for res ep start/stop

Imputation and derivation

DERIVATION
variable: rsnmovecenm (table[s]: crs_residency_episodes_clean_poly [derived table])
variable: rsnmovedep (table[s]: crs_residency_episodes_clean_poly [derived table])

<the rsnmove variables from crs_residency_episodes_clean_poly are first renamed to w_stoprsn/h_startrsn etc.>
la def strsn 1 "Baseline/present" 2 "Birth/Death" 3 "Migr - marriage" 4 "Migr - Div/Sep" 5 "Migr - educ" 6 "Migr - work" 7 "Migr - other"
foreach s in w h {
foreach st in start stop {
gen `s'_`st'rsn=1 if `s'_`st'type==1
replace `s'_`st'rsn=2 if `s'_`st'type==2
replace `s'_`st'rsn=3 if (`s'_`st'type==3 | `s'_`st'type==4) & `s'_rsn`st'==1
replace `s'_`st'rsn=4 if (`s'_`st'type==3 | `s'_`st'type==4) & `s'_rsn`st'==2
replace `s'_`st'rsn=5 if (`s'_`st'type==3 | `s'_`st'type==4) & `s'_rsn`st'==3
replace `s'_`st'rsn=6 if (`s'_`st'type==3 | `s'_`st'type==4) & `s'_rsn`st'==4
replace `s'_`st'rsn=7 if (`s'_`st'type==3 | `s'_`st'type==4) & (`s'_rsn`st'==5 | `s'_rsn`st'==.)
la val `s'_`st'rsn strsn
}
}

H_RSN: Husband's reason for res ep start/stop (D)

Data file: all_spouse_pairs_full

Overview

Valid: 31663 Invalid: 164740
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 7 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	Baseline/present	10446	33%
2	Birth/Death	803	2.5%
3	Migr - marriage	6209	19.6%
4	Migr - Div/Sep	1172	3.7%
5	Migr - educ	135	0.4%
6	Migr - work	6995	22.1%
7	Migr - other	5903	18.6%
Sysmiss		164740	

Description

DEFINITION
Husband's reason for res ep start/stop

Imputation and derivation

DERIVATION
variable: rsnmovecenm (table[s]: crs_residency_episodes_clean_poly [derived table])
variable: rsnmovedep (table[s]: crs_residency_episodes_clean_poly [derived table])

<the rsnmove variables from crs_residency_episodes_clean_poly are first renamed to w_stoprsn/h_startrsn etc.>
la def strsn 1 "Baseline/present" 2 "Birth/Death" 3 "Migr - marriage" 4 "Migr - Div/Sep" 5 "Migr - educ" 6 "Migr - work" 7 "Migr - other"
foreach s in w h {
foreach st in start stop {
gen `s'_`st'rsn=1 if `s'_`st'type==1
replace `s'_`st'rsn=2 if `s'_`st'type==2
replace `s'_`st'rsn=3 if (`s'_`st'type==3 | `s'_`st'type==4) & `s'_rsn`st'==1
replace `s'_`st'rsn=4 if (`s'_`st'type==3 | `s'_`st'type==4) & `s'_rsn`st'==2
replace `s'_`st'rsn=5 if (`s'_`st'type==3 | `s'_`st'type==4) & `s'_rsn`st'==3
replace `s'_`st'rsn=6 if (`s'_`st'type==3 | `s'_`st'type==4) & `s'_rsn`st'==4
replace `s'_`st'rsn=7 if (`s'_`st'type==3 | `s'_`st'type==4) & (`s'_rsn`st'==5 | `s'_rsn`st'==.)
la val `s'_`st'rsn strsn
}
}

W_LASTALIVE: Last known date wife alive (D)

Data file: all_spouse_pairs_full

Overview

Valid: 196345
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Last known date wife alive

Imputation and derivation

DERIVATION
variable: stepdate (table[s]: crs_residency_episodes_clean_poly [derived table])
variable: idmoth (table[s]: parents_lafd [derived table])
variable: idfath (table[s]: parents_lafd [derived table])
variable: m_lastAC (table[s]: parents_lafd [derived table])
variable: m_firstDC (table[s]: parents_lafd [derived table])
variable: f_lastAC (table[s]: parents_lafd [derived table])
variable: f_firstDC (table[s]: parents_lafd [derived table])
variable: event_date (table[s]: indices])
variable: adl (table[s]: indices])

