```
/****************************************************************************
*Malaria early morbidity and mortality                                     *
*1. Data preparation do file                                               *
*Author: Carl Higgs     Last edited: 30 October 2015                       *
****************************************************************************/

***Set up log of analysis***
capture log close
version 13.1
set linesize 100
set more off
cd "C:\Users\Carl\Google Drive\MPH\Projects\Malaria project\Data"
loc today = c(current_date)
log using "malariaproject_log_`today'.txt", append text


***Label unlabelled variables***
label variable dod "Date of first discharge"
label variable Age "Age at date of presentation"
label variable pf "P.falciparum"
label variable pv "P.vivax"
label variable pm "P.malariae"
label variable po "P.ovale"
label variable HbDGr7 "Anaemia at death (< 7g/dL)"
label variable HbDGr5 "Anaemia at death (< 5g/dL)"
label variable Malaria_Last63Gr "Presentations with malaria in the last 2 months"
label variable AGR4 "Age"
label variable Era "ACT Era"
label variable Sex "Sex"
label variable Species "Initial Species"
label variable wbcmin "White Cell Count (minimum)"
label variable wbcmax "White Cell Count (maximum)"
label variable hbmin_Min "Haemoglobin (minimum)"
label variable pltmin "Platelet count (minimum)"
label variable hbmin_First "Haemoglobin (first recorded value)"
label variable hbmin_Last "Haemoglobin (last recorded value)"
label variable aa_any "Amodiaquine+Artesunate"
label variable cq_any "Chloroquine"
label variable cl_any "Clindamycin"
label variable dhp_any "DHA-Piperaquine"
label variable dox_any "Doxycycline"
label variable ivart_any "Iv Artesunate"
label variable ivq_any "Iv Quinine"
label variable oralq_any "Oral Quinine"
label variable pq_any "Primaquine"
label variable sp_any "Sulfadoxine-Pyrimethamine"
label variable PQmgkg_Sum "Sum total dose of Primaquine (mg/kg)"
label variable MalNut "Under-nutrition"
label variable MalNutD "Under-nutrition at death"
label variable op "Outpatient"
label variable Source "Source"
label variable ip "Inpatient"
label variable Bleeding "Bleeding"
label variable EthnicGr "Ethnic Group"

format %12.1f Age
format %12.2f PQmgkg_Sum
format %9.1f predwt


***Generate observation id number (for ease of computing totals, etc)
gen long obsno = _n
order obsno
label variable obsno "Admissions"

***Generate constant in case needed
gen constant = 1
label define total 1 "Total admissions"
label values constant total

***Generate Anemia on admission
gen HbAdmGr7 = 1 if hbmin_First<7
recode HbAdmGr7 (.=0)
tab HbAdmGr7 HbDGr7
label variable HbAdmGr7 "Anaemia at admission (< 7g/dL)"
gen HbAdmGr5 = 1 if hbmin_First<5
recode HbAdmGr5 (. = 0) if hbmin_First>=5 & hbmin_First<.
label variable HbAdmGr5 "Anaemia at admission (< 5g/dL)"
tab HbAdmGr5 HbDGr5

***Generate Severe Thrombocytopenia (platelet count < 50,000/uL)
gen sevThrom = 1 if pltmin<50
recode sevThrom (.=0) if pltmin>=50 & pltmin~=.
local micro = char(181)
label variable sevThrom "Severe thrombocytopenia (platelet count < 50,000/`micro'L)"
list sevThrom pltmin in 1/10
```

```
***Generate Log of WBC, Platelets and Primaquine mg/kg
gen log10hbmin_f= log10(hbmin_First)
label variable log10hbmin_f "Haemoglobin count (first recorded value; log base 10)"
gen log10wbcmin= log10(wbcmin)
label variable log10wbcmin "White Blood Cell count (minimum; log base 10)"
gen log10pltmin= log10(pltmin)
label variable log10pltmin "Platelet count (minimum; log base 10)"
gen log10PQmgkg_Sum= log10(PQmgkg_Sum)
label variable log10PQmgkg_Sum "Sum total dose of Primaquine (mg/kg; log base 10)"
gen Age_log10= log10(Age)
label variable Age_log10 "Age (years; log base 10)"
gen log2wbcmin= log(wbcmin)/log(2)
label variable log2wbcmin "White Blood Cell count (minimum; log base 2)"
gen log2pltmin= log(pltmin)/log(2)
label variable log2pltmin "Platelet count (minimum; log base 2)"
gen log2PQmgkg_Sum= log(PQmgkg_Sum)/log(2)
label variable log2PQmgkg_Sum "Sum total dose of Primaquine (mg/kg; log base 2)"
gen Age_log2= log(Age)/log(2)
label variable Age_log2 "Age (years; log base 2)"

***Generate quintiles of variable list to check for linearity
/*this code assigns labels based on the min and max of each quintile.  For example 1 "100 to 4100" 2 "4110 to 5090" etc */
loc q5 wbcmin pltmin/*Designates variables to quintile and label without decimals*/
foreach v of varlist `q5' {
xtile q5_`v' = `v', nq(5)
su q5_`v', meanonly
loc vmin = r(min)
loc vmax = r(max)
forval j = `vmin'/`vmax' {
su `v' if q5_`v'==`j', meanonly
local call `call' `j' "`: di %3.0f `=r(min)'' to `: di %3.0f `=r(max)'''"
}
label define q5_`v' `call', replace
label values q5_`v' q5_`v'
codebook q5_`v'
}


loc q5 PQmgkg_Sum/*Designates variables to quintile and label with decimals*/
foreach v of varlist `q5' {
xtile q5_`v' = `v', nq(5)
su q5_`v', meanonly
loc vmin = r(min)
loc vmax = r(max)
forval j = `vmin'/`vmax' {
su `v' if q5_`v'==`j', meanonly
local call `call' `j' "`: di %3.2f `=r(min)'' to `: di %3.2f `=r(max)'''"
}
label define q5_`v' `call', replace
label values q5_`v' q5_`v'
codebook q5_`v'
}
label variable q5_PQmgkg_Sum "Sum total dose of primaquine (mg/kg, quintiles)"
label variable q5_wbcmin "White cell count (minimum, quintiles)"
label variable q5_pltmin "Platelet count (minimum, quintiles)"


***Generate 'dummy' variable for later Tabout table output macros
generate dummy = 1

***Generate Female only Pregnancy variable (males are missing)
gen pregWom = Preg if Sex==2
label variable pregWom "Pregnant (females only)"


***Convert Year to labelled categories starting at 1
loc convert Year
foreach v of varlist `convert' {
su `v', meanonly
loc vmin = r(min)
loc vmax = r(max)
loc n 0
forvalues j = `vmin'/`vmax' {
    local cat = `++n'
    local convi `convi' "(" `j' "=" `cat' ")"
    local def `def' `cat' "`j'"
}
recode `v' `convi', gen(`v'Cat)
label define `v'Cat `def', replace
label values `v'Cat `v'Cat
label variable `v'Cat "`v'"
codebook `v'Cat
}

***Define and assign value labels***
label define SOURCE 1 "OP" 2 "OP to IP" 3 "IP", replace
label define AGR4 1 "< 1 year" 2 "1 to < 5 years" 3 "5 to <15 years" 4 "15+ years", modify
```

```
label define yesno 0 "No" 1 "Yes"
label values aa_any cq_any cl_any dhp_any dox_any ivart_any ivq_any oralq_any pq_any sp_any ///
            MalNut Bleeding Preg pregWom HbDGr7 HbDGr5 MalNutD VivaxDeath HbAdmGr5 HbAdmGr7 sevThrom ///
            MalariaNext365 PFNext365 PVNext365 PMNext365 MIXNext365 AdmNext14 DiedNext14 DiedNext365 DiedNext7 yesno
```

