DERIVED VARIABLES LIST: INFORMATION FOR LS USERS

The following syntax files are available for researchers using the LS to use.

The syntax was written by CeLSIUS, unless otherwise stated below. Please email celsius@ucl.ac.uk if you have any questions or suggestions related to the files below.

IMPORTANT INFORMATION (PLEASE READ):

- You will need to check the syntax to ensure that it does what you require. Although the syntax has been checked, it does **not** mean that it will do what is required for your purpose. You may need to adapt it to your needs.
  ****

COMMUNAL ESTABLISHMENTS
ceresid0.do
Selects staff in communal establishments for reclassification as residents if they are:
- Age 0-15 (age not imputed)
- Age ≥75 (age not imputed)
- Age 16-74 and economically inactive (age and economic activity not imputed)
- Age is imputed and economically inactive (economic activity not imputed)

Variables needed (from ME01): posp0, agep0, agepimp, ecop0, ecopimp.

This code can be adapted to reclassify staff in communal establishments to residents using similar variables from ME11: posp11, posp11_imp, agep11, agep11_imp, ecop11
****

COUNTRY OF BIRTH
cob0_englang.do
Syntax written by Ben Wilson (LSE).
Creates a variable showing whether English is an official language in LS member's country of birth as reported in 2001.
Variables needed (from ME01): cobp0.

cob9_englang.do
Syntax written by Ben Wilson (LSE).
Creates a variable showing whether English is an official language in LS member's country of birth as reported in 1991.
Variables needed (from ME91): cobp9.

cob01_harm.do
This syntax creates a harmonised variable for cob in 2001 (replacing missing values with cob from 1991, cob from 1981 if 1991 missing and cob from 1971 if 1981 missing.
It works in two parts, the first recodes 1971, 1981 and 1991 country of birth variables to 2001 categories and the second creates the harmonised variable.
Variables needed: pob7 (from ME71), cob8 (from ME81), cob9 (from ME91), and cobp0 and cobpimp (from ME01).

cobeumem.do
Generates a variable showing when countries joined the EU.
NB: West Germany joined the EU in 1951, but the 2001 code does not distinguish between West and East Germany, therefore everyone born in Germany is included even though only West Germany was a member of the EU until reunification in 1990.
There are two versions of the syntax in the file; version 1 derives the variable based on the 2001 country of birth (i.e. cobp0 from ME01); version 2 derives the variable based on the 2011 country of birth (i.e. cobp11 from ME11).
- Variables needed for version 1 (from ME01): cobp0, cobpimp.
- Variables needed for version 2 (from ME11): cobp11, cobp11_imp.

cobgrp01.do
Generates a variable that groups country of birth in 2001 into 47 categories:
  - 1: England;2: Wales; 3: Scotland; 4: Northern Ireland; 5: UK, part not specified; 6: Republic of Ireland; 7: Ireland part not specified; 8: Channel Islands and Isle of Man; 9: France; 10: Germany; 11: Italy; 12: Netherlands; 13: Spain; 14: Other country in European Union; ;15: Other countries in Western Europe not in EU; 16: Poland; 17: Other Eastern Europe; 18: North Africa; 19: Nigeria; 20: Other Central and Western Africa; 21: Kenya; 22: South Africa; 23: Zimbabwe; 24: Other South and Eastern Africa; 25: Cyprus; 26: Iran; 27 Other Middle East; 28: China; 29: Hong Kong; 30: Japan; 31: Malaysia; 32: Singapore; 33: Other Far East; 34: Bangladesh; 35: India; 36: Pakistan; 37: Other South Asia; 38: Canada; 39: Jamaica; 40: Other Caribbean and West Indies; 41: USA; 42: Other North America; 43: South America; 44: Australia; 45: New Zealand; 46: Other Oceania; 99: Other.