<vital status for people is summarised from indices, crs data and reports of parental status, these are then merged into the main dataset for wives and husbands>
use ${gendata}\crs_residency_episodes_clean_nopoly, clear
bys ident: egen lastincrs=max(stopdate)
bys ident (stopdate): gen laststop=stoptype[_N]
format %td lastincrs
keep ident lastincrs laststop
duplicates drop
gen crs=1
save ${tempfiles}\allcrsmembers, replace
gen adl=1 if laststop==1 | laststop==4
replace adl=2 if laststop==2
rename lastincrs intdate
save ${tempfiles}\adlcrs, replace

*get listing of last alive and death dates from parents
use idmoth idfath m_lastAC m_firstDC f_lastAC f_firstDC using ${gendata}\parents_lafd, clear
expand 2, gen(exp)
drop if idmoth=="" & idfath==""
gen ident=idmoth if exp==0
replace ident=idfath if exp==1
gen lastA=m_lastAC if exp==0
replace lastA=f_lastAC if exp==1
gen firstD=m_firstDC if exp==0
replace firstD=f_firstDC if exp==1
keep ident lastA firstD
drop if ident==""
duplicates drop
save ${tempfiles}\parslafd, replace
*add in data from indices
use ${newstata}\indices, clear
replace intdate=event_date if event_date~=.
append using ${tempfiles}\adlcrs
expand 2 if adl==2, gen(exp)
replace adl=1 if exp==1
keep ident adl intdate
drop if adl>=3
bys ident adl: egen lastA=max(intdate) if adl==1
bys ident adl: egen firstD=min(intdate) if adl==2
append using ${tempfiles}\parslafd
bys ident: egen lastalive=max(lastA)

```
bys ident: egen firstdead=min(firstD)
keep ident lastalive firstdead
format %td lastalive firstdead
replace firstdead=. if lastalive>firstdead & lastalive~=.
duplicates drop
save ${tempfiles}\indices_lafd, replace
```

W_FIRSTDEAD: First known date wife dead (D)

Data file: all_spouse_pairs_full

Overview

Valid: 8105
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
First known date wife dead

Imputation and derivation

DERIVATION
variable: stepdate (table[s]: crs_residency_episodes_clean_poly [derived table])
variable: idmoth (table[s]: parents_lafd [derived table])
variable: idfath (table[s]: parents_lafd [derived table])
variable: m_lastAC (table[s]: parents_lafd [derived table])
variable: m_firstDC (table[s]: parents_lafd [derived table])
variable: f_lastAC (table[s]: parents_lafd [derived table])
variable: f_firstDC (table[s]: parents_lafd [derived table])
variable: event_date (table[s]: indices])
variable: adl (table[s]: indices])

```
<vital status for people is summarised from indices, crs data and reports of parental status, these are then merged into the main dataset for wives and husbands>
use ${gendata}\crs_residency_episodes_clean_nopoly, clear
bys ident: egen lastincrs=max(stopdate)
bys ident (stopdate): gen laststop=stoptype[_N]
format %td lastincrs
keep ident lastincrs laststop
duplicates drop
gen crs=1
save ${tempfiles}\allcrsmembers, replace
gen adl=1 if laststop==1 | laststop==4
replace adl=2 if laststop==2
rename lastincrs intdate
save ${tempfiles}\adlcrs, replace

*get listing of last alive and death dates from parents
use idmoth idfath m_lastAC m_firstDC f_lastAC f_firstDC using ${gendata}\parents_lafd, clear
expand 2, gen(exp)
drop if idmoth=="" & idfath==""
gen ident=idmoth if exp==0
replace ident=idfath if exp==1
gen lastA=m_lastAC if exp==0
replace lastA=f_lastAC if exp==1
```

```
gen firstD=m_firstDC if exp==0
replace firstD=f_firstDC if exp==1
keep ident lastA firstD
drop if ident==""
duplicates drop
save ${tempfiles}\parslafd, replace
*add in data from indices
use ${newstata}\indices, clear
replace intdate=event_date if event_date~=.
append using ${tempfiles}\adlcrs
expand 2 if adl==2, gen(exp)
replace adl=1 if exp==1
keep ident adl intdate
drop if adl>=3
bys ident adl: egen lastA=max(intdate) if adl==1
bys ident adl: egen firstD=min(intdate) if adl==2
append using ${tempfiles}\parslafd
bys ident: egen lastalive=max(lastA)
bys ident: egen firstdead=min(firstD)
keep ident lastalive firstdead
format %td lastalive firstdead
replace firstdead=. if lastalive>firstdead & lastalive~=.
duplicates drop
save ${tempfiles}\indices_lafd, replace
```

