```
/*-------------------------------------------------------------------------------------------------------------------------------
Example stata code used to produce results in the Emerging Risk Factors Collaboration (ERFC) statistical
methods paper [Int J Epidemiol 2010]
The stata programs used are available on the ERFC website: http://ceu.phpc.cam.ac.uk/research/erfc/methods/
and can be installed from stata in the standard way using net install command e.g.
net from http://ceu.phpc.cam.ac.uk/software/erfc/
net install adjmeta
net install mvshape
net install phtest
net install stsetage
...etc
-------------------------------------------------------------------------------------------------------------------------------*/

*(i) Example of code used to produce results on Table 2 (hazard ratios per 1 g/L higher baseline fibrinogen).
* do some variable transformations required below
gen ages2 = ages^2   /* quadratic term for age at screening */
egen aband5 = cut(ages), at(20(5)90) /* 5-year age bands */

* define adjustment models to be compared as global macros
global adj1 = "ages"
global adj2 = "i.aband5"
global adj3 = ", strata(aband5)"    /* add stratification by 5-year age bands */
global adj4 = "i.sex*ages"
global adj5 = "ages ages2"
global adj6 = "i.sex*ages i.sex*ages2"
global adj7 = "ages i.smallbin tchol sbp bmi"

* stage1: get study specific log RRs per 1 g/L higher fibrinogen adjusted as above
* approach1: easy specification but computationally slower due to keeping whole dataset in memory
bysort studydes: adjmeta, studyid(cohort) subjid(cohortid) studydes(studydes) ///
timevar(duration) failure(ep_chd==1) strata(sex trialarm) maxadj(7) adjseq(1(1)7) ///
fitmodel(bb) conexp(fib) saving(output\adj_rr_fib) replace
* approach2: faster computation by keeping in memory only the data for each cohort in turn
preserve
local resfile = "output\adj_rr_fib.dta"   /* file on disk to append study-specific results to */
tempfile tempres                          /* stata temporary file to save study-specific results */
local k=1                                 /* loop counter */
levelsof cohort, local(cohorts)
foreach coh of local cohorts {
   * call adjmeta program to calculate the study-specific log RRs
   keep if cohort=="`coh'"       /* keep data for current cohort in memory, computationally faster */
   adjmeta, studyid(cohort) subjid(cohortid) studydes(studydes) ///
   timevar(duration) failure(ep_chd==1) strata(sex trialarm) maxadj(7) ///
   adjseq(1(1)7) fitmodel(bb) conexp(fib) saving(`tempres') replace
   if `k'==1 {
      use `tempres', clear
      save "`resfile'", replace  /* save permanent file to disk on first call */
   }
   else {
      use "`resfile'", clear
      append using `tempres'     /* append to permanent file on disk subsequently */
      save, replace
```

```
   }
   local ++k                    /* increment loop counter */
   restore, preserve
}
restore
* stage2: combine study-specific estimates by random effects meta-analysis
use output\adj_rr_fib.dta, clear
bysort stdesign: poolmeta, randomi saving(output\adj_rr_fib_pool_randomi) replace


*(ii) Example of code used to produce results on Table 1 (hazard ratios per 1 SD higher baseline fibrinogen).
* generate exposure variables of interest and label them for nice output
gen lnfib = ln(fib)
loneway fib cohort                          /* get pooled within study SD from large one-way ANOVA on cohort */
gen sdfibov = r(sd w)                       /* the SD estimate is returned in r(sd w) macro */
loneway lnfib cohort
gen sdlnfibov = r(sd_w)
bysort cohort: egen sdfibcoh = sd(fib)      /* get study-specific SDs */
bysort cohort: egen sdlnfibcoh = sd(lnfib)
gen fibovsd = fib/sdfibov                   /* scale fibrinogen by pooled within study SD and study-specific SD */
gen fibcohsd = fib/sdfibcoh
gen lnfibovsd = lnfib/sdlnfibov             /* scale log fibrinogen by pooled within study SD and study-specific SD */
gen lnfibcohsd = lnfib/sdlnfibcoh
summ fib
gen cfib = fib - r(mean)                    /* center fibrinogen to reduce correlation betw main eff and quadratic term */
gen cfib2 = cfib^2                          /* quadratic term for fibrinogen */
label variable fib "Fibrinogen (g/L)"
label variable lnfib "Log fibrinogen (log g/L)"
label variable fibovsd "Standardised fibrinogen, overall SD (g/L)"
label variable fibcohsd "Standardised fibrinogen, study-sp SD (g/L)"
label variable lnfibovsd "Standardised log fibrinogen, overall SD (log g/L)"
label variable lnfibcohsd "Standardised log fibrinogen, study-sp SD (log g/L)"
label variable cfib "Centered fibrinogen"
label variable cfib2 "Centered fibrinogen^2"

* stage1: get study specific log RRs for each fibrinogen transformation adjusted for age
global adj1 = "ages"
local expvlist = `"fib fibovsd lnfibovsd fibcohsd lnfibcohsd "cfib cfib2""' /* list of exposure variables */
preserve
local resfile = "output\adj_rr_fib_compare.dta" /* file on disk to append study-specific results to */
tempfile tempres                            /* stata temporary file to save study-specific results */
local k=1                                   /* loop counter */
local m=1                                   /* exposure variable counter */
foreach expvar in `expvlist' {
   local rrfor = cond("`expvar'"=="cfib cfib2", 2, 1) /* interested in quadratic term only */
   levelsof cohort, local(cohorts)
   foreach coh of local cohorts {
      * call adjmeta program to calculate the study-specific log RRs
      keep if cohort=="`coh'"       /* keep data for current cohort in memory, computationally faster */
      adjmeta, studyid(cohort) subjid(cohortid) studydes(studydes) ///
      timevar(duration) failure(ep_chd==1) strata(sex trialarm) maxadj(1) ///
      adjseq(1(1)1) fitmodel(bb) conexp(`expvar') rrfor(`rrfor') saving(`tempres') replace
```

```
      if `k'==1 {
         use `tempres', clear
         gen expvarno = `m'
         save "`resfile'", replace  /* save permanent file to disk on first call */
      }
      else {
         use "`resfile'", clear
         append using `tempres'     /* append to permanent file on disk subsequently */
         replace expvarno = `m' if expvarno==.