Variables needed (from ME01): cobp0, cobpimp.
Name of derived variable: cobgroup
[The syntax has been adapted to create a syntax file that does the same for 2011. (cobgrp11.do), using the variables cobp11 and cobp11_imp from ME11.]

cobgrp71.do
Groups 1971 country of birth into 39 categories (as for 2001), but there were no categories for North Africa, Japan, Malaysia, Singapore, Bangladesh, Other North America and Other Oceania.
[These categories were not needed because all the countries that were used in pob7 could be allocated to one of the 39 categories in the derived variable. For example, no North African countries were included in pob7; Bangladesh did not exist in 1971 and therefore could not be stated as a country in pob7; Japan, Malaysia and Singapore could be categorised as "Other Far East".]
Variables needed (from ME71): pob7.
Name of derived variable: cobgrp71

cobgrp81.do
Groups 1981 country of birth into 46 categories (no categories for Other Oceania, but additional category for 98: Not stated).
Variables needed (from ME81): cob8.
Name of derived variable: cobgrp81

cobgrp91.do
Groups 1991 country of birth into 48 categories (as 2001).
Variables needed (from ME91): cob9.
Name of derived variable: cobgrp91

NB: Users could amend the code for each year to put the countries into fewer categories than the 40+ in the syntax files above, e.g. put European countries into 3 categories: UK, EU, Non-EU

cobhier.txt
This is not a syntax file. It lists the countries that cobgrp71, cobgrp81, cobgrp91, cobgroup and cobgroup11 are categorised into, giving higher levels that researchers might want to recode the 47 categories into.

cobnbir.txt
This derives variables for the country of birth of the mother of LS members who are born into the study (i.e. born on one of the four birthdays that are used to select LS members from 1971 onwards). These variables cannot be derived for LS members who were born prior to 1971.
It creates three variables from mobplbinb (place of birth of mother of new birth [in NBIR])
  - moplpre92 for 1971-1991
  - moplfr92 for 1992-2006
  - moplfr07 for 2007 onwards
Gives the three derived variables value labels.

The syntax is also available for SPSS users.

Variable needed (from NBIR): moplbinb.

Names of derived variables: moplpre92 (place of birth of mother for births prior to 1992, i.e. 1971-1991), moplfr92 (place of birth of mother for births from 1992-2006) and moplfr07 (place of birth of mother for live births from 2007).

****

ECONOMIC ACTIVITY
ecac9101.do
Generates two economic activity variables (one for 1991 [employ9] and other for 2001 [employ0]) from econpo89 and ecop80, and recodes them to:

> 1 = employed; 2 = unemployed; 3 = students; 4 = house minders; 5 = others (sick, retired, other)

The syntax includes an option to set student to employed/ unemployed and lose the student group. There is also syntax creating an employment status variable [empstat] which merges the two variables created above (employ9 and employ0), with the following categories:

> 1= employed 91 & 01; 2=employed 91 unemployed 01; 3=employed 91 student 01; 4=employed 91 home minder 01; 5=employed 91 other 01; 6=unemployed 91 employed 01; 7=unemployed 91 & 01; 8= unemployed 91 student 01; 9=unemployed 91 home minder 01; 10=unemployed 91 other 01; 11=student 91 employed 01; 12=student 91 unemployed 01; 13=student 91 student 01; 14=student 91 home minder 01; 15=student 91 other 01; 16=home minder 91 employed 01; 17=home minder 91 unemployed 01; 19=home minder 91 home minder 01; 20=home minder 91 other 01; 21=other 91 employed 01; 22=other 91 unemployed 01; 23=other 91 student 01; 24=other 91 home minder 01; 25=other 91 other 01

Variables needed: econpo89 (from ME91), ecop80 (from ME01).
Names of derived variables: employ9 (employment status in 1991), employ0 (employment status in 2001) and empstat9101 (longitudinal employment status 1991-2001).

ecac0111.do
Generates two economic activity variables (one for 2001 [employ0] and other for 2011 [employ11]) from ecop80 and ecop81, and recodes them to:

> 1 = employed; 2 = unemployed; 3 = students; 4 = house minders; 5 = others (sick, retired, other)

The syntax includes an option to set student to employed/ unemployed and lose the student group
There is also syntax creating an employment status variable [empstat] which merges the two variables created above (employ0 and employ11), with the following categories:

> 1= employed 01 & 11; 2=employed 01 unemployed 11; 3=employed 01 student 11; 4=employed 01 home minder 11; 5=employed 01 other 11; 6=unemployed 01 employed 11; 7=unemployed 01 & 11; 8= unemployed 01 student 11; 9=unemployed 01 home minder 11; 10=unemployed 01 other 11; 11=student 01 employed 11; 12=student 01 unemployed 11; 13=student 01 student 11; 14=student 01 home minder 11; 15=student 01 other 11; 16=home minder 01 employed 11; 17=home minder 01 unemployed 11; 19=home minder 01 home minder 11; 20=home minder 01 other 11; 21=other 01 employed 11; 22=other 01 unemployed 11; 23=other 01 student 11; 24=other 01 home minder 11; 25=other 01 other 11.