H_LASTALIVE: Last known date husband alive (D)

Data file: all_spouse_pairs_full

Overview

Valid: 195687
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
Last known date husband alive

Imputation and derivation

DERIVATION
variable: stepdate (table[s]: crs_residency_episodes_clean_poly [derived table])
variable: idmoth (table[s]: parents_lafd [derived table])
variable: idfath (table[s]: parents_lafd [derived table])
variable: m_lastAC (table[s]: parents_lafd [derived table])
variable: m_firstDC (table[s]: parents_lafd [derived table])
variable: f_lastAC (table[s]: parents_lafd [derived table])
variable: f_firstDC (table[s]: parents_lafd [derived table])
variable: event_date (table[s]: indices])
variable: adl (table[s]: indices])

<vital status for people is summarised from indices, crs data and reports of parental status, these are then merged into the main dataset for wives and husbands>

```
use ${gendata}\crs_residency_episodes_clean_nopoly, clear
bys ident: egen lastincrs=max(stopdate)
bys ident (stopdate): gen laststop=stoptype[_N]
```

```
format %td lastincrs
keep ident lastincrs laststop
duplicates drop
gen crs=1
save ${tempfiles}\allcrsmembers, replace
gen adl=1 if laststop==1 | laststop==4
replace adl=2 if laststop==2
rename lastincrs intdate
save ${tempfiles}\adlcrs, replace

*get listing of last alive and death dates from parents
use idmoth idfath m_lastAC m_firstDC f_lastAC f_firstDC using ${gendata}\parents_lafd, clear
expand 2, gen(exp)
drop if idmoth=="" & idfath==""
gen ident=idmoth if exp==0
replace ident=idfath if exp==1
gen lastA=m_lastAC if exp==0
replace lastA=f_lastAC if exp==1
gen firstD=m_firstDC if exp==0
replace firstD=f_firstDC if exp==1
keep ident lastA firstD
drop if ident==""
duplicates drop
save ${tempfiles}\parslafd, replace
*add in data from indices
use ${newstata}\indices, clear
replace intdate=event_date if event_date~=.
append using ${tempfiles}\adlcrs
expand 2 if adl==2, gen(exp)
replace adl=1 if exp==1
keep ident adl intdate
drop if adl>=3
bys ident adl: egen lastA=max(intdate) if adl==1
bys ident adl: egen firstD=min(intdate) if adl==2
append using ${tempfiles}\parslafd
bys ident: egen lastalive=max(lastA)
bys ident: egen firstdead=min(firstD)
keep ident lastalive firstdead
format %td lastalive firstdead
replace firstdead=. if lastalive>firstdead & lastalive~=.
duplicates drop
save ${tempfiles}\indices_lafd, replace
```

H_FIRSTDEAD: First known date husband dead (D)

Data file: all_spouse_pairs_full

Overview

Valid: 21121
Type: Discrete Width: 10 Range: - Format: character

Description

DEFINITION
First known date husband dead

Imputation and derivation

DERIVATION
variable: stepdate (table[s]: crs_residency_episodes_clean_poly [derived table])
variable: idmoth (table[s]: parents_lafd [derived table])
variable: idfath (table[s]: parents_lafd [derived table])
variable: m_lastAC (table[s]: parents_lafd [derived table])
variable: m_firstDC (table[s]: parents_lafd [derived table])
variable: f_lastAC (table[s]: parents_lafd [derived table])
variable: f_firstDC (table[s]: parents_lafd [derived table])
variable: event_date (table[s]: indices])
variable: adl (table[s]: indices])