         save, replace
      }
      local ++k                     /* increment loop counter */
      restore, preserve
   }
   local ++m                        /* increment exposure variable counter */
}
restore
* stage2: combine study-specific estimates by both random and fixed effects meta-analysis
use output\adj_rr_fib_compare.dta, clear
bysort stdesign expvarno expvar: poolmeta, randomi saving(output\adj_rr_fib_compare_pool_randomi) replace
bysort stdesign expvarno expvar: poolmeta, fixedi saving(output\adj_rr_fib_compare_pool_fixedi) replace


*(iii) Example of code used to produce results on Table 3 (interactions between fibrinogen and risk factors).
* stage1: get study specific log RRs for continuous interaction with fibrinogen
global adj1 = "fib"
global adj2 = "fib ages"
global adj3 = "fib ages i.smallbin"
global adj4 = "fib ages i.smallbin sbp"
global adj5 = "fib ages i.smallbin sbp bmi"
global adj6 = "fib ages i.smallbin sbp bmi tchol"
gen femsex = sex==2 if sex~=.
preserve
local intvlist = `"ages sbp bmi tchol femsex"' /* list of interaction variables */
local resfile = "output\inter_rr_fib.dta" /* file on disk to append study-specific results to */
tempfile tempres                          /* stata temporary file to save study-specific results */
local k=1                                 /* loop counter */
local m=1                                 /* exposure variable counter */
foreach var in fib `intvlist' {
   if "`var'"=="fib" {
      local expvar = "`var'"              /* also get fibrinogen main effect to be used for estimating between study interaction for sex */
      local rrfor = 1
   }
   else {
      gen fib_`var' = fib*`var'           /* generate continuous interaction with fibrinogen */
      local expvar = "`var' fib_`var'"    /* intvar main effect and fib interaction */
      local rrfor = 2                     /* interested in log RR for interaction term only, fib main effect is in the global adjustments
macro */
   }
   tempfile cohfile
   save `cohfile'
   levelsof cohort, local(cohorts)
```

```
   foreach coh of local cohorts {
      * call adjmeta program to calculate the study-specific log RRs
      use `cohfile', clear
      keep if cohort=="`coh'"       /* keep data for current cohort in memory, computationally faster */
      adjmeta, studyid(cohort) subjid(cohortid) studydes(studydes) ///
      timevar(duration) failure(ep_chd==1) strata(sex trialarm) maxadj(6) ///
      adjseq(6(1)6) fitmodel(bb) conexp(`expvar') rrfor(`rrfor') saving(`tempres') replace
      if `k'==1 {
         use `tempres', clear
         gen expvarno = `m'
         save "`resfile'", replace  /* save permanent file to disk on first call */
      }
      else {
         use "`resfile'", clear
         append using `tempres'     /* append to permanent file on disk subsequently */
         replace expvarno = `m' if expvarno==.
         save, replace
      }
      local ++k                     /* increment loop counter */
   }
   local ++m                        /* increment exposure variable counter */
   restore, preserve
}
collapse (mean) femprop=femsex, by(cohort)   /* get proportion of females within each cohort */
tempfile femprop
save `femprop'
use output\inter_rr_fib.dta, clear
joinby cohort using `femprop'
sort cohort usemax model expvarno
save output\inter_rr_fib.dta, replace
restore
* stage2: combine study-specific estimates by random effects meta-analysis
use output\inter_rr_fib.dta, clear
bysort stdesign expvarno expvar: poolmeta, randomi saving(output\inter_rr_fib_pool_randomi) replace
metareg lnrr femprop if expvar=="fib", mm wsse(selnrr)      /* between study sex interaction */
preserve
keep if expvar=="fib_femsex"
set obs `=_N+1'
replace cohort = "BETW" in `=_N'
replace lnrr = _b[femprop] if cohort=="BETW"
replace selnrr = _se[femprop] if cohort=="BETW"
gen inttype = cohort=="BETW"
label define inttype 1 "Between study" 0 "Within study"
label values inttype inttype
metan lnrr selnrr, randomi by(inttype) lcols(cohort nfail)  /* combine between and within study interaction terms for sex */
restore


*(iv) Example of code used to test proportional hazard assumption (Table 4).