Variables needed: ecop80 (from ME01) and ecop81 (from ME11).
Names of derived variables: employ0 (employment status in 2001), employ11 (employment status in 2011) and empstat0111 (longitudinal employment status 2001-2011).
****

SOCIAL CLASS

epsclas8.do

Creates a new social class variable [epsclas8], with a category for unemployment in 1981
Categories in new variable:

1=Professional; 2=Managerial; 3=Skilled Non-Manual; 4=Skilled Manual; 5=Part Skilled;
6=Unskilled; 7=Armed Forces; 9=Unemployed; 10=In education; 11=Housewives/Husbands;
12=Other inactive.

Variables needed (from ME81): econact8, soc8.
Name of new variable: epsclas8.

epsclas9.do

Creates a new social class variable [epsclas9], with a category for unemployment in 1991.
Categories in new variable:

1=Professional; 2=Managerial; 3=Skilled Non-Manual; 4=Skilled Manual; 5=Part Skilled;
6=Unskilled; 7=Armed Forces; 8=Other employed; 9=Unemployed; 10=Student; 11=Looking
after home; 12=Other.

Variables needed (from ME91): econpo89, sclas9.
Name of new variable: epsclas9

epsclas0.do

Creates a variable [epsclas0] similar to roserg0 with an additional category for unemployed.
Categories in new variable:

1=Professional; 2=Managerial; 3=Skilled Non-Manual; 4=Skilled Manual; 5=Part Skilled;
6=Unskilled; 7=Armed Forces; 9=Unemployed; 10=Student; 11=Looking after home;
12=Other.

Variables needed (from ME01): roserg0, ecop80.
Name of new variable: epsclas0.

epsclas11.do

Creates a variable [epsclas11] similar to rgsc11 with an additional category for unemployed.
Categories in new variable:

1=Professional; 2=Managerial; 3=Skilled Non-Manual; 4=Skilled Manual; 5=Part Skilled;
6=Unskilled; 7=Armed Forces; 9=Unemployed; 10=Student; 11=Looking after home; 12=Other.

Variables needed (from ME11): rgsc11, ecop81.
Name of new variable: epsclas11.

****

ETHNICITY
ethgrp7.do
This syntax generates an ethnicity variable for LS members in 1971 (ethgrp7) using their place of birth (if both their parents are in a different place to the LS member, then their place of birth over-rides the LS member's own).
Variables needed (from ME71): pob7, pobf7, pobm7.
The new variable is coded: 1 White, 2 Black/mixed, 3 IndPakBang, 4 ChinJapEtc, 5 Other.
Name of new variable: ethgrp7.

ethgrp8.do
This syntax generates an ethnicity variable for LS members in 1981 (ethgrp8) using their country of birth.
Variables needed (from ME81): cob8.
The new variable is coded: 1 White, 2 Black/mixed, 3 IndPakBang, 4 ChinJapEtc, 5 Other.
Name of new variable: ethgrp8.

ethgrp9.do
This syntax generates an ethnicity variable for LS members in 1991 (ethgrp9) using their ethnicity variable.
Variables needed (from ME91): ethnic9.
The new variable is coded: 1 White, 2 Black/mixed, 3 IndPakBang, 4 ChinJapEtc, 5 Other.
Name of new variable: ethgrp9.

ethgrp0.do
This syntax generates an ethnicity variable for LS members in 2001 (ethgrp0) using their ethnicity variable.
Variables needed (from ME01): ethgrp0.
The new variable is coded: 1 White, 2 Black, 3 IndPakBang, 4 ChinJapEtc, 5 Mixed, 6 Other (i.e. as per 1971/1981/1991, but with an additional category for Mixed).
Name of new variable: ethgrp0.