<vital status for people is summarised from indices, crs data and reports of parental status, these are then merged into the main dataset for wives and husbands>
use ${gendata}\crs_residency_episodes_clean_nopoly, clear
bys ident: egen lastincrs=max(stopdate)
bys ident (stopdate): gen laststop=stoptype[_N]
format %td lastincrs
keep ident lastincrs laststop
duplicates drop
gen crs=1
save ${tempfiles}\allcrsmembers, replace
gen adl=1 if laststop==1 | laststop==4
replace adl=2 if laststop==2
rename lastincrs intdate
save ${tempfiles}\adlcrs, replace

*get listing of last alive and death dates from parents
use idmoth idfath m_lastAC m_firstDC f_lastAC f_firstDC using ${gendata}\parents_lafd, clear
expand 2, gen(exp)
drop if idmoth=="" & idfath==""
gen ident=idmoth if exp==0
replace ident=idfath if exp==1
gen lastA=m_lastAC if exp==0
replace lastA=f_lastAC if exp==1
gen firstD=m_firstDC if exp==0
replace firstD=f_firstDC if exp==1
keep ident lastA firstD
drop if ident==""
duplicates drop
save ${tempfiles}\parslafd, replace
*add in data from indices
use ${newstata}\indices, clear
replace intdate=event_date if event_date~=.
append using ${tempfiles}\adlcrs
expand 2 if adl==2, gen(exp)
replace adl=1 if exp==1
keep ident adl intdate
drop if adl>=3
bys ident adl: egen lastA=max(intdate) if adl==1
bys ident adl: egen firstD=min(intdate) if adl==2
append using ${tempfiles}\parslafd
bys ident: egen lastalive=max(lastA)
bys ident: egen firstdead=min(firstD)
keep ident lastalive firstdead
format %td lastalive firstdead
replace firstdead=. if lastalive>firstdead & lastalive~=.
duplicates drop
save ${tempfiles}\indices_lafd, replace

IDENT: Unique identifier

Data file: marital_status_episodes

Overview

Valid: 134177 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Imputation and derivation

DERIVATION
variable: wife (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: husband (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: ident (table[s]: all_marstatus_reports [derived table])

<on all_spouse_pairs_clean_summarised:>
expand 2, gen(exp)
gen ident=wife if exp1==0
replace ident=husband if exp1==1
drop if ident==""

SEX: Sex

Data file: marital_status_episodes

Overview

Valid: 134175 Invalid: 2
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Imputation and derivation

DERIVATION
variable: sex (table[s]: gen_identity_clear [derived table])

<on all_spouse_pairs_clean_summarised:>
expand 2, gen(exp1)
gen sex=0 if exp1==1
replace sex=1 if exp1==0
<on all_marstatus_reports:>
merge m:1 ident using ${newstata}\gen_identity_clean, keep(master match) keepusing(sex) nogen

DATE1: Start date of episode (D)

Data file: marital_status_episodes

Overview

Valid: 134177
Type: Discrete Width: 10 Range: - Format: character

Imputation and derivation

DERIVATION
variable: mardate (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: fstknwmardate (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: stillmardate (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: sepdate (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: intdate (table[s]: all_marital_reports [derived table])

The summary dates for all_spouse_pairs_clean_summarised converted to episodes and then combined with episodes created from all_marital_reports and are then processed to remove over-lapping time.

DTEST1: if date1 is estimated (D)

Data file: marital_status_episodes

Overview

Valid: 134177 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 5 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	No	109115	81.3%
1	Day est	0	0%
2	Day&month	20509	15.3%
3	est due to overlap	198	0.1%
4	ep censor	0	0%
5	ep split	4355	3.2%

Imputation and derivation

DERIVATION
variable: mardtest (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: fstknwmardtest (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: stillmardtest (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: sepdtest (table[s]: all_spouse_pairs_clean_summarised [derived table])

This variable is taken from the all_spouse_pairs_clean_summarised and the all_marital_reports datasets, depending on which record is retained for the episode, if the episode starts or ends due to episode censor or split, this variable reflects this.

DATATYPE1: data type of date1 (D)

Data file: marital_status_episodes

Overview

Valid: 134177 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 6 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	Report	87850	65.5%
2	Residency	19294	14.4%
3	Both	0	0%
4	Death	1451	1.1%
5	ep censor	21388	15.9%
6	ep split	4194	3.1%

Imputation and derivation

DERIVATION
variable: datatypeM (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: datatypeFM (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: datatypeSM (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: datatypeS (table[s]: all_spouse_pairs_clean_summarised [derived table])

This variable is taken from the all_spouse_pairs_clean_summarised and the all_marital_reports datasets, depending on which record is retained for the episode, if the episode starts or ends due to episode censor or split, this variable reflects this.