Sort database by hrn, date of admission and date of discharge, in that order
```
sort hrn doa dod
```

*Round age to two decimal places (the current long data format causes difficulties with 'if' statements using scalars where Age < 1, apparently)
```
replace Age = round(Age, .01)
```

*Generate sex variable differentiating between pregant and not pregnant females
```
gen sexPreg= Sex+Preg
recode sexPreg (1=3) (2=1) (3=2)
lab def sexPreg 1 "Female (not pregnant)" 2 "Female (pregnant)" 3 "Male"
lab val sexPreg sexPreg
lab var sexPreg "Sex and pregnancy status"
codebook sexPreg
```

*Generate combined HRN and MalCluster variable for generating Riskset graphics

```
gen hrnmal = hrn + (MalCluster/10) if MalCluster <10
tempvar maxmal
by hrn: egen `maxmal' = max(MalCluster)
replace hrnmal = hrn +(MalCluster/(`maxmal'+1)) if `maxmal' >=10
list hrnmal if `maxmal' >10
```

*Generate early exit variables with "t1" as day zero instead of "t.5", for equal intervals when plotting

```
gen AdmFU15 = Adm_FU14+1
recode AdmFU15 (1.5=1)
lab var AdmFU15 "Admission: Follow up to 15 days"

gen DiedFU15 = Died_FU14+1
recode DiedFU15 (1.5=1)
lab var DiedFU15 "Died: Follow up to 15 days"
```