ethgrp11.do
This syntax generates an ethnicity variable for LS members in 2011 (ethgrp11) using their ethnicity variable.
Variables needed (from ME11): ethgrp11.
The new variable is coded: 1 White, 2 Black, 3 IndPakBang, 4 ChinJapEtc, 5 Mixed, 6 Other (i.e. as per 1971/1981/1991, but with an additional category for Mixed).
Name of new variable: ethgrp11.
****

HIGHER EDUCATION INDICATOR
heducat0.do
Generates a binary variable indicating if an LS member has a degree or higher education qualification in 2001 using hlqp0 and qup0 (string variable with 12 digits [1 in position 5 or 6 denotes a degree or higher education qualification]).
The variable has two categories:
- − 0=No degree
- − 1=Has degree

Any individuals with imputed values in hlqpimp or qupimp are coded to missing (.).
Variables needed (from ME01): hlqp0, hlqpimp, qup0 qupimp
Name of new variable: heducat0

heducat11.txt
As above, this syntax generates a binary variable indicating if an LS member has a degree or higher education qualification in 2011 using hlqp11, qup11 (string variable with 13 digits [1 in position 8, 9 or 10 denotes a degree or higher education qualification]) and qup11_imp.
The variable has two categories:
- − 0=No degree
- − 1=Has a degree

Any individuals with imputed values in qup11_imp are coded to missing (.).
Variables needed (from ME11): hlqp11, qup11 qup11_imp
Name of new variable: heducat11
        ****

HOUSEHOLD LEVEL EMPLOYMENT VARIABLE

hhemp9.do

This syntax generates a binary variable (hhemp9) indicating if an LS member lives in a household with economically active individuals in 1991.

Variables needed (from ME91): eonaem9, econp09.

Name of new variable: hhemp9.

hrpemp0.do

This syntax generates a binary variable (hrpemp0) indicating if an LS member lives in a household where the household reference person is economically active 2001.

Variables needed: agep0, ageimp, nssp0, nsspimp, hrpp0, hrppimp, ecop0 and ecopimp (from ME01); and nagep0, nageimp, nnssp0, nnsspimp, nhrpp0, necop0 and necopimp (from NM01).

Name of new variable: hrpemp0.

This syntax can be adapted to derive a similar binary variable for 2011, using similar variables from ME11 (agep11, agep11_imp, nssp11, hrpp11 and ecop11) and NM11 (nagep11, nagep11_imp, nnssp11, nhrpp11 and necop11).

****

HOUSEHOLD AND FAMILY TYPE INDICATOR

hhfam9.do
Derives a variable (housefam9) indicating the household and family type of the LS member in 1991.
The variable is categorised into 12 categories:

> 1=solitary; 2=couple only; 3=couple + children; 4=couple + others; 5=couple, child + other; 6=lone parent; 7=lone parent + other; 8=2+ families; 9=no family + other; 10=child in family; 11=communal establishment; 12=Other

Variables needed: geninfm9, hhfmtyt9, mhucom9, resclas9 and uresind9 (from ME91) and hiscen91 (from CORE1).
Name of new variable: housefam9.

hhfam0.do
Derives a variable (housfam0) indicating the household and family type of the LS member in 2001.
The variable is categorised into 11 categories:

> 1=solitary; 2=couple only; 3=couple + children; 4=couple + others; 5=couple, child + other; 6=lone parent; 7=lone parent + other; 8=2+ families; 9=no family + other; 10=child in family; 11=communal establishment

Variables needed: geninf0 (generation in family), mhucom0 (combination of minimal household units within household), relfimp (imputation flag for family relationship) (from ME01), and hiscen01 (from CORE1).
Name of new variable: housefam0.

hhfam11.do
Derives a variable (housfam11) indicating the household and family type of the LS member in 2011.
The variable is categorised into 11 categories:

> 1=solitary; 2=couple only; 3=couple + children; 4=couple + others; 5=couple, child + other; 6=lone parent; 7=lone parent + other; 8=2+ families; 9=no family + other; 10=child in family; 11=communal establishment

Variables needed: geninfam11 (generation in family), mhucom11 (combination of minimal household units within household), rel2p111_imp (imputation flag for relationship to person 1) (from ME11), and hiscen11 (from CORE1).
Name of new variable: housefam11.

hhfam01c91.do
This syntax requires two datasets, which you will need to ask the USO for:

> – housfam1: contains coreno and hiscen01 (from CORE1); faminhh0, fmtfam0, geninf0, ieoind, lsfnum0, mhucom0, mstp0, mstpimp, relpimp, relhimp, relfimp, stup0, stupimp, ttind0 and ttimp (from ME01).
> – housfam2: contains coreno, nlsfnum0, lsrelat0, nmstp and nmstpimp (from NM01).
> – The USO will move both datasets to your project in the SRS so that you can run the syntax on them.