QUES1: source if marital status report1 (D)

Data file: marital_status_episodes

Overview

Valid: 58561 Invalid: 0
Type: Discrete Width: 3 Range: - Format: character

Imputation and derivation

DERIVATION
variable: ques (table[s]: all_marital_reports [derived table])

This variable is taken from the all_marital_reports dataset if the start if the episode is due to a report

INFORMANT1: informant of date1 (D)

Data file: marital_status_episodes

Overview

Valid: 134177 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 9 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	proxy	41270	30.8%
1	self	23340	17.4%
2	spouse	7215	5.4%
3	self&spouse	15163	11.3%
4	res data	20745	15.5%
5	ep censor	21388	15.9%
6	ep split	4646	3.5%
9	nk if self or proxy	410	0.3%

Imputation and derivation

DERIVATION
variable: informantM (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: informantFM (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: informantSM (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: informantS (table[s]: all_spouse_pairs_clean_summarised [derived table])

the informant variables are taken from the all_spouse_pairs_clean_summarised dataset, they are only changed if there is more than one record per date (the average value is used)

RSN1: more detailed description of date1 (D)

Data file: marital_status_episodes

Overview

Valid: 134177 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 9 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	StatusRep	59022	44%
2	Marriage	25373	18.9%
3	1stRepOfMar	6571	4.9%
4	LastRepOfMar	7140	5.3%
5	Divorce	9122	6.8%
6	Death (sp)	1367	1%
7	Death (own)	0	0%

8	ep censor	21388	15.9%
9	ep split	4194	3.1%

Imputation and derivation

DERIVATION
variable: mardate (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: fstknwmardate (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: stillmardate (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: sepdate (table[s]: all_spouse_pairs_clean_summarised [derived table])

This variable is taken from the all_spouse_pairs_clean_summarised and the all_marital_reports datasets, depending on which record is retained for the episode, if the episode starts or ends due to episode censor or split, this variable reflects this.

MARITAL: marital status for episode (D)

Data file: marital_status_episodes

Overview

Valid: 134167 Invalid: 10
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 9 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	Never married	35166	26.2%
1	Married	53656	40%
2	Divorced/separated	17841	13.3%
3	Widowed	5040	3.8%
4	N/A (dead)	0	0%
8	Unknown	0	0%
9	Missing	22464	16.7%
Sysmiss		10	

Imputation and derivation

DERIVATION
variable: mardate (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: fstknwmardate (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: stillmardate (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: sepdate (table[s]: all_spouse_pairs_clean_summarised [derived table])
variable: marital (table[s]: all_marital_reports [derived table])

For episodes created using all_spouse_pairs_clean_summarised within the main marriage dates episode the marital status is married, an episode created using the end date of the main episode has the marital status as divorced or widowed, or unknown depending what is know about the marriage. The episodes created using all_marital_reports use the original marital variable appended. Following cleaning and adding spouse ids marital is updated if required.

SPOUSENUM: number of spouses for episode (D)

Data file: marital_status_episodes

Overview

Valid: 55950 Invalid: 78227
Type: Discrete Decimal: 0 Width: 2 Range: 1 - 99 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category
97	Not applicable
98	Unknown
99	Missing

Imputation and derivation

DERIVATION
variable: spousenum (table[s]: all_marital_reports [derived table])
variable: idspouse (table[s]: all_spouse_pairs_clean_summarised [derived table])

this come directly from all_marital_reports and is updated once the spouse ids have been added into the episode data.