*generate estimated date of admission / death
```
gen doX_adm = doa+AdmFU15
gen doX_died = doa+DiedFU15
```

*Generate pltCat
```
recode pltmin  (20/50=2 "20 to 50"), gen(pltCat)
replace pltCat = 1 if pltmin<20
replace pltCat = 3 if pltmin>50&pltmin<.
list pltmin pltCat
label define pltCat 2 "20 to 50" 1 "< 20" 3 "> 50", replace
label variable pltCat "Platelet count (minimum))"
label variable pltCat "Platelet count ( x 1000s/µL)"
```

*Generate reduced categories of MalPres
```
recode MalPres (1=1 "1") (2=2 "2") (3/9=3 "3+"), gen(MalPres_cat)
label variable MalPres_cat "Number of malaria events in episode"
```

*Gen white blood cell above, below or within normal range indicator
```
generate whiteCat = .
replace whiteCat=1 if Age<0.0192 & (wbcmin<20000) & wbcmin!=.
replace whiteCat=2 if Age<0.0192 & (wbcmax>40000) & wbcmin!=.
replace whiteCat=1 if (Age>=0.0192 & Age<0.0385) & (wbcmin<5000) & wbcmin!=.
replace whiteCat=2 if (Age>=0.0192 & Age<0.0385) & (wbcmax>21000) & wbcmin!=.
replace whiteCat=1 if (Age>=0.0385 & Age<0.25) & (wbcmin<5000) & wbcmin!=.
replace whiteCat=2 if (Age>=0.0385 & Age<0.25) & (wbcmax>20000) & wbcmin!=.
replace whiteCat=1 if (Age>=0.25 & Age<1) & (wbcmin<5000) & wbcmin!=.
replace whiteCat=2 if (Age>=0.25 & Age<1) & (wbcmax>15000) & wbcmin!=.
replace whiteCat=1 if (Age>=1 & Age<5) & (wbcmin<50000) & wbcmin!=.
replace whiteCat=2 if (Age>=1 & Age<5) & (wbcmax>12000) & wbcmin!=.
replace whiteCat=1 if Age>=5 & (wbcmin<4000) & wbcmin!=.
replace whiteCat=2 if Age>=5 & (wbcmax>10000) & wbcmin!=.
replace whiteCat=0 if whiteCat==. & wbcmin!=.
replace whiteCat=0 if whiteCat==1 & whiteout==0
label define whiteCat 0 "within normal range" 1 "below normal range" 2 "above normal range"
label values whiteCat whiteCat
tab whiteCat
label variable whiteCat "WBC count"
```

*Generate Species variable omitting P.ovale
```
gen SpeciesX = Species if Species !=3
label variable SpeciesX "Species"
label copy Species SpeciesX
label values SpeciesX SpeciesX
label define SpeciesX 3 "", modify
```

```
*Generate AGR4 with age category 4 hard coded as baseline (for conveniance)
recode AGR4 (4=0), gen(AGR4_4b)
label copy AGR4 AGR4_4b
label values AGR4_4b AGR4_4b
label variable AGR4_4b "Age"
label define AGR4_4b 1 "< 1 year" 2 "1 to < 5 years" 3 "5 to <15 years" 0 "15+ years (baseline)", replace

*create EthnicX with Highland as baseline reference
codebook Ethnic
recode Ethnic (1=3) (2=1) (3=2), gen(EthnicX)
codebook EthnicX
label copy Ethnic EthnicX
label variable EthnicX EthnicX
codebook EthnicX
label values EthnicX EthnicX
label define EthnicX 3 "Non Papuan" 1 "Highland" 2 "Lowland", replace
codebook EthnicX
tab Ethnic EthnicX
tab Ethnic EthnicX, missing
label variable EthnicX "Ethnicity"

*IV Quinine and Artesunate variable
gen ivArt = 1 if TreatGr_First ==5
replace ivArt = 0 if TreatGr_First ==6
lab copy TreatGr_First ivArt
label define ivArt 1 "ivArtesunate +/-DHP,Cq,SP" 0 "ivQuinine +/-DHP,Cq,SP", replace
label values ivArt ivArt
lab var ivArt "IV treatment"
codebook ivArt

*oral quinine and DHP variable
gen oral_v_dhp = 1 if TreatGr_First==4
replace oral_v_dhp =0 if TreatGr_First == 2
tab oral_v_dhp TreatGr_First
```