The syntax generates a variable (fam01c91) describing living arrangements of LS members in 2001, which is compatible with the 1991 categories.
The variable has 12 categories:

> 1=solitary; 2=couple only; 3=couple + children; 4=couple + others; 5=couple + children + other; 6=lone parent; 7=lone parent + other; 8=2+ families; 9=no family + other; 10=child in family; 11=communal est.; -8=imputed

Name of new variable: fam01c91.

There is also syntax that creates alternative versions of the above that exclude (set to -8) cases with imputation at the household, family or person level.

hhfam11c91.do

This syntax requires two datasets, which you will need to ask the USO for:

- housfam1: contains coreno and hiscen11 (from CORE1); faminhh11, famno11, fmtfam11, geninfam11, mhucom11, mstp11, mstp11_imp, rel2p111_imp, stup11, stup11_imp, ttind11 and ttind11_imp (from ME11).
- housfam2: contains coreno, nfmtfam11, nfamno11, lsrelat11, nmstp11 and nmstpi11_imp (from NM11).
- The USO will move both datasets to your project in the SRS so that you can run the syntax on them.

The syntax generates a variable (fam11c91) describing living arrangements of LS members in 2011, which is compatible with the 1991 categories.

The variable has 12 categories:

1=solitary; 2=couple only; 3=couple + children; 4=couple + others; 5=couple + children + other; 6=lone parent; 7=lone parent + other; 8=2+ families; 9=no family + other; 10=child in family; 11=communal est.; -8=imputed

Name of new variable: fam11c91.

There is also syntax that creates alternative versions of the above that exclude (set to -8) cases with imputation at the household, family or person level.

****

INTERNATIONAL CLASSIFICATION OF DISEASES
icd20.txt

Generates one cause of death variable (udcod) based on the underlying cause of death (UCD) information from the various versions of the ICD that are used to code underlying cause of death of LS members.

The derived variable has 20 categories:

> 0=Alive; 1=IHD; 2=Stroke; 3=Pulmonary Disease; 4=Lung Cancer; 5=Other Cancers; 6=Infectious & Parasitic; 7=Diabetes; 8=Gastro-intestinal tract disease; 9=Liver Disease; 10=Mental & Behavioural; 11=Abnormalities & Lab results; 12=Other Circulatory Disease; 13=Accidents, self-harm; 14=Muscular Diseases; 15=Benign Neoplasms; 16=Nervous System; 17=Genito-Urinary; 18=Other Endocrine; 19=Skin disease; 20=Other Cause

It also generates a finer classification (finecause) with 35 categories:

- 0=Alive; 1= Infectious & Parasitic; 2=Diabetes; 3=Pneumonia; 4=Stomach neoplasm; 5=Neoplasm: Colon, Rectum, Anus; 6= Neoplasm: Pancreas; 7= Neoplasm: Larynx, Trachea, Lung; 8= Neoplasm: Breast; 9= Other Neoplasms; 10= Intestinal disease; 11 = Liver disease; 12= IHD; 13= CerebroVascu; 14=Mental&Behavioural; 15= Abnormalities and Lab results; 16= Respiratory system; 17= Other Circulatory Disease; 18= Muscular diseases; 19= Accident&Violence; 20= Accidental poisoning; 21=Suicide, self-harm; 22= Assault; 23= Undetermined intent; 24= Other Accident&Violence; 25= Benign Neoplasm; 26= Nervous System; 27= Other respiratory; 28= Genito-Urinary; 29= Other Endocrine; 30= Metabolic; 31= Skin disease; 32= Other digestive; 33= Pregnancy complications; 35= Cause unknown.