REPBY1: who reported the union [1] (D)

Data file: marital_status_episodes

Overview

Valid: 45241 Invalid: 88936
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 4 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	8156	18%
2	husband	3380	7.5%
3	both	33705	74.5%
4	Residency data	0	0%
Sysmiss		88936	

Imputation and derivation

DERIVATION
variable: repby (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and repby for the spouse assigned to the record where they were current spouses

PARLINK1: whether union only known about through co-parent link and co-residency [1] (D)

Data file: marital_status_episodes

Overview

Valid: 48270 Invalid: 85907
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Imputation and derivation

DERIVATION
variable: newparlink (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and parlink for the spouse assigned to the record where they were current spouses

DATE2: End date of episode (D)

Data file: marital_status_episodes

Overview

Valid: 134177
Type: Discrete Width: 10 Range: - Format: character

Imputation and derivation

DERIVATION
variable: date1 (table[s]:this dataset)

The end of date of each episode is the start date of the next episode or 1 year after date1 if the next episode is over 12 months later (or doesn't exist)

DTEST2: if date2 is estimated (D)

Data file: marital_status_episodes

Overview

Valid: 134177 Invalid: 0
Type: Discrete Decimal: 0 Width: 5 Range: 0 - 17800 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	No	123505	92%
1	Day est	0	0%
2	Day&month	6176	4.6%
3	est due to overlap	164	0.1%
4	ep censor	0	0%
5	ep split	4331	3.2%
17800		1	0%

Imputation and derivation

DERIVATION
variable: dtest1 (table[s]:this dataset)

This variable is taken from the all_spouse_pairs_clean_summarised and the all_marital_reports datasets, depending on which record is retained for the episode, if the episode starts or ends due to episode censor or split, this variable reflects this.

DATATYPE2: data type of date2 (D)

Data file: marital_status_episodes

Overview

Valid: 134177 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 6 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	Report	25812	19.2%
2	Residency	18685	13.9%
3	Both	0	0%
4	Death	1909	1.4%
5	ep censor	83577	62.3%
6	ep split	4194	3.1%

Imputation and derivation

DERIVATION
variable: datatype1 (table[s]:this dataset)

This variable is taken from the all_spouse_pairs_clean_summarised and the all_marital_reports datasets, depending on which record is retained for the episode, if the episode starts or ends due to episode censor or split, this variable reflects this.

QUES2: source if marital status report2 (D)

Data file: marital_status_episodes

Overview

Valid: 18727 Invalid: 0
Type: Discrete Width: 3 Range: - Format: character

Imputation and derivation

DERIVATION
variable: ques (table[s]: all_marital_reports [derived table])

This variable is taken from the all_marital_reports dataset if the start if the episode is due to a report

INFORMANT2: informant of date2 (D)

Data file: marital_status_episodes

Overview

Valid: 134177 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 9 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	proxy	12596	9.4%
1	self	7638	5.7%
2	spouse	2399	1.8%
3	self&spouse	1204	0.9%
4	res data	20594	15.3%
5	ep censor	83577	62.3%
6	ep split	6094	4.5%
9	nk if self or proxy	75	0.1%

Imputation and derivation

DERIVATION
variable: informant1 (table[s]:this dataset)

This variable is taken from the all_spouse_pairs_clean_summarised and the all_marital_reports datasets, depending on which record is retained for the episode, if the episode starts or ends due to episode censor or split, this variable reflects this.

RSN2: more detailed description of date2 (D)

Data file: marital_status_episodes

Overview

Valid: 134177 Invalid: 0
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 9 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	StatusRep	19429	14.5%
2	Marriage	8434	6.3%
3	1stRepOfMar	2058	1.5%
4	LastRepOfMar	5811	4.3%
5	Divorce	9986	7.4%
6	Death (sp)	688	0.5%
7	Death (own)	0	0%
8	ep censor	83577	62.3%
9	ep split	4194	3.1%

Imputation and derivation

DERIVATION
variable: rsn1 (table[s]:this dataset)

This variable is taken from the all_spouse_pairs_clean_summarised and the all_marital_reports datasets, depending on which record is retained for the episode, if the episode starts or ends due to episode censor or split, this variable reflects this.