phtest fib, method(sch2stage) studydes(studydes) timevar(duration) failure(ep_chd) ///
strata(sex trialarm) adjvar(ages) studyid(cohort) saving(output\phtest_res_sch2stage)
```

```
phtest fib, method(int2stage) studydes(studydes) timevar(duration) failure(ep_chd) ///
strata(sex trialarm) adjvar(ages) studyid(cohort) saving(output\phtest_res_int2stage)


*(v) Example of code used to assess the shape of association between fibrinogen and CHD (Figure 2).
local epvar = "ep_chd"  /* endpoint */
local ngrp = 5          /* quintile exposure groups */
local adjvar = "ages"   /* adjustment variables, here age only */
mvshape fib, studyid(cohort) subjid(cohortid) studydes(studydes) timevar(duration) failure(`epvar'==1) ngrp(`ngrp') ///
matchvar(match) adjvar("`adjvar'") strata(sex trialarm) method1(mm) method2(mm) esave(N_sub N_fail) noauglist ///
savedata("output\mvshape_fib_`epvar'_data") saveres("output\mvshape_fib_`epvar'_results")


*(vi) Example of code used to assess interaction between fibrinogen and age-at-risk for CHD (Figure 3).
preserve
keep if inlist(studydes, 1, 2, 5)   /* only cohorts, case-cohorts, and clinical trials contribute to time-to-event analysis */
xtile fib5 = fib, nq(5)             /* baseline fibrinogen quintiles */
foreach epvar in ep_chd {
   capture log close
   log using logfiles\mvshape_fib_agespec_`epvar'.log, replace

   * call my stsetage program to set up the time-to-event data in format suitable for estimating age-at-risk specific log RRs
   stsetage, subjid(subjectid) agevar(ages) agerisk(40 60 70) agecent(55 65 75) timevar(duration) failure(`epvar'==1)

   * call mvshape program to get study-specific age-at-risk log RRs and then combine them by multivariate meta-analysis
   gen group = agegrp*100 + fib5       /* code combination of age-at-risk group and baseline fibrinogen quintiles */
   local refgrp = 205                  /* choose reference category as top fifth on age 60-69, observed in most studies! */
   local transref = 101                /* to transform reference to bottom fifth in age 40-59, for easier interpretation */
   local adjvar = "nuis1 nuis2 nuis3"  /* nuisance covariates to adjust for continuous age effect within the age-at-risk groups */
   mvshape fib, studyid(cohort) subjid(subjectid) studydes(studydes) timevar(duration) failure(`epvar'==1) nostset noidcheck ///
   catexp(group) refgrp(`refgrp') transref(`transref') adjvar("`adjvar'") strata(sex trialarm) method1(mm) method2(mm) ///
   esave(N_sub N_fail) noauglist savedata("output\mvshape_fib_agespec_`epvar'_data") saveres("output\mvshape_fib_agespec_`epvar'_results")

   * replot the graph with regression lines fitted through the age-specific log RRs
   use "output\mvshape_fib_agespec_`epvar'_results.dta", clear
   gen fib5 = mod(fib15, 100)          /* recover fibrinogen quintiles from the coded combination of age-group*quintile variable */
   gen agegrp = (fib15 - fib5)/100     /* recover age-group coding as well */
   sort agegrp fib5
   levelsof agegrp, local(agegrps)
   foreach agegp of local agegrps {
      regress flnrr pmfib if agegrp==`agegp' [aw = invvarf]
      predict _linfit
      replace linfit = _linfit if agegrp==`agegp'
      drop _linfit
   }
   replace explinfit = exp(linfit)
   label define agegrp 1 "40-59" 2 "60-69" 3 "70+"
   label values agegrp agegrp
   decode agegrp if fib5==5, gen(label)
   sort agegrp fib5
   save, replace
```

```
   * call mvshapegr program to replot the graph with age-specific regression lines
   mvshapegr, yscale(log) ylabel(1 2 4 8 16, angle(horiz) nogrid) xlabel(2.0(0.5)4.5, angle(horiz) format(%9.1f)) ///
   fitline sort(agegrp pmfib) mlabel(label) ytitle("Hazard ratio (log scale)") xtitle("Baseline fibrinogen (g/L)")

   log close
   restore, preserve
}
restore
```