Variables needed (from DETH): ic10ude (ICD10 code for UCD), ic10ufde (final ICD10 code for UCD), ucde3dde (ICD9 3-digit final UCD (1979-1992)), icd9ude (ICD9 (1993 onwards) for UCD), icd9ufde (ICD9 (1993 onwards) for final UCD), ucde3cde (ICD8 first 3 digits of UDC (1971-1981)), deyrbde (year of death of LS member).

Names of new variables: udcod and finecause.

****

GOVERNMENT OFFICE REGION (GOR)
gor7.do
Derives a Government Office Region variable for 1971 data (gor7) that is compatible with the codings used for GOR in 2001.

Requires use of the GOR look-up dataset, which you will need to ask a CeLSIUS USO for.

Categories in new variable:

1=North East; 2=North West; 3=Yorks and Humberside; 4=East Midlands; 5=West Midlands; 6=East of England; 7=London; 8=South East; 9=South West; 10=Wales.

Variables needed (from ME71): co7 (county of usual address).

Name of new variable: gor7

gor8.do
Derives a Government Office Region variable for 1981 data (gor8) that is compatible with the codings used for GOR in 2001.

Requires use of the GOR look-up dataset, which you will need to ask a CeLSIUS USO for.

Categories in new variable:

1=North East; 2=North West; 3=Yorks and Humberside; 4=East Midlands; 5=West Midlands; 6=East of England; 7=London; 8=South East; 9=South West; 10=Wales.

Variables needed (from ME81): coua8 (county of usual address).

Name of new variable: gor8

gor9.do
Derives a Government Office Region variable for 1991 data (gor9) that is compatible with the codings used for GOR in 2001.

Requires use of the GOR look-up dataset, which you will need to ask a CeLSIUS USO for.

Categories in new variable:

1=North East; 2=North West; 3=Yorks and Humberside; 4=East Midlands; 5=West Midlands; 6=East of England; 7=London; 8=South East; 9=South West; 10=Wales.

Variables needed (from ME91): ctyuaea9 (county of usual address).

Name of new variable: gor9.

gor0.do
Syntax to replace the string variable Government Office Region in 2001 with numerical values (gor0).

The letters A, B, C, D, E, F, G, H, J, K, W are recoded to:

1=North East; 2=North West; 3=Yorks and Humberside; 4=East Midlands; 5=West Midlands; 6=East of England; 7=London; 8=South East; 9=South West; 10=Wales.

Variables needed (from ME01): gors0 (2001 GOR).

Name of new variable: gor0.

gor11.do
Syntax to replace the string variable Government Office Region in 2011 (gor11) with numerical values.

The letters A, B, C, D, E, F, G, H, J, K, W are recoded to:

1=North East; 2=North West; 3=Yorks and Humberside; 4=East Midlands; 5=West Midlands; 6=East of England; 7=London; 8=South East; 9=South West; 10=Wales.

Variables needed: gor11 in ME11.

Name of new variable: gor11.

****

RELIGION

religion0-check.do

Derives a grouped variable for religion in 2001 which groups the 300 categories in relp0 into nine (religion01). The categories in this new variable are compatible with the categories in the 2011 religious group variable (relgp11).

The 300 categories were combined into nine:

1= No religion; 2=Christian; 3=Buddhist; 4=Hindu; 5=Jewish; 6=Muslim; 7=Sikh; 8=Other religion; 9=Not stated.

Variable needed (from ME01): relp0.

Name of new variable: religion01.

****

IMPUTATION FLAG FOR RELATIONSHIP TO OTHER HOUSEHOLD MEMBERS
relimp01.txt

This syntax derives one variable (relimp) for each LS member that shows if any of the relationship to other members of their household variables has been imputed (coded 1 in the variables rel01imp – rel30imp).

Variables needed:

- rel01imp, rel02imp, rel03imp, rel04imp, rel05imp, rel06imp, rel07imp, rel08imp, rel09imp, rel10imp, rel11imp, rel12imp, rel13imp, rel14imp, rel15imp, rel16imp, rel17imp, rel18imp, rel19imp, rel20imp, rel21imp, rel22imp, rel23imp, rel24imp, rel25imp, rel26imp, rel27imp, rel28imp, rel29imp, rel30imp from ME01.
- npersno0 from NM01.
  ****