IDSPOUSE1: spouses's id [1] (D)

Data file: marital_status_episodes

Overview

Valid: 48270 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Imputation and derivation

DERIVATION
variable: idspouse (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and spouse ids assigned to the record where they were current spouses

AGEDIFF1: age difference between individual and their spouse [1] (D)

Data file: marital_status_episodes

Overview

Valid: 48270 Invalid: 85907
Type: Continuous Decimal: 0 Width: 2 Range: 0 - 44 Format: Numeric

Imputation and derivation

DERIVATION
variable: agediff (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and agediff for the spouse assigned to the record where they were current spouses

IDSPOUSE2: spouses's id [2] (D)

Data file: marital_status_episodes

Overview

Valid: 2444 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Imputation and derivation

DERIVATION
variable: idspouse (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and spouse ids assigned to the record where they were current spouses

AGEDIFF2: age difference between individual and their spouse [2] (D)

Data file: marital_status_episodes

Overview

Valid: 2444 Invalid: 131733
Type: Continuous Decimal: 0 Width: 2 Range: 0 - 44 Format: Numeric

Imputation and derivation

DERIVATION
variable: agediff (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and agediff for the spouse assigned to the record where they were current spouses

REPBY2: who reported the union [2] (D)

Data file: marital_status_episodes

Overview

Valid: 2372 Invalid: 131805
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 4 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	249	10.5%
2	husband	211	8.9%
3	both	1912	80.6%
4	Residency data	0	0%
Sysmiss		131805	

Imputation and derivation

DERIVATION
variable: repby (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and repby for the spouse assigned to the record where they were current spouses

PARLINK2: whether union only known about through co-parent link and co-residency [2] (D)

Data file: marital_status_episodes

Overview

Valid: 2444 Invalid: 131733
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Imputation and derivation

DERIVATION
variable: newparlink (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and parlink for the spouse assigned to the record where they were current spouses

IDSPOUSE3: spouses's id [3] (D)

Data file: marital_status_episodes

Overview

Valid: 98 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Imputation and derivation

DERIVATION
variable: idspouse (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and spouse ids assigned to the record where they were current spouses

AGEDIFF3: age difference between individual and their spouse [3] (D)

Data file: marital_status_episodes

Overview

Valid: 98 Invalid: 134079
Type: Continuous Decimal: 0 Width: 2 Range: 0 - 41 Format: Numeric

Imputation and derivation

DERIVATION
variable: agediff (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and agediff for the spouse assigned to the record where they were current spouses

REPBY3: who reported the union [3] (D)

Data file: marital_status_episodes

Overview

Valid: 96 Invalid: 134081
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 4 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	10	10.4%
2	husband	13	13.5%
3	both	73	76%
4	Residency data	0	0%
Sysmiss		134081	

Imputation and derivation

DERIVATION
variable: repby (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and repby for the spouse assigned to the record where they were current spouses

PARLINK3: whether union only known about through co-parent link and co-residency [3] (D)

Data file: marital_status_episodes

Overview

Valid: 98 Invalid: 134079
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

Imputation and derivation

DERIVATION
variable: newparlink (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and parlink for the spouse assigned to the record where they were current spouses

IDSPOUSE4: spouses's id [4] (D)

Data file: marital_status_episodes

Overview

Valid: 12 Invalid: 0
Type: Discrete Width: 7 Range: - Format: character

Imputation and derivation

DERIVATION
variable: idspouse (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and spouse ids assigned to the record where they were current spouses

AGEDIFF4: age difference between individual and their spouse [4] (D)

Data file: marital_status_episodes

Overview

Valid: 12 Invalid: 134165
Type: Continuous Decimal: 0 Width: 2 Range: 1 - 33 Format: Numeric

Imputation and derivation

DERIVATION
variable: agediff (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and agediff for the spouse assigned to the record where they were current spouses

REPBY4: who reported the union [4] (D)

Data file: marital_status_episodes

Overview

Valid: 12 Invalid: 134165
Type: Discrete Decimal: 0 Width: 1 Range: 1 - 4 Format: Numeric

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	wife	0	0%
2	husband	0	0%
3	both	12	100%
4	Residency data	0	0%
Sysmiss		134165	

Imputation and derivation

DERIVATION
variable: repby (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and repby for the spouse assigned to the record where they were current spouses

PARLINK4: whether union only known about through co-parent link and co-residency [4] (D)

Data file: marital_status_episodes

Overview

Valid: 12 Invalid: 134165
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 0 Format: Numeric

Imputation and derivation

DERIVATION
variable: newparlink (table[s]: all_spouse_pairs_clean_summarised [derived table])

Data from all_spouse_pairs_clean_summarised is converted to a wide listing of spouses per person with start and end dates, this is merged into the episode data and parlink for the spouse assigned to the record where they were current spouses